Incentive Schemes in Customer Rating Systems - Comparing the Effects of Unconditional and Conditional Rebates on Intrinsic Motivation

D. van Straaten[a,]

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

Abstract

The success of e-commerce is strongly affected by customer review systems, since they are the most important trust-building device in online shopping. However, participation in these systems is biased as certain customer groups do not provide insights from their experience with the purchased goods. This study aims to stimulate the participation in customer review systems by implementing financial incentives. By conducting an economic laboratory experiment, subjects are treated with different types of incentives, namely unconditional rebates (i.e., gifts) or conditional rebates (i.e., rebates in return) for providing customer reviews. Through this process, providing gifts triggers reciprocity and rebates in return rather address economic motives for providing reviews. Finding no evidence for peer reciprocity, we identify seller reciprocity, economic incentives, and altruism as the drivers for providing customer reviews. In particular, we find that conditional rebates strongly increase the quantity of customer ratings but crowd out ratings that are motivated by altruism. In comparison, unconditional rebates result in a lower increase of customer ratings but show the advantage of not crowding out intrinsically motivated customer ratings.

Key Words: Customer Rating, Altruism, Reciprocity, Incentives, Crowding out

JEL Classification: C91, D82, L81, L86

Email address: dirk.van.straaten@uni-paderborn.de (D. van Straaten)

1. Introduction

Customer review systems are one of the most important trust-building devices in online purchasing. Many studies have shown the strong effects of customer reviews on sales and other economic outcomes in online market places (e.g., Chevalier and Mayzlin 2006; Floyd et al. 2014). However, research shows that not every customer group writes reviews, leading to biases in product evaluation. Hu et al. (2009) identifies the under-reporting bias, that is, only very (un)satisfied customers post a review, whereby average satisfied customers do not, thus making middle categories under-reported. It seems that intrinsic motivation of customers with moderate experiences is not sufficient to exceed the costs (i.e., the time and effort) for writing reviews. Hence, one approach is to use extrinsic motivation factors to elicit customer reviews.

The literature has identified various motives for publishing customer reviews, such as altruism, peer reciprocity, seller reciprocity, economic incentives, reputation, self-expression, and social affiliation (Hennig-Thurau et al. 2004; Cheung and Lee 2012; Munzel and H. Kunz 2014; Wu 2019; Peddibhotla and Subramani 2007). As platforms and sellers of high quality are interested in customers providing reviews, incentives are used that address the above-mentioned economic incentives. The literature shows the positive effect of conditional rebates, i.e., economic incentives, on the quantity of customer reviews (Burtch et al. 2017; Chen et al. 2017). However, it also identifies drawbacks such as the decreasing quality of reviews or a lower grade of participation of former top reviewers after the implementation of conditional rebates for reviews (Khern-am nuai et al. 2018; Chen et al. 2010; Sun et al. 2017).

Addressing these drawbacks, this study investigates how monetary incentives should be implemented in customer review systems to maximize provided information for future customers. In particular, the implementation of unconditional rebates (i.e., gifts) vs. conditional rebate mechanisms (i.e., rebate in return) are investigated. The underlying psychological processes behind these mechanisms differ fundamentally: While conditional rebates aim for economic motives, gifts such as unconditional rebates can be used to initialize reciprocal behavior. Our results indicate that conditional economic incentives elicit the highest quantity of ratings. However, altruism is crowded out as a motivational factor when conditional rebates are implemented. In comparison, unconditional rebates lead to a smaller raise in customer ratings. However, unconditional rebates do not result in a decrease of customer ratings that are driven by altruism. This result implies that although conditional rebates work pretty well to raise the quantity of customer ratings, unconditional rebates can be superior to conditional rebates when the goal of an incentive program in reputation systems is to collect a more diverse picture of opinions by reducing the amount of intrinsically motivated customer reviews that are crowded out. The paper is organized as follows: Chapter 2 presents a survey of the literature on motives for writing customer reviews and the derived research hypotheses. Chapter 3 provides an explanation of the experimental design. The results of the hypotheses tests are given in chapter 4. Chapter 5 discusses the results and concludes the article.

2. Motives for Writing Customer Reviews

Several studies investigate the motives for providing customer reviews. The enjoyment of helping others (altruism) is identified in most articles (cf. Peddibhotla and Subramani 2007; Hennig-Thurau et al. 2004; Munzel and H. Kunz 2014; Cheung and Lee 2012; Wu 2019). Reciprocity towards peers (cf. Peddibhotla and Subramani 2007; Hennig-Thurau et al. 2004; Munzel and H. Kunz 2014) and sellers (cf. Munzel and H. Kunz 2014; Wu 2019) is also observed. Economic incentives (cf. Peddibhotla and Subramani 2007; Hennig-Thurau et al. 2004; Wu 2019) are a further affective motive for providing reviews. Furthermore, self-expression, building up a reputation, and social affiliation are found to drive reviewing behavior (cf. Peddibhotla and Subramani 2007; Hennig-Thurau et al. 2004; Munzel and H. Kunz 2014; Cheung and Lee 2012; Wu 2019). As our interest is mainly on the effect of monetary incentives on reviewing behavior, we therefore focus on altruism, peer reciprocity, seller reciprocity, and economic incentives. We also address further motives in brief and discuss the dependencies between these motives.

2.1. Altruism

Eisenberg (1996) defines altruism as a "voluntary behavior that is intended to benefit another and is not motivated by the expectation of external reward". Thereby, altruism is a motivational concept that might produce pro-social behavior, but does not have to (Batson and Powell 2003). By conducting an experiment, Andreoni and Miller (1993) find support that pure altruism exists, although it is difficult to discriminate between altruism and warm-glow (Andreoni et al. 2008). A further experiment, run by Cox (2004), finds in an experiment clean and separable evidence for trust, reciprocity, and altruism.[1] Later, Engel (2011) performed a meta analysis on dictator games, which shows that if there is more than one recipient, the share given to the receiver increases on average by over 10%. Investigating altruism as a motive for customer reviews, Reimer and Benkenstein (2016) held a scenario experiment and found that highlighting the opportunity to help others or to choose the destination of a donation increases the intention of customers to write a customer review. Peddibhotla and Subramani (2007) analyzed profiles of reviewers on amazon.com and classified them regarding their described motives for providing reviews. Using this classification, they show that the helpfulness of reviews (as a proxy for quality) published by altruistic subjects is weakly significant higher in comparison with reviewers who do not mention altruistic reasons. However, they do not find an effect of altruism on the quantity of reviews. Studies employing questionnaires show in different domains that altruism is a robust motive for providing customer ratings (Hennig-Thurau et al. 2004; Munzel and H. Kunz 2014; Cheung and Lee 2012). Conducting interviews with top-reviewers from amazon.co.uk, Wu (2019) identified the focus of intrinsic motivation to be rather the perceived enjoyment

