UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

FACULTY OF BUSINESS ADMINISTRATION AND ECONOMICS

Working Paper Series

Working Paper No. 2021-02

Anonymity and Self-Expression in Online Rating Systems - An Experimental Analysis

Britta Hoyer and Dirk van Straaten

April 2021

WiWi

Anonymity and Self-Expression in Online Rating Systems - An Experimental Analysis

B. Hoyer[a,*], D. van Straaten[b]

[a] *Paderborn University, Department of Economics, Warburger Str. 100, 33098 Paderborn, Germany*
[b] *Paderborn University, Department of Management, Heinz Nixdorf Institute, Warburger Str. 100, 33098 Paderborn, Germany*

Abstract

Customer reviews are a fundamental part of online markets to establish trust between customers and sellers. Sharing experiences about products and services, however, is accompanied with the (sometimes involuntary) sharing of personal information. This is in conflict with growing concerns of data security calling for more privacy in online settings. Anonymous reputation systems could be one instrument to ensure more privacy in such markets. However, the impact of anonymity on the propensity to leave reviews is unclear. In this experimental study we therefore analyze whether the degree of anonymity of customers affects their propensity to leave reviews in an online market. We find that the amount of ratings drops significantly when subjects are anonymous pointing to self-expression as a driver of customer reviews. Moreover, we find that altruistic subjects are not affected by the introduction of anonymity and, hence, provide significantly more reviews compared to non-altruists under anonymity. When we remove the veil of anonymity, this difference between altruists and non-altruists disappears and, overall, market outcomes increase.

Key Words: Customer Rating, Altruism, Public Good, Anonymity, Reputation

JEL Classification: C91, D64, H41, L86

This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Centre 901 "On-The-Fly Computing" (SFB 901) under the project number 160364472-SFB901.

* Corresponding author.
Email addresses: britta.hoyer@uni-paderborn.de (B. Hoyer), dirk.van.straaten@uni-paderborn.de (D. van Straaten)

1. Introduction

Within the last two decades, online marketplaces have been one of the biggest success stories of the internet. Today E-commerce accounts for almost 20% of global retail sales, with a sales value of 3.5 trillion US Dollar in 2020. Over 50% of these sales are made via online market platforms such as amazon.com (Clement, 2020). One of the enabling factors of this rise of e-commerce was the introduction of feedback and reputation systems, first used by eBay and then copied in one form or another by virtually every online market platform (Tadelis, 2016). The inherent anonymity of online markets leads to issues of hidden information (sellers may know that the object they sell is broken but do not reveal it) and hidden action (a seller might not pack a good correctly, increasing the chance of it arriving damaged) and the usual trust building devices of traditional marketplaces are often either impossible (e.g. the buyer cannot touch the good before buying) or less effective in large scale online settings (Dellarocas, 2003). Consequently, this may lead to a so-called *market for lemons* (Akerlof, 1970) and therewith to market failure. Thus, well-functioning reputation systems are vital for online markets. However, leaving a rating costs time and effort for the individual writing the review. At the same time, society as a whole can benefit from the review once it is published. Therefore, ratings are a public good and free-riding behavior can be expected (Bolton et al., 2004). Nevertheless, numerous reviews can be observed for any kind of product being sold online. For example, the Amazon Review Data 2018 (Ni, 2019) includes 233.1 million reviews for the time span from 1996 to 2018. Given the importance of reputation systems, the amount of reviews written, and the unknown motives for customers to write reviews, the question of what actually drives customers to write reviews has been analyzed frequently in the literature. Studies show that the main motives for writing customer reviews are a sense of belonging, altruism, self-expression, and economic incentives (Cheung and Lee, 2012; Hennig-Thurau et al., 2004). [1]

While rating systems are important for online markets and can significantly influence buyers' and sellers' behavior (e.g., Chevalier and Mayzlin (2006); Lafky (2014)), by leaving reviews customers also post personal information online. Additionally, reputation systems are used to monitor users' behavior and they thus pose a threat to privacy (Voss, 2004). To avoid this privacy issue, research in computer science shows the technical feasibility of anonymous reputation systems, in which additionally it is guaranteed that only customers who bought the product can also rate it (e.g., Blömer et al. (2015); Bemmann et al. (2018)).[2] However, while anonymous rating systems may be better in terms of privacy, the impact of anonymity on the propensity of customers to write reviews is not clear (Rockenbach and Sadrieh, 2012). Even though

[1] Bolton et al. (2013) show that in markets where both buyers and sellers can give feedback, reciprocity also plays a large role, both for writing reviews as well as for the content of the review. Halliday and Lafky (2019) show that side payments of sellers also initiate reciprocal behavior of customers.

[2] While Bemmann et al. (2018) show the technical feasibility of such a review system, we are not aware of any real world implementation of it.

anonymity should not have an impact on the purely altruistic motives for writing reviews it might affect the propensity to rate by inhibiting self-expression.[3] Therefore, in this first study, we focus on the blunting of the motive of self-expression through anonymity. We furthermore investigate whether anonymity impacts the behavior of altruistic and non-altruistic subjects in the same way.

As we do not see anonymous reputation systems implemented in online markets so far, we cannot collect data from existing platforms to analyze the effect of anonymity on customers' likelihood to leave reviews. Therefore, we conduct a laboratory experiment, which enables us to vary the degree of anonymity in reputation systems – from pseudonymity to anonymity – and compare the propensity of customers to leave reviews. Since the participants' possibility to express themselves is manipulated in this dichotomous manner, we eliminate potentially distorting effects of heterogeneous preferences for sharing more or less private information by design. Additionally, by using a laboratory experiment, we can control for other outside factors that might influence results such as prices, product groups, visualization effects etc. Therefore, we can clearly answer the question whether anonymity inhibits the self-expression motive in reputation systems and, hence, affects customers' propensity to write reviews. Additionally, we can also show whether anonymity impacts the behavior of altruistic subjects and non-altruistic subjects to the same degree, and the overall impact on market outcomes.

The rest of the paper is organized as follows: In Section 2 the hypotheses on the rating behavior of the participants with respect to the motives self-expression and altruism and their effect on market outcomes are derived. In Section 3 we present our experimental design and procedure. Section 4 reports the results with focus on the tests of the hypotheses. Discussing the results and limitations, Section 5 concludes the paper. Appendix A provides the instructions of the experiment.