[1] Although both reciprocity and altruism can lead to pro-social behavior, Suls et al. (1981) identify that pro-social behavior resulting from an altruistic behavior is valued more compared to positive reciprocity.

than pure altruism. Bearing this in mind, we, however, phrase this motive altruism and hypothesize the following:

Research Hypothesis 1. *In reputation systems without additional instruments for eliciting customer ratings there is a significant number of customer ratings.*

In situations with the option to share the budget with a recipient, the amounts given vary over a wide range (cf. Engel 2011). Thereby, Andreoni and Miller (2002) show that this can be explained by heterogeneous attitudes towards altruism. We hypothesize the following:

Research Hypothesis 2. *In reputation systems without additional instruments for eliciting customer ratings, altruists publish significantly more customer ratings compared to non-altruists.*

2.2. Reciprocity

We follow the definition of reciprocity by Fehr and Gächter (2000, p.159): "Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently much more nasty and even brutal. (...) Reciprocity is an in-kind response to beneficial or harmful acts." Conducting a gift exchange game in a market setting, Fehr et al. (1993) observed that prices offered by buyers were higher than market prices (and can therefore be perceived as gifts), leading to high quality sent by sellers. Falk (2007) found evidence for reciprocity generated by gifts in the field. Sending solicitation letters with or without gifts, donations increased by up to 75% when large gifts (i.e., 4 postcards with envelopes) were included. Illustrating the robustness of reciprocity, Berg et al. (1995) identify reciprocity even in an anonymous setting by conducting an experiment. Kube et al. (2012) show that the reaction on a reciprocal act is not only determined by the price equivalent but also by the type of a gift and its evaluation of the receivers' perception. Reciprocity can be present in many different constellations as there are different market participants such as customers, sellers or the market provider. In the context of customer reviews, prior studies mainly investigated reciprocity initiated by the customer (e.g., Hennig-Thurau et al. 2004: "platform assistance") or reciprocity initiated by the seller not with an extra service or gift, but by providing satisfying products ("helping the company"). In our study, described below in more detail, two types of reciprocity can appear: indirect peer reciprocity and seller reciprocity.

2.2.1. Indirect peer reciprocity

Peer reciprocity of the customer results from a kind act (i.e., a review) of a peer helping the participant to make a more informed decision. Since we prevent direct reciprocal behavior between customers by design, participants can only react indirectly reciprocally by providing other customers with a review as a kind act. In Munzel and H. Kunz (2014), 32% of the participants state that they publish a review as they want to give something back to the community. Peddibhotla and Subramani (2007) show that this motive has a significant positive impact on the quality of the reviews. Hennig-Thurau et al. (2004) also identifies advice seeking as a motive to publish customer reviews.

Simpson et al. (2018) investigate generalized and indirect reciprocity and their impact on pro-social behavior. Greiner and Levati (2005) illustrate that not even a direct connection between sender and receiver has to be present to establish reciprocity. Hence, we hypothesize the following:

Research Hypothesis 3. *In reputation systems without additional instruments for eliciting customer ratings, customers who receive more ratings also publish more ratings.*

2.2.2. Seller reciprocity

Askalidis et al. (2017) propose sending emails to customers and inviting them to submit a review.[2] Evaluating that this approach works in principal, they show that self-motivated reviews are perceived to be more helpful and more negative than retailer-prompted reviews. They also mention that the self-selection of reviewers is mitigated by the retailer-prompted emails as these elicit other reviews from different buyer groups. Munzel and H. Kunz (2014) ask reviewers on TripAdvisor.com about their motives after reviewing and find seller reciprocity to be a motive for 16.7% of the subjects as they react on the hotel's request to provide a review. Gutt et al. (2017) examine the effect of gifts provided with a request to publish a customer review on reviewing behavior, finding a positive effect on reviewing behavior only in the group of older customers. Wu (2019) identify the existence of seller reciprocity as a motive for providing customer reviews. The relation is initiated by the seller, who provides the reviewer with a sample that motivates the reviewer to write a review in return. Additionally, Chen et al. (2017) find that free sampling of software on one platform leads to higher volume of customer reviews on another, which also affects the number of sales. We hypothesize the following:

Research Hypothesis 4. *The implementation of reciprocal instruments increases the quantity of customer ratings in reputation systems.*

2.3. Economic incentives

Hennig-Thurau et al. (2004) show that economic incentives increase the quantity of reviews in reputation systems. Supporting these results, Peddibhotla and Subramani (2007) also find that economic incentives do not affect the quality of reviews. Wu (2019) finds further evidence that customer reviews are motivated by economic incentives.

[2] Pointing in a similar direction, Picazo-Vela et al. (2010) observe that the level at which customers perceive they are pressured by sellers to rate is positively correlated with their intention to write a review.

Cabral and Li (2015) analyze the effect of feedback-conditional rebates on the buyers' likelihood to give feedback. They find weak evidence that feedback increases when conditional rebates are provided. When service quality is low, seller rebates help to avert negative feedback.

In a natural experiment, Khern-am nuai et al. (2018) show that introducing financial incentives leads to increasing numbers of reviews. However, the quality of these reviews decreases as the valence increases. Li and Xiao (2014) conduct a lab experiment to investigate rebate mechanisms and their effect on the market outcomes. Sellers have an option to offer rebates for providing feedback and thereby the possibility to signal their buyers to provide high quality, an approach that works as the market efficiency increases. Conducting an experiment on MTurk, Burtch et al. (2017) identified that incentives lead to higher numbers of reviews but that these incentivized reviews are shorter. Neumann and Gutt (2019) identifies that the source of elicitation matters in a B2B setting: Investigating review elicitation, they find that a seller-initiated elicitation leads to an increase in reviews. However, if the reviews are elicited by the market platform there is a decrease in reviews. Thereby, elicitation mechanisms affect review length negatively. Although the evidence on the quality of incentivized reviews is mixed, we expect the quantity of reviews to be affected positively by economic incentives:

Research Hypothesis 5. *The implementation of economic incentives in reputation systems increases the quantity of customer ratings.*

To the best of our knowledge, there is no study present that quantitatively compares the effect of both instruments. However, induced value theory (cf. Smith 1976) is a well-established concept and central pillar of experimental economics. Hence, we expect that the economic incentives will have a stronger impact on reviewing behavior in comparison with implemented reciprocity:

Research Hypothesis 6. *The implementation of economic incentives in reputation systems has a stronger impact on the quantity of customer ratings in comparison with reciprocal instruments.*

2.4. Crowding-out and crowding-in motives

Writing customer reviews can be interpreted as a pro-social act (Wu 2019), which is defined as "voluntary behavior intended to benefit another" (Eisenberg 1996). Behavior in general and pro-social behavior, in particular, can be evoked by various types of motivation. In particular, the source of motivation can be rather intrinsic or extrinsic. Intrinsic motivation is defined as "doing an activity simply for the enjoyment of the activity itself, rather than its instrumental value" (Ryan and Deci 2000, 60). In contrast, extrinsic motivation is defined as "doing something because it leads to a separable outcome" (Ryan and Deci 2000, 55).

Crowding-out describes the phenomenon that intrinsic motivation is crowded out by extrinsic motivational factors such as rewards (Deci et al. 1999). In a meta analysis on intrinsic motivation and incentives, Cerasoli et al. (2014) identify that intrinsic motivation can be crowded out when incentives are implemented in a performance-salient manner.[3] In an experiment on contributions to a social networking service, Vilnai-Yavetz and Levina (2018) find support for crowding out of intrinsic motivation by extrinsic motivation triggered by financial incentives. Khern-am nuai et al. (2018) analyze (within-subject) how introducing financial incentives affect the reviewing behavior of intrinsically motivated customers, finding that the number of reviews decreases, but the level of reviews' quality is not affected.

Research Hypothesis 7. *The implementation of economic incentives in reputation systems crowds out altruism as a driver of customer ratings.*

The impact of pro-social attitude and reciprocity is investigated in Simpson and Willer (2008), who find that egoists react more sensitively to reciprocal incentives. If no reciprocal incentives are implemented, altruists give substantially more than egoists. Wu (2019) finds evidence that reciprocity to the seller crowds out intrinsic motivation to publish reviews. Gutt et al. (2017) also find evidence that younger customers provide less customer reviews when unconditional gifts are provided, which points to crowding-out of intrinsic motivation. Hence, we hypothesize the following:

Research Hypothesis 8. *The implementation of reciprocal instruments in reputation systems crowds out altruism as a driver of customer ratings.*

Crowding-out in markets is identified by Fehr and Gächter (2001), where cooperation based on reciprocity is driven out by financial incentives. Indirect peer reciprocity is the motive arising from a moral obligation to give something back to the community. When the received information is no longer perceived as a kind act, but as a rational and profit maximizing strategy, the feeling of being morally obliged might decrease or entirely cease. We hypothesize the following:

Research Hypothesis 9. *The implementation of economic incentives in reputation systems crowds out moral obligation as a driver of customer ratings.*

Research Hypothesis 10. *The implementation of reciprocal instruments in reputation systems crowds out moral obligation as a driver of customer ratings.*

2.5. Further motives

As described above, further motives are identified to affect the propensity to publish a review. Social affiliation is one of these observed motives. For example, Cheung and Lee (2012) identify sense of belonging as a driver for providing a review. Although further evidence is found in Hennig-Thurau et al. (2004)

[3] Surveying the literature, Frey and Jegen (2001) show the robustness of crowding-out.

and Munzel and H. Kunz (2014), in their more abstract setting Solow and Kirkwood (2002) find that provision to a public good is at least not primarily affected by group identity. Chen et al. (2010) show that implementing instruments to stimulate social affiliation work in principle, but also can backfire. Providing users with the median user's number of reviews leads to a 530% increase of reviews from customers below the median. However, the number of users above the median decrease by 62 percent.

In several studies, building up a reputation was identified to drive customer ratings (Hoyer and van Straaten 2021; Cheung and Lee 2012; Wu 2019). However, by using data from TripAdvisor, Liu et al. (2016) show that glory-based incentives on online platforms yield a lower quality of reviews as the reviewers' status increases. In addition, reviewers with a higher status are less likely to post extreme reviews. Self-expression of feelings and opinions also motivates customers to publish reviews (Peddibhotla and Subramani 2007; Hennig-Thurau et al. 2004; Munzel and H. Kunz 2014).

Crowding effects. Non-monetary instruments can also cause crowding-in and crowding-out of other motives for providing customer reviews: Sun et al. (2017) investigates the impact of implementing monetary incentives on the likelihood of contributing reviews subject to the social connectedness in the community. They find an increase of 1400% for less-connected members, but also a decrease of 90% for more-connected members. Hence, a sense of belonging seems to be crowded out when extrinsic monetary rewards are implemented. In the context of customer review systems, Wu (2019) finds evidence for the crowding-in of intrinsic motivation by extrinsic motivation, as status recognition crowd in enjoyment while writing reviews. Burtch et al. (2017) combine financial incentives and social norms to stimulate online reviews. Showing participants how many customers already provided reviews increases the likelihood of eliciting further and longer customer reviews. As already mentioned above, the number of reviews is affected more strongly by financial incentives. Hence, financial incentives and social norms can be seen as complementary instruments to elicit reviews.

3. Experimental Design

This study introduces a novel experimental design to investigate the impact of economic incentives and reciprocity on customers' reviewing behavior. To describe the structure of the experiment, we first mention that all participants act as customers on a marketplace and are faced with two decisions in each period. In the first step, each participant buys one out of seven products. Afterwards, they decide whether to rate the bought product, thereby helping other customers to make a more informed decision in the subsequent periods.

Buying and experiencing the seven products is associated to static, but unknown average satisfaction levels, as provided in Table 1. Since the satisfaction is accumulated and determines the payoff at the end of the experiment, participants are interested in finding the best product (i.e., product A, cf. Table 1)[4].