2. Literature and Predictions

Writing a review can be considered as providing a public good and therefore, while being useful for society as a whole, it actually does not have any direct benefits for the reviewer. Hence, analyzing the motives for writing reviews is extremely important when considering changes to reputation systems. In the literature self-expression has been identified as one of the major reasons for writing reviews. In particular, Hennig-Thurau et al. (2004) identify beside other motives extraversion and positive self-enhancement as drivers of reviewing behavior. The results of Cheung and Lee (2012) point into a similar direction by identifying reputation, i.e., the improvement of status and reputation in the profession, as driving motives. In Wu (2019) interviews with reviewers from Amazon UK are conducted to investigate their motives for

[3] We do not expect the other main motives for writing customer reviews (sense of belonging and economic incentives) to be affected by anonymity. However, even if other motives such as sense of belonging are affected by anonymity its impact should be negative as well resulting in an overall negative impact of anonymity on the willingness to publish a review.

providing customer reviews. Among other motives, ranking and status recognition are identified as important factors. Additionally, the data propose a crowding-in of enjoyment and status recognition. All this shows that self-expression is a strong motive to rate a product. We argue that introducing anonymity to a rating system blunts this motive of self-expression, as the provision of ratings is no longer linked to ones identity. Hence, we hypothesize:

Hypothesis 1. *In anonymous reputation systems fewer ratings are observed in comparison with non-anonymous reputation systems.*

Besides self-expression also altruism has been shown to be affective when customer reviews are provided. As reviews are a public good, customers providing reviews act - to some extent - altruistically. Indeed, enjoyment of helping other customers is identified as one of the motivating factors in writing online reviews by studies employing quantitative (Cheung and Lee, 2012) and qualitative (Wu, 2019) methods. Additionally, Hennig-Thurau et al. (2004) provides evidence that concern for other customers is an important motive for the rating of products. They show that negative and positive experiences can trigger customer reviews and result in customers warning others of bad products or helping them to find the right products. Altruism based on positive and negative experiences is identified in Munzel and Kunz (2014) as the most important driver for writing reviews.

Altruism is defined by Eisenberg (1996) as "voluntary behavior that is intended to benefit another and is not motivated by the expectation of external reward". Hence, by definition an altruistic motive should not be influenced by anonymity. If a customer provides a review for altruistic reasons or to help other customers, it does not matter whether he can be identified.

Hypothesis 2. *The introduction of anonymity does not affect altruists' publishing of customer reviews.*

An important characteristic of reputation systems is that providing information is similar to contributing to a public good (Avery et al. 1999). In public good games, the welfare is maximized when every subject contributes. However, the individuals' profit maximizing strategy is free-riding. The results of Bardsley and Moffatt (2007) indicate that subjects can be classified into four groups in their public goods game: strategists, free-riders, reciprocators, and altruists, among which altruists are the smallest group. Assuming support for Hypothesis 2, we expect that altruists are not affected by the introduction of anonymity to the reputation system. However, assuming support for Hypothesis 1, we do expect to see less free-riders and an increase of strategists and reciprocators in non-anonymous reputation systems, which would overall lead to more information on the market. Corresponding to the welfare maximizing perspective in public goods games, economic theory states that especially on markets with information asymmetries efficacy increases when information is provided (see, e.g. Akerlof (1970)). In particular, Resnick et al. (2006) identifies that customers' willingness-to-pay in auctions increases when more information on the seller is available.

That is, the more information is on the market, the higher are the reservation prices, which correspond to utilities of the customers. Hence, we hypothesize:

Hypothesis 3. *Increasing the information on the market will lead to a higher market outcome.*

Figure 1 graphically depicts the reasoning behind our three hypotheses. As leaving a rating after the purchase can be considered a public good, the incentives for writing reviews need to be taken into account. The main motives for writing customer reviews that have been discussed in the literature are a sense of belonging, altruism, self-expression, and economic incentives (Cheung and Lee, 2012; Hennig-Thurau et al., 2004). In this experiment, we focus on the aspect of self-expression, which might be blunted through anonymity (Hypothesis 1) and on altruism, which should lead to more ratings but should not be affected by anonymity (Hypothesis 2). Finally, we also consider the market outcome as such, as we know from standard economic theory (see, e.g., Akerlof (1970)) that market outcomes increase when more information is provided. Since more ratings mean that there is significantly more information on the market, we expect market outcomes to increase (Hypothesis 3).

Figure 1: Overview of Hypotheses

3. Experimental Design and Procedure

In each of the eight conducted sessions, all 24 participants are customers on a stylized market place, who are asked to choose one out of 7 heterogeneous computerized sellers in each of 10 periods. The sellers all sell the same product. However, the satisfaction gained from the product differs for the participants. Therefore, the participants are randomly allocated to 3 different, evenly sized customer groups (X, Y and Z), which differ only in the satisfaction they receive from the sellers' products. We implement these customer groups to reflect different expectations and demands customers may have with respect to a product.[4] The level of satisfaction with a product customers receive directly translates into their payoff. To avoid issues of different prices, or not buying the product at all, in this stylized market place, there are no prices for products. After choosing the seller, participants can decide whether they want to publish their satisfaction with the seller. In the experiment, customers simply have the option to publish their satisfaction with a

4 This may depict differences between customers who focus , e.g., on quality-per-money ratio and those who buy the product with the highest quality.

seller or not. To prevent untruthful ratings or different information value of different ratings, we implement the review process as a binary choice. However, it is costly for them to publish their satisfaction. We implement the costs, as usually publishing a review is associated with costs as writing a review costs time and effort. If customers decide to publish a rating, it will become public, helping other participants to make an informed decision in the following periods.

In order to analyze the impact of anonymity on the propensity of customers to leave a rating, we designed two treatments. In the *anonymity treatment*, subjects are assigned a buyer number. This ID is private information and will not be published when the subject decides to reveal his satisfaction. In the *identity treatment*, subjects choose a personal identifier at the beginning of the experiment (pseudonym). If they reveal their satisfaction, it will be published with their pseudonym. Also addressing the self-expression motive, in the identity treatment a ranking of the participants' number of published ratings is shown at the end of each period, enabling them to build up a reputation as a frequent reviewer. In the anonymity treatment this is not the case, as ratings are anonymous.

After the personal identifier is determined at the very beginning of the market experiment, two phases follow, which are repeated in each of the 10 periods (cf. Figure 2).

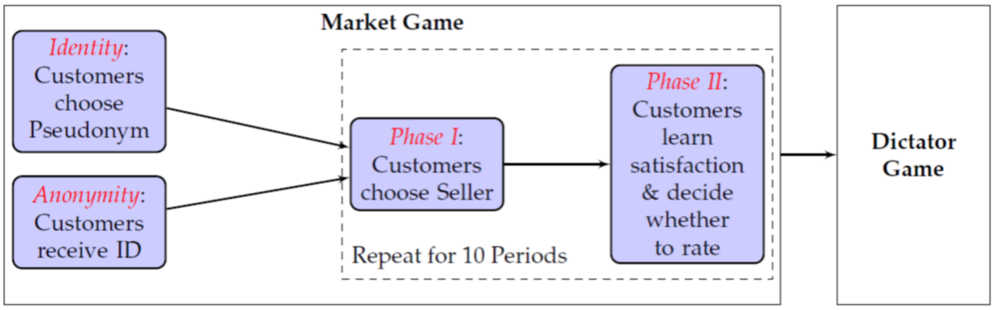

Figure 2: Set-up of the Experiment

In *Phase I*, each customer can choose exactly one product and receives an uncertain satisfaction between 2 and 20 points for the chosen product. The satisfaction u_i of customer i is determined as follows:

$$u_i = S_{ij} + \epsilon_i,$$

where $S_{ij} \in [4, 18]$[5] is the expected quality provided by seller j for i's customer group. I.e., the 7 heterogeneous sellers, called A to G, provide different satisfaction levels to the customers depending on their customer group (cf. Table 1).[6] The quality variation term $\epsilon_i \in \{-2, -1, 0, 1, 2\}$ is randomly drawn for each

5 We chose this range so that buyers cannot earn negative points even if they decide to rate the seller's product.

6 Note that the actual satisfaction of customers still depends on the quality variation term as well, such that satisfaction between 2 and 20 points are in fact possible.

customer in each period to increase uncertainty. Whereas the different customer groups reflect different general attitudes towards a product, the quality variation term is used to reflect personal satisfaction with a product as well as the chance of receiving a lemon good from an otherwise good seller. Since we abstract from prices, there are thus only three sources of information available for customers to base their decision on: First, sellers give an indication of how satisfied customers will be with the product on average across all three customer groups (Average Satisfaction Levels).[7] This information is fixed for the whole experiment. Second, customers can observe their past decisions and the associated satisfaction. Third, published ratings of other customers from preceding periods can be available.