Table 1: Average satisfaction levels for the products A to G

	Provided Satisfaction by Products						
Product	A	B	C	D	E	F	G
S_j	18	8	11	14	7	13	6

To increase uncertainty in the experiment, the level of satisfaction also varies over time, indicated by $\epsilon_i(t) \in \{-2, -1, 0, 1, 2\}$. Hence, the satisfaction from choosing a product j for customer i in period t is $s_{i,j}(t) = S_j + \epsilon_i(t) \in [4, 20]$. The customers do not know the information provided in Table 1. At the beginning of the experiment in period 1 they only know the interval of the average satisfaction level of each product $S_j \in [6, 18]$, resulting in 13 possible satisfaction states for each of the seven products. The initial setting, hence, can be interpreted as a new online shop without any rated products.[5]

After buying a product (in period 1 randomly, since no information is available to discriminate between the products), customer i is informed about the actual experienced satisfaction $s_{i,j}$. Then, this person can decide to publish a review associated with costs $C = 2$ and, thereby, reduce the uncertainty (cf. Table 2) about the chosen product for four other customers by four states (cf. Figure 1). Before their rating decision, the subjects are informed about the recipient's current level of uncertainty, i.e., their possible satisfaction states, and whether their rating would reduce the recipients' uncertainty for the chosen seller. If no review is published, $C = 0$ in this period.

Table 2: Effect of customer reviews: Each of the first three reviews eliminates the uncertainty of the underlying average satisfaction of the reviewed product by four states. The resulting possible satisfaction states are also depicted.

Number of reviews	0	1	2	3	4 or more
Eliminated satisfaction states	0	4	8	12	12
Possible satisfaction states	13	9	5	1	1
Exemplary Uncertainty Interval D	[6,18]	[7,15]	[11,15]	14	14

To test for the efficiency of different incentive mechanisms, we implemented three treatments.

Baseline Treatment (BL). In the baseline treatment, no incentives are implemented. Participants are asked to publish a review, pointing out the utility for the other participants.

[4] We randomized the products in each session and run to control for order effects.

[5] Participants also know that there is one single best product with $S_j = 18$, one worst product with $S_j = 6$, and that $S_i \neq S_j$ for $i, j \in \{A, ..., G\}$ and $i \neq j$.

Incentive Treatment (IN). In the incentive treatment, participants receive a conditional rebate when they opt in to publish a review. Then, participants receive a rebate $I = 4\ ECU$. Since $I > C$, it is rational to publish a review, as additional revenues can be generated.

Gift Treatment (GI). In the gift treatment, participants are asked to publish a review, pointing out the utility for the other participants. With this request, participants receive an unconditional rebate, i.e., a gift $I = 4\ ECU$. Overcompensating the costs for publishing a review, participants can decide whether to keep the extra gift all for themselves or to publish a review and, thereby, let four other customers benefit as well.

The earnings in the experiment are the accumulated satisfaction points minus costs for publishing a review plus potential incentives:

$$E_i = \sum_{t=1}^{6} E_i(t),\ with\ E_i(t) = s_{i,j}(t) - C(t) + I(t). \quad (1)$$

3.1. Potential motives

As explained in chapter 2, various motives can affect the reviewing behavior. Subsequently, it is lined out, which motives are present in the different treatments. Direct peer reciprocity is not effective in the whole experiment as customer group g receives reviews from customer group g-1 and publishes reviews for group g+1. As the experiment is conducted for six periods and seven customer groups are implemented, reviews by members of group g cannot initiate reciprocal reviewing behavior of group g-1 for group g (cf. Figure 1).

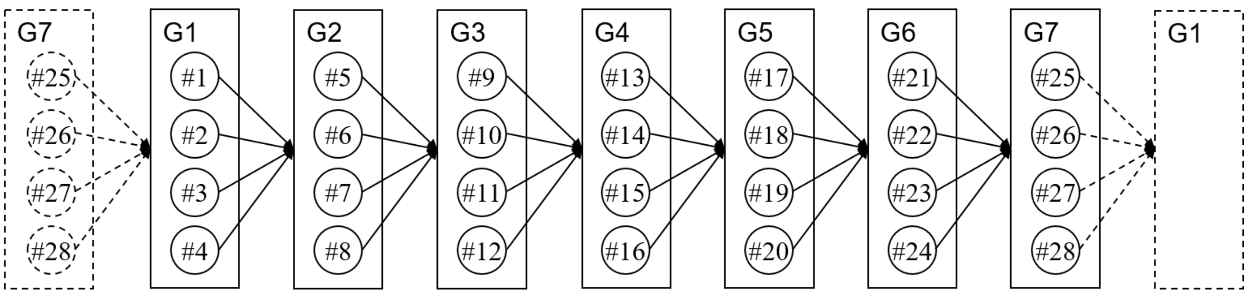

Figure 1: Participants are matched in groups of four and these groups are ordered randomly in the depicted manner to prevent direct peer reciprocity as, for instance, G1 provides information to G2 and receives information from G7.

Further, participants cannot build up a reputation as they act completely anonymous and their decision cannot be linked to them, eliminating this motive. Neither is group identity a motive for reviewing, as participants do not interact with their matched group partners.

Altruism (including the enjoyment of helping others) and indirect peer reciprocity (moral obligation) generated by receiving information on the products by members of g-1, leading to own provision of a review for group g+1, can be prevalent motives in each treatment, since they are independent of financial motives.

The economic incentive is effective in the incentive treatment as the costs for writing a review are smaller than the provided incentives: $C = 2 < I = 4$. In the baseline treatment and gift treatment, it is economically rational to not write a review.

Seller reciprocity (between seller and customers) is a potential motive in the gift treatment. There, customers receive a request with a gift (i.e., a kind act), which could lead to a reciprocally kind reaction.

In contrast to the baseline treatment, crowding out of intrinsic motivation can be effective in the gift treatment and the incentive treatment. Taking all this together, motives are effective as illustrated in Table 3.

Table 3: Potential motives in BL (baseline treatment), IN (incentive treatment), and GI (gift treatment). Note that self-enhancement/reputation, a sense of belonging, and direct peer reciprocity are discarded by the experimental design. Indirect peer reciprocity might be effective in all treatments, whereas seller reciprocity is only effective in GI treatment. *CO* abbreviates *crowding-out*.