		Seller							
		A	B	C	D	E	F	G	Average
Customer group	X	18	8	4	14	6	10	13	10.4
	Y	4	12	16	10	13	8	10	10.4
	Z	4	14	11	6	13	16	9	10.4
	Average Satisfaction Level	8.7	11.3	10.3	10	10.7	11.3	10.7	

Table 1: Satisfaction levels provided by sellers A to G for the three customer groups X, Y, and Z.

In *Phase II*, each customer is informed about the actual satisfaction he received from the product and can choose whether to leave a truthful rating afterwards. When rating, he reveals which customer group he belongs to and how much satisfaction he received from the product of the chosen seller. How the ratings are published depends on the treatment, as summarized in Table 2.

Published Rating in Identity Treatment	Published Rating in Anonymity Treatment
Nick: I belong to customer group *X* and my level of satisfaction with seller *D* was *12* points.	A customer from customer group *X* chose seller *D* and the level of satisfaction was *12* points.

Table 2: Different presentation of published satisfaction between treatments

Revealing the satisfaction costs $c = 2$ Experimental Currency Units (*ECUs*). As explained above, we implement costs for ratings to model the fact that writing a review involves time and effort costs. Each

[7] Average Satisfaction Levels are reported for each seller j as $\overline{S_j} = \frac{1}{3} * (S_j^X + S_j^Y + S_j^Z)$, where S_j^X is seller j's expected quality provided to customer group X.

customer i's payoff per period t thus depends on the satisfaction received from the product they chose and whether they incurred costs c for rating the seller. At the end of the 10 periods, total payoff Π_i for customer i is calculated as the sum of the payoffs per period (in ECUs):

$$\Pi_i = \sum_{t=1}^{10} \Pi_i^t, \text{where } \Pi_i^t = u_i^t - c_i^t.$$

After the market game we conduct a dictator game to elicit participants' attitudes towards altruism. In the dictator game each participant is completely anonymous – the pseudonyms or assigned IDs are no longer displayed. Each participant acts simultaneously as dictator and receiver: i.e, he is asked to split 20 ECUs (in steps of 2 ECUs) between himself and an unknown randomly matched partner. After everybody decided how to split the amount, it is randomly determined whether the own or the partner's decision is implemented and the according ECUs are added to the payoff Π_i from the market experiment. Finally, a questionnaire is used to gather demographics and individual characteristics.

The experiment was conducted in the computerized experimental lab (BaER-Lab) of Paderborn University in June and October 2018. The instructions can be found in the appendix.[8] We used the experimental software 'z-Tree' (Fischbacher, 2007). Subjects were recruited via the online recruiting system ORSEE (Greiner, 2015) at Paderborn University from a pool of approximately 2800 voluntary students. We conducted eight experimental sessions (four per treatment) in which 192 subjects participated. The sessions lasted 58 minutes on average and the participants generated an average payment of 12.07 Euros (including the show-up fee of 2.50 Euros, an average earning of 8.90 Euros in the market game resp. 0.67 Euros in the dictator game).

4. Analysis and Results

We provide some overall summary statistics across both treatments concerning demographics as well as average earnings in Table 3. We do not find treatment effects for gender ($\chi^2: \ z = 0.7840, \ p = 0.376$), age (Mann-Whitney-U test: $z = -1.369, \ p = 0.1709$), or field of studies ($\chi^2: \ z = 3.4034, \ p = 0.334$).

The overall rating behavior is summarized in Table 4. Subjects published on average 1.56 ratings in the identity treatment and 1.03 ratings in the anonymity treatment. The histogram of ratings per subject are depicted in Figure 3 separately for both treatments.

[8] As the experiment was conducted in German, we translated the instructions. Appendix A.1 gives the translations of the instructions for the anonymity treatment and Appendix A.2 gives the translation of the instructions for the identity treatment.

Summary Statistics	Identity Treatment (n=96)	Anonymity Treatment (n=96)	Total Sample
Male	43%	36%	40%
Mean Age	23.95	22.96	23.45
Economics student	32.3%	28.1%	30.2%
Education student	39.6%	31.3%	35.4%
Engineering student	15.6%	22.9%	19.3%
Humanities student	12.5%	17.7%	15.1%
Average Earning Market Game	EUR 8.97	EUR 8.83	EUR 8.90
Average Earnings	EUR 12.14	EUR 12.00	EUR 12.07

Table 3: Summary statistics of both treatments

Rating Behavior	Identity Treatment (n=96)	Anonymity Treatment (n=96)	Total Sample
Published Ratings	150 (15.6%)	99 (10.3%)	249 (13.0%)
Averaged Published Ratings per Subject	1.56	1.03	1.30
Averaged Published Satisfaction Level	12.41	12.20	12.33
Averaged Unpublished Satisfaction Level	14.03	13.60	13.80

Table 4: Rating behavior in market game

Effect of self-expression. To test Hypothesis 1, we need to compare rating behavior between treatments. Therefore we look at the behavior of subjects in a comparative analysis. Comparing the total number of published ratings per subject with a Mann-Whitney-U test (MWU), we find that subjects in the anonymity treatment reveal their satisfaction significantly less often than subjects in the identity treatment ($p = 0.0065$, $z = -2.723$). For robustness, we repeat our analysis separately for each customer group and find significant treatment effects in customer groups X ($p = 0.0644$, $z = -1.849$) and Z ($p = 0.0347$, $z = -2.112$), but not in group Y ($p = 0.4767$, $z = -0.712$). Primary objective of the rating system is to provide information on the quality provided by the sellers. Hence, treatment effects can also arise by earlier information disclosure in the anonymity treatment, which would expire this reason to provide further ratings. We examine this potential explanation and find that the treatment effect is not driven by earlier information disclosure in the anonymity treatment compared to the identity treatment, as in the anonymity treatment all 3 optimal sellers are rated by the relevant customer groups on average in period 5 compared to period 3.83 in the identity

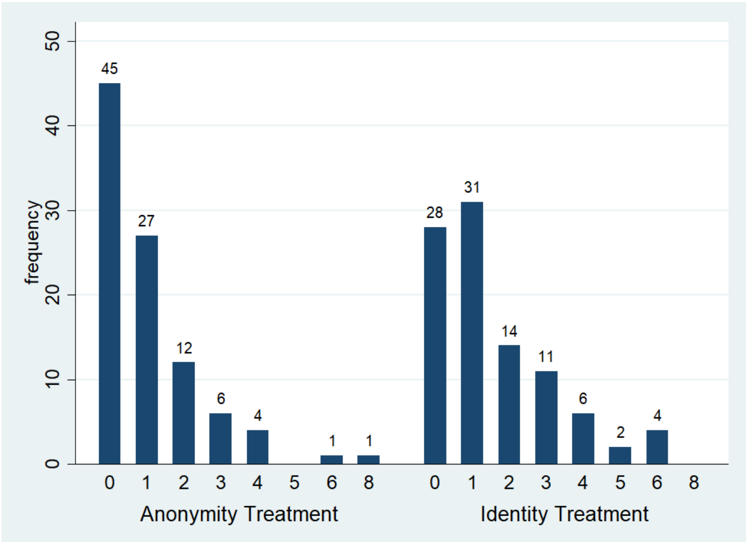

Figure 3: Histogram of ratings per subject in anonymity treatment and identity treatment

treatment. Further, in the anonymity treatment on average 13.25 out of 21 satisfaction levels (cf. Table 1) were published compared to 16.25 satisfaction levels in the identity treatment. Also luckier subjects in the anonymity treatment who pick by chance the best seller and do not see any benefit of sharing information and receiving information in return (i.e., reciprocity) might explain the treatment effect. We test for this and refute that idiosyncratic satisfaction levels in the early periods might have driven the different rating behavior, since we do not find different levels of satisfaction between treatments until period 6 (MWU: $p = 0.0660$, $z = -1.839$). Since only ratings of group X for seller A could reach the maximum satisfaction value of 20 points, leading to a certain best seller for this group, we exclude in a further robustness test ratings of seller A (13 in anonymity and 23 in identity treatment) and still find a significant treatment effect (MWU: $p = 0.0171$, $z = -2.385$).

Analyzing the rating behavior over time, we find a negative trend in rating behavior in both treatments. We test the difference in rating behavior in periods 1-5 and 6-10 and find, independent of the treatment, a strongly significant decrease over time (χ^2 test in anonymity treatment: $p < 0.001$, $z = 20.8243$; χ^2 test in identity treatment: $p = 0.004$, $z = 8.0909$). Comparing the accumulated ratings between treatments in each period, the significant difference occurs for the first time in Period 5 (MWU: $p = 0.0795$, $z = -1.754$) and increases in each subsequent period (cf. Figure 4). The difference in treatments thus does not depend on different rating behavior over time between treatments.

Therefore, Result 1 follows directly.

Result 1. *As compared to the identity treatment, subjects reveal their satisfaction with sellers significantly less often in the anonymity treatment.*

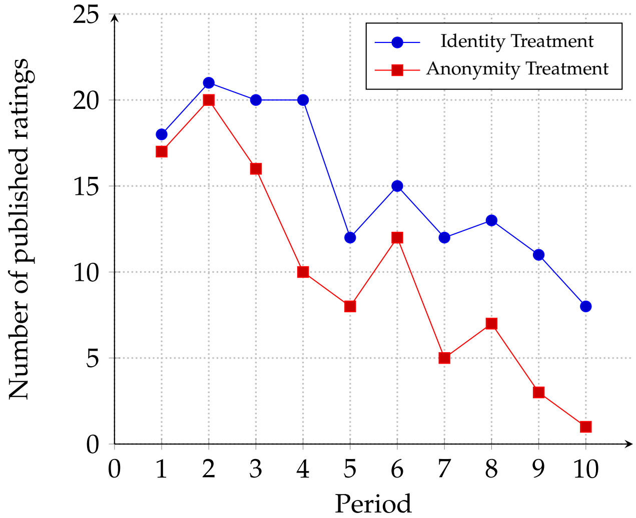

Figure 4: Number of ratings over time by treatment

Effect of Altruism. Hypothesis 2 is concerned with the effect of anonymity on altruistic subjects. To analyze whether altruism indeed plays a role in providing reviews and how it interacts with anonymity, we included the dictator game in our experiment as a measure of altruism (Engel, 2011). As we do not focus on the separation of warm-glow and altruism, as pointed out in Andreoni et al. (2008), the dictator game is the appropriate choice for testing the effect of altruism on the reviewing behavior in an incentivized set-up. We classify the participants into two groups according to their behavior. *Altruists* (n=91) are defined by giving more ECUs to their matched partners in the dictator game than the median amount (6 ECUs, cf. Table 5), whereas *non-altruists* (n=101) give the median amount or less.

Dictator Game	**Identity Treatment (n=96)**	**Anonymity Treatment (n=96)**	**Total Sample**
Mean ECU Given	5.38	5.90	5.64
Median ECU Given	6	7	6
Mean ECU Given by Altruists	9.63	9.54	9.58
Mean ECU Given by Non-Altruists	1.92	2.25	2.08
Altruists' Averaged Published Ratings	1.63 (n=43)	1.42 (n=48)	1.52 (n=91)
Non-Altruists' Averaged Published Ratings	1.51 (n=53)	0.65 (n=48)	1.10 (n=101)

Table 5: Behavior in dictator game and ratings of altruists and non-altruists

We do not observe any differences between treatments with regard to the amount given in the dictator game (MWU: $p = 0.2696$, $z = 1.104$).[9] Thus, with respect to altruism our random allocation of subjects to treatments worked. Testing for differences regarding the published ratings per subject (cf. Table 5), we find that, overall, altruists publish more ratings (MWU: $p = 0.0136$, $z = -2.467$). Conducting this analysis separately for each treatment, we find that this result holds under anonymity (MWU: $p = 0.0129$, $z = -2.487$). In the identity treatment, where self-concerning motives such as self-expression also play a role this effect disappears (MWU: $p = 0.2831$, $z = -1.073$). The publishing behavior for altruists and non-altruists is depicted in Figure 5.

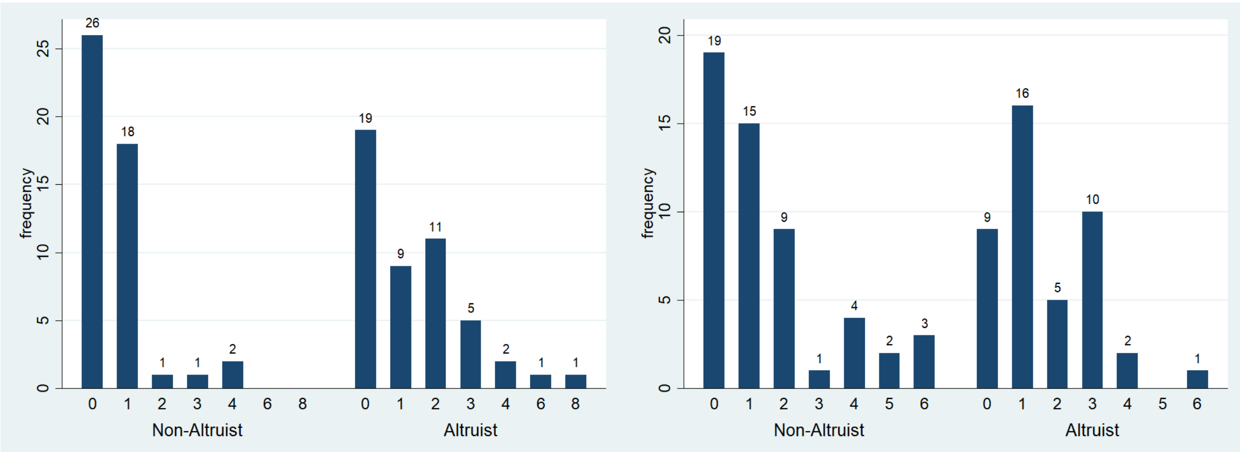

Figure 5: Histogram of ratings per altruistic and non-altruistic subjects in anonymity treatment (left) and identity treatment (right)

Testing whether the number of ratings published by altruistic subjects is independent of the treatment (Hypothesis 2), we compare altruists' averaged published ratings between treatments. As we do not find a significant difference (MWU: $p = 0.2234$, $z = -1.217$), Result 2 follows directly.