Motives	**BL**	**IN**	**GI**
Altruism	Yes	Yes (CO?)	Yes (CO?)
Peer Reciprocity	Yes	Yes (CO?)	Yes (CO?)
Seller reciprocity	No	No	Yes
Economic incentive	No	Yes	No

3.2. *Questionnaire*

The design allows us to directly observe the overall impact of the various incentive programs in customer review systems. To investigate the driving motives, disentangling altruism, and moral obligation in more detail and to quantify potential crowding-out effects, we additionally conduct a questionnaire by using the HEXACO items (16 items on honesty-humility and four items on altruism), to elicit pro-social and altruistic attitudes from the participants (Lee and Ashton 2004).

3.3. *Operationalization of hypotheses*

Considering the effective motives in our treatment (cf. Table 3) we operationalize the research hypotheses of chapter 2 in Table 4. Superscripts *A* and *NA* indicate *Altruists* resp. *Non-Altruists* who are classified regarding their answers in the questionnaire. Superscripts *M* and *L* indicate *More-Receivers* resp. *Less-Receivers*, i.e., whether subjects received more than the median amount of ratings (potentially causing peer reciprocity) or less than the median amount. Research Hypotheses 7 to 10 on crowding-out effects are identified by comparing subgroups across different treatments. The identification strategy is that altruists are intrinsically motivated and non-altruists are not. Comparing the difference between altruists and non-altruists in the BL treatment provides an unbiased difference between these groups through intrinsic motivation. Decreasing differences in the IN (*RH7*) or GI treatment (*RH8*) indicate evidence that intrinsic motivation is crowded out. Crowding-out of peer reciprocity (*RH9, RH10*) is investigated in a similar manner.

Table 4: Operationalization of research hypotheses (RH)

	Name	RH	Formula
Incentive Schemes	Existence of pro-social behavior	1	$reviews_{BL} > 0$
	Effect of conditional rebates	5	$reviews_{IN} > reviews_{BL}$
	Effect of unconditional rebates	4	$reviews_{GI} > reviews_{BL}$
	Efficiency of incentive scheme	6	$reviews_{IN} > reviews_{GI}$
Altruism	Effect of altruism	2	$reviews^{A}_{BL} > reviews^{NA}_{BL}$
	Crowding-out by conditional rebates	7	$reviews^{A}_{IN} - reviews^{A}_{BL} < reviews^{NA}_{IN} - reviews^{NA}_{BL}$
	Crowding-out by unconditional rebates	8	$reviews^{A}_{GI} - reviews^{A}_{BL} < reviews^{NA}_{GI} - reviews^{NA}_{BL}$
Peer Reciprocity	Effect of peer reciprocity	3	$reviews^{M}_{BL} > reviews^{L}_{BL}$
	Crowding-out by conditional rebates	9	$reviews^{M}_{IN} - reviews^{M}_{BL} < reviews^{L}_{IN} - reviews^{L}_{BL}$
	Crowding-out by unconditional rebates	10	$reviews^{M}_{GI} - reviews^{M}_{BL} < reviews^{L}_{GI} - reviews^{L}_{BL}$

3.4. Procedure

The experiment was carried out online with z-Tree (Fischbacher 2007) unleashed (Duch et al. 2020).[6] We recruited the participants with Orsee (Greiner 2015) from the subject pool of BaER-Lab in Paderborn, Germany. A total of 8 sessions were conducted in December 2020 and March 2021. In each session we conducted two treatments. Before the experiment, subjects were informed that the session was separated in two parts and that they receive further instructions after the first part. Hence, in the first run, subjects did not know the objective of the experiment. However, the instructions of the second part pointed out differences between the first and second part, allowing inferences about the experiment's objective. The within-subject attribute allows us to analyze, ceteris paribus, the effect of two incentive schemes on reviewing behavior since pro-social attitudes are constant. Because everything is reset and constellations are re-matched after the first run, there should not be differences regarding expectations of the behavior of the new partners. Controlling for this and other biasing effects, we varied the order of treatments. Overall, we conducted the combinations illustrated in Table 5.

[6] Figures A.6 and A.7 provide screenshots of the implemented decisions in the experiment.

Table 5: In each session two treatments were conducted. This allows an investigation of crowding-out effects in a within-subject manner.

Combination	BL-IN	IN-BL	GI-BL	BL-GI	IN-GI	GI-IN
1st run	BL	IN	GI	BL	IN	GI
2nd run	IN	BL	BL	GI	GI	IN
Sessions	1	2	3 / 7	4 / 8	5	6

The sequence of each session is illustrated in Figure 2. There are 28 participants acting as customers in each session. Participants are matched in groups of four, resulting in seven groups in each session.[7]

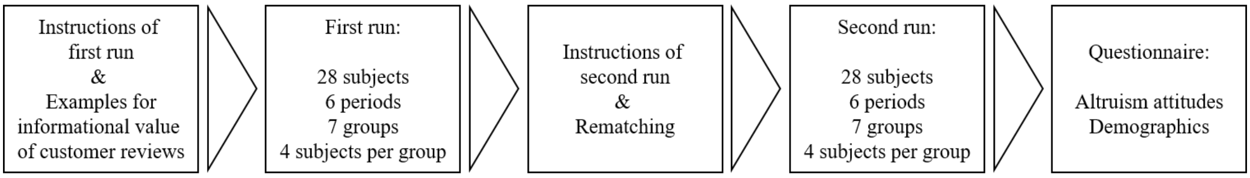

Figure 2: Procedure of experiment sessions

The sessions lasted 75 minutes on average and participants earned EUR 16.40 on average (including the show-up fee of EUR 2.50).

4. Results

In the subsequent sections, we provide a summary of the statistics and describe the sample in Table 6.[8]

4.1. Incentive schemes

For parts of analysis we pool the observations of all combination groups. As the order of treatments is changed between combination groups, we assume that learning and order effects are ruled out. The resulting publishing rates over treatments are depicted in Figure 3.[9]

[7] Due to technical problems we conducted the first session with only 24 participants. Note that when 6 groups are involved, direct peer reciprocity could also be excluded as a potential motive for publishing reviews.

[8] For sessions 1-6, we have seen unexpected heterogeneity in the data. In particular, we found session effects for gender ($\chi^2 : p = 0.023$, $z = 12.89$). For age, field of studies or altruism we have not found significant session effects. A summary of subject samples on a session level is provided in the appendix (cf. Table A.9). Due to confounded results and insufficient data of male subjects in the BL-GI and GI-BL combination groups, we conducted sessions 7 and 8. On the level of combination groups (cf. Table 5) these differences disappear. That is, by pooling data of sessions 3 and 7, resp. session 4 and 8, we obtain comparable subject samples. Hence, subsequent analysis is conducted between combination groups.