Result 2. *Altruists' publishing behavior is not affected by anonymity.*

This implies that altruistic subjects provide ratings to give information to others, independent of whether they can be identified by their pseudonyms, while in the identity treatment, non-altruists averaged published ratings significantly increase (MWU: $p = 0.0066$, $z = -2.718$). The treatment effects are thus driven by non-altruists, emphasizing the importance of different sources of motivation when investigating the effects of design elements on rating behavior.

Effect of information on market outcome. The utility of published ratings for other subjects depends on the previously published ratings. The first rating for each seller and customer group (cf. Table 1) has the

[9] However, we observe a gender effect with regard to the amount given in the dictator game, as women give significantly more (MWU: $p = 0.0496$, $z = 1.964$), which is in line with the general literature on altruism (e.g., Brañas-Garza et al. 2018).

highest informative value. Subsequent ratings provide less information although reducing the uncertainty induced by the quality variation term ϵ_i. Hence, we define the first ratings for each seller and customer group as *fully informative ratings*, meaning that 21 fully informative ratings are possible in each session or 84 fully informative ratings in each treatment, respectively.[10] The other published ratings are classified as *less-informative ratings*. In the sessions of the anonymity treatment, there are overall 57 fully informative and 42 less-informative ratings, while in the sessions of the identity treatment, overall 76 fully informative and 74 less-informative ratings are provided. As we observe overall more ratings in the identity treatment, it is reasonable that we find both more fully informative ratings ($p = 0.088,\ z = 2.9163$), and more less-informative ratings ($p = 0.001,\ z = 10.1833$). Comparing the classified ratings between treatments using a χ^2-test, we do not find a significant difference ($p = 0.285,\ z = 1.1441$). This means that the share of fully informative ratings does not explain the differences in ratings between the treatments. We also test within subjects whether rather fully informative or less-informative ratings are published. Because the Wilcoxon signed-rank test does neither reject equality in the identity treatment ($p = 0.9020,\ z = 0.123$) nor in the anonymity treatment ($p = 0.1332,\ z = 1.502$), on average the informative value of a rating is similar in both treatments. This means that the higher number of ratings also corresponds to more information in the identity treatment, in comparison with the anonymity treatment.

Comparing the published and unpublished satisfaction levels provides a measure of the *representativeness* of the ratings (cf. Table 4). The ratings exhibit a negative bias in the representativeness, as the published satisfaction levels are significantly lower in comparison with the unpublished satisfaction levels ($z = 4.959;\ p < 0.0001$), meaning that the negative experiences are more often published. In particular, the representativeness of the ratings is similar in both treatments as the deviations between published and unpublished satisfaction levels are similar in the anonymity treatment ($z = 3.233;\ p = 0.0012$), and the identity treatment ($z = 3.958;\ p = 0.0001$). Controlling for time effects, we separately investigate the published and unpublished satisfaction levels for each period and find significant differences only for the periods 6, 8, and 9.[11]The bias in the representativeness might thus be driven by the interaction of saturation of information and the higher satisfaction levels in the later periods. Testing for differences between the accumulated published and unpublished satisfaction levels after each period provides support for this reasoning as significant differences occur the first time for observations from periods 1 to 6 ($z = 2.041;\ p = 0.0413$). However, we can conclude that neither the informative value nor the representativeness of the ratings are affected by the treatment variations. We thus focus on the quantity of the ratings and its impact on market outcomes in the subsequent analysis.

[10] When different subjects of the same customer group rate the same seller in the same period, all of these ratings are classified as fully informative, which means that more than 84 fully informative ratings are theoretically possible.

[11] The test statistics of the MWU test are as follows: periods 6 ($z = 1.955;\ p = 0.0506$); period 8 ($z = 2.167;\ p = 0.0302$); period 9 ($z = 1.829;\ p = 0.0674$).

We analyze the effect of the anonymous reputation system on the satisfaction of customers in our artificial market. There is empirical evidence that market outcomes such as sales of vendors (Chevalier and Mayzlin 2006) and reservation prices of customers (Resnick et al. 2006) increase with more (positive) information. As mentioned above, more ratings (i.e., more information) are published in the identity treatment, meaning that we expect higher satisfaction levels of subjects in the identity treatment in comparison with the anonymity treatment. Using the MWU test, we find evidence that the level of satisfaction is indeed higher in the identity treatment in comparison with the anonymity treatment ($z = -2.372;\ p = 0.0177$).

Analyzing the overall effect of our anonymous rating system, we also consider the costs to publish ratings. We thus analyze whether the lower costs for publishing fewer ratings overcompensate the lower satisfaction of subjects in the anonymity treatment, on average. Comparing the monetary net benefit of the subjects (i.e., satisfaction minus costs for publishing) between both treatments, we find that the payoffs of the subjects are weakly significantly higher in the identity treatment (MWU: $z = -1.787;\ p = 0.0739$). This means that also under consideration of costs, the implementation of pseudonyms in our anonymous reputation system corresponds to higher market outcomes (i.e., higher net-benefits of the subjects) at least in our experiment.

Result 3. *Implementing pseudonyms in an artificial anonymous reputation system induces more information, which corresponds to higher market outcomes.*

5. Conclusion

Reputation systems are a vital part of online commerce. Without well-functioning reputation systems, online markets will fail, as the usual trust-building devices that work on regular market places are not available in an online setting. At the same time, however, a lack of privacy is becoming a steadily increasing concern in online environments in general as well as in reputation systems in particular. Therefore, given the need for more privacy in online markets, computer scientists have shown the technical feasibility and effectiveness of anonymous review systems. While this may not have any impact on reviews published for purely altruistic reasons, another motive that has been identified for writing reviews is self-expression. As self-expressing motives would not be effective in anonymous systems, this might lead to fewer ratings and, therefore, to a lower information value of such reputation systems. We have designed a laboratory experiment to test under ceteris paribus conditions whether anonymity indeed inhibits the potential of self-expression and, therefore, negatively affects the propensity of customers to leave ratings. Additionally, we have tested whether the impact is the same for altruistic and non-altruistic subjects and the overall effect on the market outcomes.

Varying the degree of anonymity from pseudonymity in the *identity treatment* to anonymity in the *anonymity treatment*, we show that under anonymity significantly fewer ratings are published. This decrease in ratings leads to lower market outcomes. Looking at the rating behavior in general, we, thereby, find that overall altruistic subjects publish more ratings compared to non-altruists. In particular, we find that altruists publish ratings on a constant level independent of the treatment. In contrast, non-altruists are strongly affected by the introduction of anonymity in reputation systems as they publish significantly less ratings. I.e., when designing a reputation system, altruism and self-expression must be considered as powerful, and to some degree complementary, motives of reviewers. As there are different groups of reviewers regarding the underlying motives, a platform should analyze in advance which customer groups are intended to be stimulated by its reputation systems to publish customer reviews.

We also find the tendency of ratings to decrease over time. This was to be expected as information value of ratings in our setting reduces over time. Once a seller has been identified as a good seller for a particular customer group, there is no benefit to rate him anymore. Although we do not face a problem in this experiment, given the constant quality levels of the sellers, concerns about this issue may arise in other settings when quality shifts might take place. In particular, in anonymous reputation systems with a lower level of customer reviews this risk must not be neglected.

In this experiment we focused on the effect of anonymity on the number of ratings, as this is the first and most basic question that needs to be answered when considering anonymous reputation systems. However, while we have thus shown downsides of anonymous reputation systems, we explicitly have not investigated potential upsides of allowing customers to review without revealing personal information. While we have shown that in our setting anonymity led to fewer ratings, we have not identified whether rating behavior itself is impacted. For example, the truthfulness of reviews might be affected by anonymity as well. Assuming that altruists publish more truthful ratings under anonymity, our results, then, would suggest that the implementation of anonymous review systems could lead, indeed, to more efficient reputation systems, as the number of untruthful reviews drops.

Additionally, in other contexts, such as for example B2B markets, anonymity might actually increase the number of ratings, as companies no longer divulge private information on their suppliers. Also in contexts such as eBay or Uber where rating is reciprocal anonymity might actually increase the number of ratings or the truthfulness of ratings, as the fear of retaliatory ratings may decrease (Resnick and Zeckhauser, 2002). Therefore, further research is needed to answer the question whether the overall impact of anonymity on reputation systems is positive or negative taking into account the specifics determinants and needs of the market in question.

Taking into account the importance of reputation systems for e-commerce platforms, designing a good reputation system is of utmost importance to avoid market failure due to information asymmetries. Given the trillion dollar value of worldwide e-commerce simply making reputation systems anonymous might thus lead to disastrous consequences. As privacy concerns remain, however, the simple statement that anonymity has a negative impact on rating systems is non-sufficient. Therefore, further research should investigate how anonymous rating systems can be modified efficiently to increase participation. For example the design of anonymous reputation systems could include instruments of gamification (such as obtaining badges for a certain amount of reviews). Bemmann et al. (2018) show that it is possible to limit rating a product to those customers who actually bought the product, while keeping the system anonymous. Adding gamification elements in such an environment should also work in this context. However, further tests are needed to analyze whether gamification is the correct measure to ensure a sufficient number of ratings in a system or whether other modifications are better approaches.

Acknowledgements

We would like to thank participants at the GfeW 2018 for their helpful comments. The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

Akerlof, G. A. (1970). The market for" lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Andreoni, J., Harbaugh, W. T., and Vesterlund, L. (2008). Altruism in experiments. *The New Palgrave Dictionary of Economics: Volume 1–8*, pages 134–138.

Avery, C., Resnick, P., and Zeckhauser, R. (1999). The market for evaluations. *American economic review*, 89(3):564–584.

Bardsley, N. and Moffatt, P. G. (2007). The experimetrics of public goods: Inferring motivations from contributions. *Theory and Decision*, 62(2):161–193.

Bemmann, K., Blömer, J., Bobolz, J., Bröcher, H., Diemert, D., Eidens, F., Eilers, L., Haltermann, J., Juhnke, J., Otour, B., Porzenheim, L., Pukrop, S., Schilling, E., Schlichtig, M., and Stienemeier, M. (2018). Fully-featured anonymous credentials with reputation system. In Doerr, S., Fischer, M., Schrittwieser, S., and Herrmann, D., editors, *Proceedings of the 13th International Conference on Availability, Reliability and Security, ARES 2018, Hamburg, Germany, August 27-30, 2018*, pages 42:1–42:10. ACM.

Blömer, J., Juhnke, J., and Kolb, C. (2015). Anonymous and publicly linkable reputation systems. In *International Conference on Financial Cryptography and Data Security*, Berlin. Springer.

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

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

Brañas-Garza, P., Capraro, V., and Rascon-Ramirez, E. (2018). Gender differences in altruism on mechanical turk: Expectations and actual behaviour. *Economics Letters*, 170:19–23.

Cheung, C. M. and Lee, M. K. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision Support Systems*, 53(1):218–225.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354.

Clement, J. (2020). E-commerce worldwide - statistics & facts. *https://www.statista.com/topics/871/online-shopping/*.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10):1407–1424.

Eisenberg, N. (1996). Caught in a narrow kantian perception of prosocial development: Reactions to campbell and christopher's critique of moral development theory. *Developmental Review*, 16(1):48 – 68.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

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

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

Halliday, S. D. and Lafky, J. (2019). Reciprocity through ratings: An experimental study of bias in evaluations. *Journal of Behavioral and Experimental Economics*, 83:101480.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, 18(1):38–52.

Lafky, J. (2014). Why do people rate? Theory and evidence on online ratings. *Games and Economic Behavior*, 87:554–570.

Munzel, A. and Kunz, W. H. (2014). Creators, multipliers, and lurkers: who contributes and who benefits at online review sites. *Journal of Service Management*, 25(1):49–74.

Ni, J. (2019). Amazon review data 2018. *https://nijianmo.github.io/amazon/index.html*.

Resnick, P. and Zeckhauser, R. (2002). Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *The Economics of the Internet and E-commerce*, 11(2):23–25.

Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental economics*, 9(2):79–101.

Rockenbach, B. and Sadrieh, A. (2012). Sharing information. *Journal of Economic Behavior & Organization*, 81(2):689–698.

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

Voss, M. (2004). Privacy preserving online reputation systems. In Deswarte, Y., Cuppens, F., Jajodia, S., and Wang, L., editors, *Information Security Management, Education and Privacy*, pages 249–264, Boston, MA. Springer US.

Wu, P. F. (2019). Motivation crowding in online product reviewing: A qualitative study of amazon reviewers. *Information & Management*.

Appendix A. Instructions

These instructions are English translations from the originally German version.[12]

Appendix A.1. Treatment Anonymity

General Information

- There is no communication allowed.
- All cell phones must be turned off during the entire experiment.
- All decisions you make in the context of this experiment are made anonymously, i.e., none of the other participants knows the identity of the one who has made a specific decision.
- Please stay seated until the end of the experiment.
- Please wait until you are called for getting your payoff.

The Experiment

- During the experiment, all amounts are given in the experimental currency unit "Taler".
- The experiment consists of two parts. The decisions in both parts are independent.
- At the end of the experiment, we ask you to fill out a questionnaire. Your answers in this questionnaire will not affect your payout in this experiment.