[9] Values are provided in Table A.10.

Table 6: Subject sample of combination groups

	BL-IN	IN-BL	GI-BL	BL-GI	IN-GI	GI-IN	Total
n	24	28	55	56	28	28	219
Age	22.8 (2.3)	23.0 (2.6)	24.8 (4.5)	24.4 (3.2)	24.1 (5.8)	23.6 (3.0)	24.0 (3.8)
Male	20.8%	46.4%	34.6%	30.4%	42.9%	42.9%	35.6%
Studies: Economics	45.8%	57.1%	37.0%	58.9%	32.1%	53.6%	47.5%
Studies: Education	20.8%	25.0%	40.7%	30.4%	39.3%	17.9%	30.6%
Studies: Engineering	12.5%	7.1%	3.7%	5.4%	7.1%	17.9%	7.8%
Studies: Humanities	16.7%	10.7%	18.5%	5.4%	21.4%	10.7%	13.2%
Altruists	45.8%	42.9%	43.6%	46.4%	28.6%	57.1%	44.3%
Treatment 1	BL	IN	GI	BL	IN	GI	
Ratings	27.8%	92.9%	40.9%	31.0%	87.5%	31.0%	48.3%
Ratings per subject	1.67	5.57	2.45	1.86	5.25	1.86	2.90
Treatment 2	IN	BL	BL	GI	GI	IN	
Ratings	95.1%	14.3%	21.5%	31.9%	19.1%	85.1%	39.1%
Ratings per subject	5.71	0.86	1.29	1.91	1.14	5.11	2.35

Pro-social Behavior. In BL treatment we analyze whether there are ratings provided that could be triggered by altruism, indirect peer reciprocity (i.e., moral obligation) or game effects. A binomial test shows that the rating frequency of 24.4% is significantly higher than 0%. In particular, a one-sided binomial test with assumed probability of 0.1 indicates $p < 0.0001$. That is, even in the absence of other motives such as social affiliation, reputation or self-expression we see a substantial rating behavior.

Seller Reciprocity. Analyzing the effect of reciprocity on publishing behavior, we test for differences between published ratings in the GI treatment and BL treatment. We find strong evidence for the efficacy of seller reciprocity ($\chi^2: \; p < 0.0001, \; z = 15.91$). These results are also supported by comparing the accumulated publishing decisions of the subjects (Mann-Whitney U test: $p = 0.0603, \; z = 1.879$). Controlling for heterogeneity, we only consider subjects from combination groups *GI-BL* and *BL-GI* and find similar results (MWU: $p = 0.0702, \; z = 1.810$).

Economic Incentives. Comparing the difference between published ratings in the IN treatment and BL treatment, we test for the overall impact of economic incentives on the propensity to publish a rating. We find strong evidence for the efficacy of economic incentives ($\chi^2: \; p < 0.001, \; z = 669.59$). The comparison of the accumulated publishing decisions of the subjects between both treatments also supports this result (MWU: $p < 0.001, \; z = 8.983$). When we repeat this analysis only for subjects from combination groups *IN-BL* and

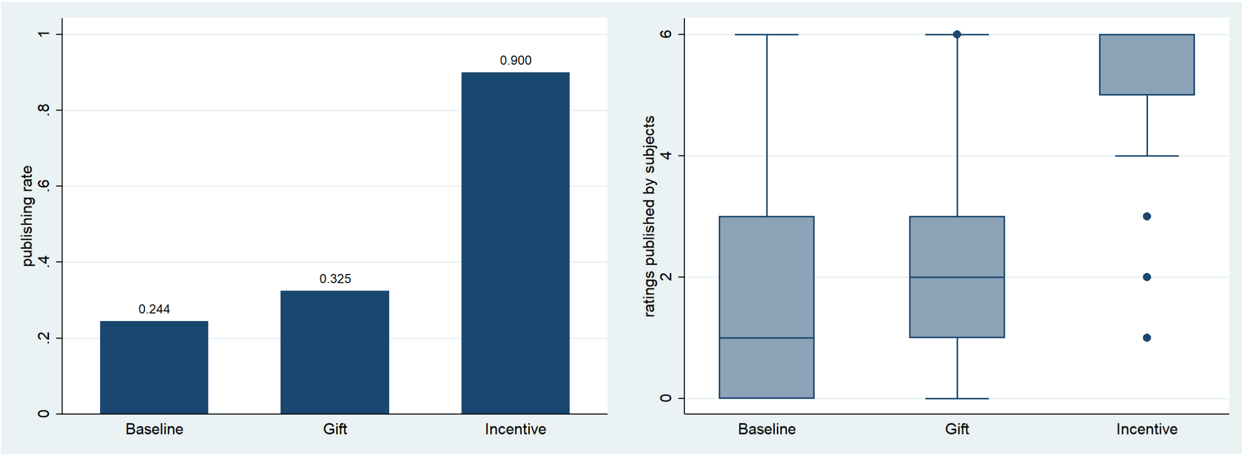

Figure 3: left: Publishing rates over treatments; right: Box plots of accumulated ratings of subjects over treatments.

BL-IN, the results receive further support (MWU: $p < 0.0001,\ z = 5.926$).

Efficiency of Incentive Scheme. Comparing the absolute difference between published ratings in the GI treatment and IN treatment, we can test which mechanism has a higher overall impact on the propensity to publish a rating. Results show the unambiguously stronger effect of economic incentives in our experiment ($\chi^2 : p < 0.001,\ z = 524.67$). Analysis with the accumulated ratings provides further evidence (MWU with all observations: $p < 0.0001,\ z = -8.847$; MWU with observations from combination groups *IN-GI* and *GI-IN*: $p < 0.0001,\ z = -5.733$).

Result 1. *Economic incentives elicit most reviews, gifts the second most reviews. Even without stimulating instruments, subjects publish reviews. All differences are statistically significant.*

4.2. Altruism

Using data from the questionnaire, we can estimate the level of altruism. For a construct for altruism, we use two items from the questionnaire, as these are closest to our research design, and then calculate the arithmetic mean. Subjects with higher than median values are then classified as altruists (cf. Table A.8). This measure is independent of combination groups ($\chi^2 : p = 0.436,\ z = 4.84$).

Effect of Altruism. Comparing the absolute difference between published ratings of altruists and non-altruists in the BL treatment, we can test for the overall impact of altruism on the propensity to publish a rating (cf. Figure 4 *left*). Using this classification, we see that in BL treatment the publishing rate of altruists is 31.3%. Non-altruists publish significantly less often in 18.9% of the cases ($\chi^2 : p < 0.0001,\ z = 20.10$). Figure 4 *(right)* provides box plots of the number of ratings for altruists and non-altruists. Testing with a

Mann-Whitney-U test for differences between altruists and non-altruists in the BL treatment provides further evidence that altruism affects publishing ratings (Mann-Whitney U test: $p = 0.0038$, $z = -2.898$).

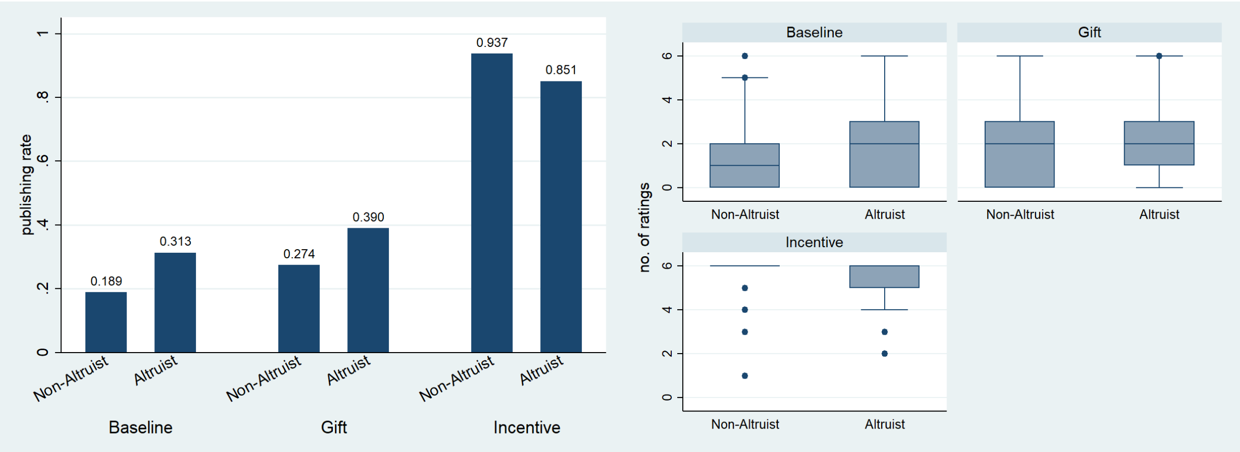

Figure 4: left: Publishing rates over treatments by altruistic attitude; right: Box plots of number of ratings over treatments by altruistic attitude

Crowding-out Altruism by Economic Incentives. Next, we test Hypothesis 7 on crowding-out intrinsic motivation by introducing economic incentives for providing ratings. Allowing within-analysis, we first test with data from combination groups *IN-BL* and *BL-IN* as subjects made decisions under both treatment conditions. We find strong evidence that altruism is crowded out, as the increase in ratings of altruists in the IN treatment is on average 1.2 ratings smaller in comparison with non-altruists. Note that the difference cannot be solely explained by the higher rating level of altruists in BL treatment, as altruists also publish less in the IN treatment (Altruists: 5.4 ratings, non-altruists: 5.8 ratings, MWU: $p < 0.0001$, $z = 5.01$). Furthermore, these results are supported when we consider all subjects in IN and BL treatment. On average, altruists publish 1.9 ratings in BL treatment and 5.1 ratings in IN treatment, while non-altruists publish 1.1 ratings in BL treatment and 5.6 in IN treatment. The MWU test shows that altruists, in comparison with non-altruists, publish significantly more ratings in BL treatment ($p = 0.0038$, $z = -2.90$) and significantly less ratings in IN treatment ($p = 0.0096$, $z = 2.59$). That is, altruism is crowded out by economic incentives.

Crowding-out Altruism by Seller Reciprocity. We test the introduction of a reciprocal instrument on the rating behavior of altruists and non-altruists. Allowing within-analysis, we only use data from combination groups *GI-BL* and *BL-GI* as subjects made decisions under both conditions. On average, altruists publish 0.42 ratings more in GI treatment and non-altruists 0.75 ratings more. We do not find significant differences between altruists and non-altruists regarding the difference between both treatments (MWU: $p =$

0.1136, $z = 1.582$): i.e., the increase in ratings from BL to GI treatment does not differ between both groups. When we also consider observations from other combination groups, we find that altruists publish significantly more ratings in comparison with non-altruists in GI treatment (MWU: $p = 0.0064$, $z = -2.729$). The size of the effect is comparable to BL treatment (MWU: $p = 0.0038$, $z = -2.898$), which speaks against the crowding-out of intrinsic motivation. When we test for differences between GI and BL treatment, we find that non-altruists publish 0.51 ratings more in GI treatment (MWU: $p = 0.0125$, $z = 2.498$) on average. Similarly, altruists publish 0.46 ratings more in GI treatment (MWU: $p = 0.0699$, $z = 1.813$) on average. Hence, there is no significant evidence that introducing reciprocity through unconditional rebates crowds out altruism.

Result 2. *Altruism is a significant motive for publishing reviews. Economic incentives crowd out this intrinsic motivation. In contrast, gifts do not crowd out altruism as a motive for publishing reviews.*

4.3. Peer reciprocity

To check the relevance of indirect peer reciprocity on rating behavior, we analyze the dependency of received and provided reviews. High correlation would support the existence of indirect peer reciprocity, leading to the received information being forwarded due to a moral obligation. Therefore, for each combination group and period we generate the median value of ratings received to control for time and treatment effects. Rating decisions in which subjects receive more than the median information are then classified as *more-receiving*. Analogously, decisions in which subjects receive less than the median information or the median information are classified as *less-receiving* or *median-receiving*, respectively.[10] Overall, we characterize 617 observations as more-receiving, 434 observations as less-receiving, and 1577 observations as median-receiving.[11] Using this classification, Figure 5 *(left)* illustrates the publishing rates over these groups. We classify subjects with regard to the number of publishing decisions in which they received more than the median information beforehand. Figure 5 *(right)* provides the resulting box plots of the number of ratings by treatments over a number of decisions in which subjects received more than the median information.