Description of Part 1

- In the first part, all participants assume the role of a customer on a marketplace. The sellers are not participants of the experiment, instead they are generated by the computer.
- As a participant, you buy a product that is offered by different sellers that differ with regard to the specific implementation of the product.
- For the duration of the first part, you will be assigned an ID that you use to act on the market.
- You will also be randomly assigned to one of three different groups of customers for the duration of the first part: customer group X, customer group Y, or customer group Z. Your group will be displayed at the top of the screen for the duration of the first part of the experiment.

[12] Additional to the instructions screenshots from the decision screen were provided for clarification (cf. Figure A.6).

- The first part of the experiment consists of 10 periods.
- In each period there are two choices to make:
  1. Choosing a seller
  2. Publication of realized satisfaction

1. **Choosing a seller**

- In each period, you have the choice between 7 sellers whose products satisfy you to varying degrees.
- Each seller offers exactly one product.
- For each seller, a constant, average expected satisfaction with the product is given for the duration of the experiment. This is calculated as follows:

$$\textit{Average Satisfaction}\varnothing = \frac{1}{3} * \left\{ \begin{array}{r} \textit{Average satisfaction of customer group X} \\ + \textit{Average satisfaction of customer group Y} \\ + \textit{Average satisfaction of customer group Z} \end{array} \right\}.$$

- The realized satisfaction with the product of a seller is between 2 and 20 points.
- The realized satisfaction with a seller's product depends on the customer group.
- In addition, the realized satisfaction in each period randomly varies by up to 2 points according to the following scheme:

Probability	**Variation**
20%	-2 points
20%	-1 point
20%	0 points
20%	+1 point
20%	+2 points

- The maximum variation of 2 points is the same for each seller.
- Your payoff in "Taler" per period is equal to the realized satisfaction you received with the purchased product. There are no costs for the purchase.

Example for choosing the seller

Seller J sells a product that results in an average expected satisfaction of 8 points for customer group X, an average expected satisfaction of 4 points for customer group Y, and an average expected satisfaction of 6 points for customer group Z. The average expected satisfaction shown in the experiment for this product is therefore $\varnothing = (\frac{8+4+6}{3}) = 6$ *points.*

If you choose this seller, this will result in the following realized satisfaction, depending on your known customer group and the random, unknown variation. Therefore, the realized satisfaction in this example can be any of the outcomes in the following table.

Probability	Variation	Realized satisfaction of customer group X	Realized satisfaction of customer group Y	Realized satisfaction of customer group Z
20%	-2 points	6 points	2 points	4 points
20%	-1 point	7 points	3 points	5 points
20%	0 points	8 points	4 points	6 points
20%	+1 point	9 points	5 points	7 points
20%	+2 points	10 points	6 points	8 points

If you belong to customer group X, your realized satisfaction in this example is between 6 and 10 points. If you belong to customer group Y, your realized satisfaction is between 2 and 6 points. If you belong to customer group Z, in this example, your realized satisfaction is between 4 and 8 points. Your payoff corresponds to your realized satisfaction, with the exchange rate: 1 point = 1 Taler.

2. **Publication of realized satisfaction**

- After you have experienced your realized satisfaction in the period, you have the opportunity to publish your realized satisfaction with the seller. Thereby, the satisfaction you have experienced with the seller will anonymously be published. When doing so, it will be published that a customer from your customer group has received your realized satisfaction from the seller you have chosen. Every customer has this opportunity.
- If you share information about your satisfaction, you will incur costs of 2 Taler. Your payoff for this period will be reduced by 2 Taler.
- The shared information is visible to every customer in all subsequent periods.

-> In summary, you buy a product in each period and afterwards decide whether to publish your realized satisfaction with the chosen product.

-> In each period your payoff is equal to your realized satisfaction minus costs of 2 Taler, if you have chosen to publish your satisfaction.

Payoff per Period = Realized Satisfaction - Costs for Publication

Description of Part 2

- The second part of the experiment is independent of the first part and you will receive a budget of 20 Taler for this part.
- You are anonymously assigned to a partner.
- Your task is to divide a budget of 20 Taler between yourself and your partner. Your partner has no influence on your decision and must accept your allocation.
- At the same time, you are anonymously assigned to another participant who decides how to divide 20 Taler between himself and you.
- Afterwards, you will see your allocation and the allocation of the participant you have been allocated to. Here, a decision is randomly marked as "A" and the other decision as "B".
- Then, a participant rolls for everyone a six-sided dice, thereby deciding, which of the two decisions will be paid out. If an even number (2, 4 or 6) is rolled, decision "A" will be implemented. If an uneven number (1, 3 or 5) is rolled, decision "B" will be implemented.
- At the end, you will either receive the payoff for which you have divided the 20 Taler, or you receive the payoff in accordance with the decision of the participant, who had to divide 20 Taler between himself and you.

Payoff

- Your earnings resulting from your realized satisfaction from the 10 periods of the first part and your earnings from the second part of the experiment are added up.
- These earnings will be paid to you at the end of the experiment at the **exchange rate: 15 Taler = € 1.**
- In addition, you will receive a **show-up fee of € 2.50** at the end of the experiment.

Good luck and thank you for participating in this experiment!

Appendix A.2. Treatment Identity

General Information

- There is no communication allowed.
- All cell phones must be turned off during the entire experiment.
- All decisions you make in the context of this experiment are made anonymously, i.e., none of the other participants knows the identity of the one who has made a specific decision.
- Please stay seated until the end of the experiment.
- Please wait until you are called for getting your payoff.

The Experiment

- During the experiment, all amounts are given in the experimental currency unit "Taler".
- The experiment consists of two parts. The decisions in both parts are independent.
- At the end of the experiment, we ask you to fill out a questionnaire. Your answers in this questionnaire will not affect your payout in this experiment.

Description of Part 1

- In the first part, all participants assume the role of a customer on a marketplace. The sellers are not participants of the experiment, instead they are generated by the computer.
- As a participant, you buy a product that is offered by different sellers that differ with regard to the specific implementation of the product.
- First, you are asked to create a pseudonym (nickname) for the duration of the first part, which you use to act on the market.
- You will also be randomly assigned to one of three different groups of customers for the duration of the first part: customer group X, customer group Y, or customer group Z. Your group will be displayed at the top of the screen for the duration of the first part of the experiment.
- The first part of the experiment consists of 10 periods.
- In each period there are two choices to make:
    1. Choosing a seller

2. Publication of realized satisfaction

1. **Choosing a seller**

- In each period, you have the choice between 7 sellers whose products satisfy you to varying degrees.
- Each seller offers exactly one product.
- For each seller, a constant, average expected satisfaction with the product is given for the duration of the experiment. This is calculated as follows:

$$\textit{Average Satisfaction}\varnothing = \frac{1}{3} * \left\{ \begin{array}{r} \textit{Average satisfaction of customer group X} \\ + \textit{Average satisfaction of customer group Y} \\ + \textit{Average satisfaction of customer group Z} \end{array} \right\}.$$

- The realized satisfaction with the product of a seller is between 2 and 20 points.
- The realized satisfaction with a seller's product depends on the customer group.
- In addition, the realized satisfaction in each period randomly varies by up to 2 points according to the following scheme:

Probability	**Variation**
20%	-2 points
20%	-1 point
20%	0 points
20%	+1 point
20%	+2 points

- The maximum variation of 2 points is the same for each seller.
- Your payoff in "Taler" per period is equal to the realized satisfaction you received with the purchased product. There are no costs for the purchase.

Example for choosing the seller

Seller J sells a product that results in an average expected satisfaction of 8 points for customer group X, an average expected satisfaction of 4 points for customer group Y, and an average expected satisfaction of 6 points for customer group Z. The average expected satisfaction shown in the experiment for this product is therefore $\varnothing = (\frac{8+4+6}{3}) = 6$ *points.*

If you choose this seller, this will result in the following realized satisfaction, depending on your known customer group and the random, unknown variation. Therefore, the realized satisfaction in this example can be any of the outcomes in the following table.

Probability	Variation	Realized satisfaction of customer group X	Realized satisfaction of customer group Y	Realized satisfaction of customer group Z
20%	-2 points	6 points	2 points	4 points
20%	-1 point	7 points	3 points	5 points
20%	0 points	8 points	4 points	6 points
20%	+1 point	9 points	5 points	7 points
20%	+2 points	10 points	6 points	8 points

If you belong to customer group X, your realized satisfaction in this example is between 6 and 10 points. If you belong to customer group Y, your realized satisfaction is between 2 and 6 points. If you belong to customer group Z, in this example, your realized satisfaction is between 4 and 8 points. Your payoff corresponds to your realized satisfaction, with the exchange rate: 1 point = 1 Taler.

2. **Publication of realized satisfaction**

- After you have experienced your realized satisfaction in the period, you have the opportunity to publish your realized satisfaction with the seller. Using your pseudonym, it will be published which seller you chose, which customer group you belong to and the level of satisfaction you have realized. Every customer has this opportunity.
- If you share information about your satisfaction, you will incur costs of 2 Taler. Your payoff for this period will be reduced by 2 Taler.
- The shared information is visible to every customer in all subsequent periods.

-> In summary, you buy a product in each period and afterwards decide whether to publish your realized satisfaction with the chosen product.

-> In each period your payoff is equal to your realized satisfaction minus costs of 2 Taler, if you have chosen to publish your satisfaction.

Payoff per Period = Realized Satisfaction - Costs for Publication

Description of Part 2

- The second part of the experiment is independent of the first part and you will receive a budget of 20 Taler for this part.
- You are anonymously assigned to a partner.
- Your task is to divide a budget of 20 Taler between yourself and your partner. Your partner has no influence on your decision and must accept your allocation.
- At the same time, you are anonymously assigned to another participant who decides how to divide 20 Taler between himself and you.
- Afterwards, you will see your allocation and the allocation of the participant you have been allocated to. Here, a decision is randomly marked as "A" and the other decision as "B".
- Then, a participant rolls for everyone a six-sided dice, thereby deciding, which of the two decisions will be paid out. If an even number (2, 4 or 6) is rolled, decision "A" will be implemented. If an uneven number (1, 3 or 5) is rolled, decision "B" will be implemented.
- At the end, you will either receive the payoff for which you have divided the 20 Taler, or you receive the payoff in accordance with the decision of the participant, who had to divide 20 Taler between himself and you.

Payoff

- Your earnings resulting from your realized satisfaction from the 10 periods of the first part and your earnings from the second part of the experiment are added up.
- These earnings will be paid to you at the end of the experiment at the **exchange rate: 15 Taler = € 1.**

- In addition, you will receive a **show-up fee of € 2.50** at the end of the experiment.

Good luck and thank you for participating in this experiment!

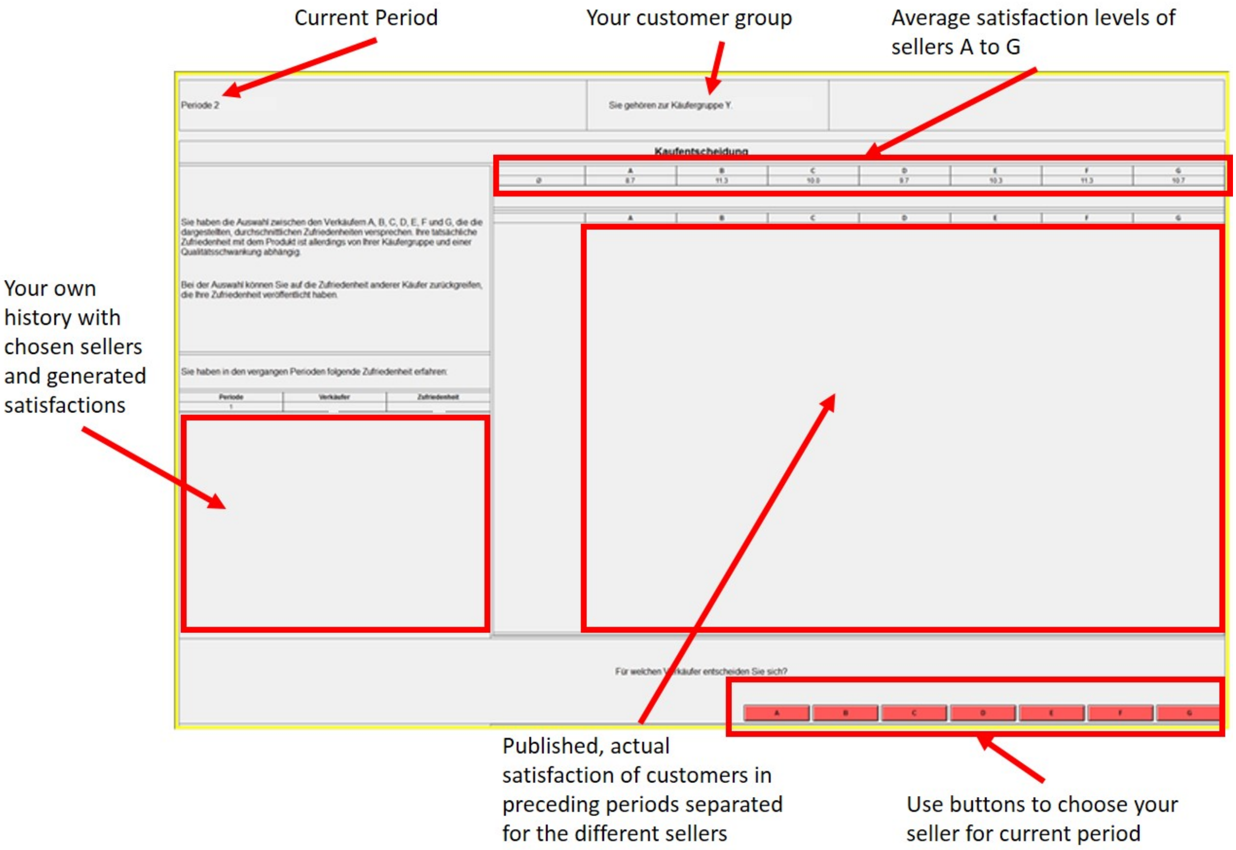

Figure A.6: Explaining screenshot handed out additionally with instructions