Effect of Peer Reciprocity. In BL treatment, subjects published 54 ratings in 247 more-receiving situations and 33 ratings in 133 less-receiving situations. We do not find significant differences between more-receiving and less-receiving decisions in publishing ratings (χ^2-test: $p = 0.514$, $z = 0.4261$). When peer reciprocity is effective in BL treatment, subjects with a higher number of more-receiving decisions should publish more ratings. Testing this with Kruskal-Wallis test, we do not find significant differences ($p = 0.1963$, $z = 6.038$): i.e., we do not find evidence that indirect peer reciprocity is effective.

[10] In Table A.11 the received information is summarized for each combination group and period.

[11] Table A.12 provides data on more-receiving, median-receiving, and less-receiving decisions and the ratings provided in these situations.

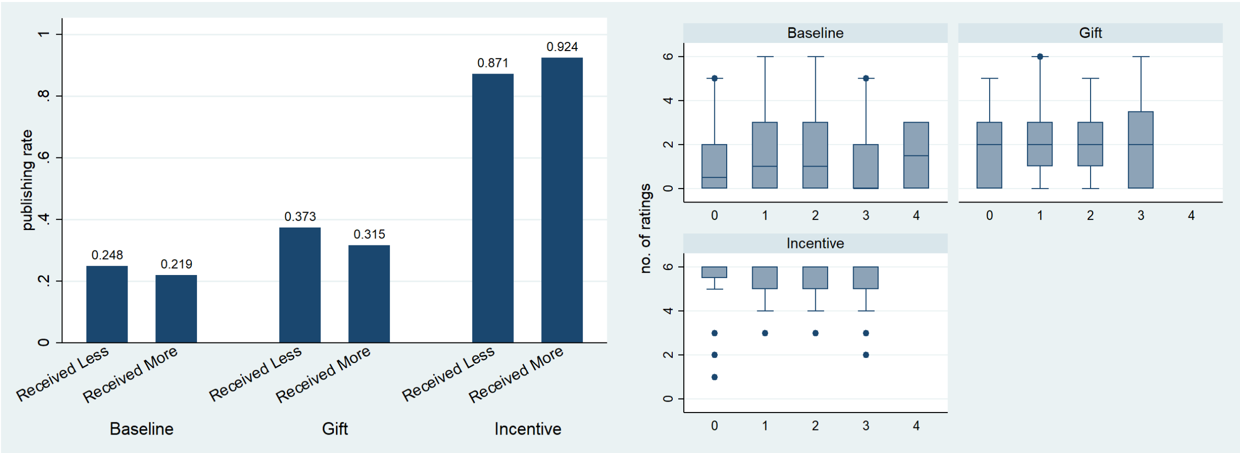

Figure 5: left: Publishing rates by treatments in situations of receiving more or less than the median information; right: Box plots of number of ratings by treatments over the number of situations in which subjects received more than the median information.

Crowding-out Peer Reciprocity by Economic Incentives. Research Hypothesis 9 states that economic incentives can crowd out indirect peer reciprocity as the moral obligation might decrease in the IN treatment. We find no evidence for indirect peer reciprocity in BL treatment. However, we test whether introducing economic incentives also has an indirect effect on providing ratings by indirect peer reciprocity. We neither find evidence of an effect of peer reciprocity on the provision of ratings with χ^2-test ($p = 0.219,\ z = 1.5094$) nor with Kruskal-Wallis on an accumulated level ($p = 0.9692,\ z = 0.250$).

Crowding-out Peer Reciprocity by Seller Reciprocity. We test whether seller reciprocity affects indirect peer reciprocity. Also with seller reciprocity, we find no effect of indirect peer reciprocity on publishing decisions with χ^2-test ($p = 0.9692,\ z = 0.250$) nor with Kruskal-Wallis on an accumulated level ($p = 0.8700,\ z = 0.713$). Hence, we conclude the following:

Result 3. *Independent of stimulating incentives, indirect peer reciprocity is not a significant motive for publishing ratings.*

4.4. Further results

Quality of Ratings. The quality of ratings differs between treatments. By design, the quality of a seller is certain after three ratings of the sender group. When subjects publish although the quality is certain or when they publish in period 6 of a run, these ratings do not help the receiver group. The accordingly classified ratings are shown for the treatments in Table 7. By conducting a pairwise comparison of the helpfulness of ratings between treatments, we find that ratings in IN treatment are significantly less helpful in comparison with BL treatment ($\chi^2:\ p < 0.001,\ z = 97.18$) and GI treatment ($\chi^2:\ p < 0.001,\ z = 111.51$).

Table 7: (Un)helpfulness of ratings in treatments

Treatment	BL	GI	IN
Helpful	207	277	289
Unhelpful	32	49	294

Ratings of treatments GI and BL are not significantly different in their helpfulness ($\chi^2: \; p = 0.582, \; z = 0.3026$).

Drop of Ratings in the Second Run. Our within-subject design allows us to observe the subjects' behavior for different treatments. Although we changed the order in the sessions to rule out the order effect, these order effects do exist. Over all combination groups, there are 634 ratings in the first run and 514 ratings in the second run. This difference is highly significant ($\chi^2: \; p < 0.001, \; z = 22.27$).

5. Conclusion

Online market platforms need the participation of customers in their reputation systems. As a result, instruments for stimulation are proposed and empirically tested in the literature. However, as it is difficult to compare the various instruments in the field, we built reputation systems with various incentive structures in an economic experiment and compared reciprocal instruments with direct economic incentives.

Without extra stimulation we find participation that can be explained by altruism, but not by indirect peer reciprocity. We further find that reciprocity initiated by sellers with unconditional rebates and requests to rate the experience also motivates additional ratings. We also find evidence that economic incentives are a strong motive to publish a rating in reputation systems.

Although both financial instruments increase participation in reputation systems and, hence, are reasonable implementations, the impact of economic incentives on the quantity of ratings is much higher. However, this instrument also crowds out ratings that are driven by intrinsically motivated customers. Given the evidence that the quality of customer reviews elicited by economic incentives is lower, this drawback can be severe, as overall information in reputation systems might even decrease when economic incentives for elicitation are implemented because high quality reviews might decrease and cannot be compensated by more low quality reviews.

Therefore, unconditional rebates might be more appropriate to elicit customer ratings. This instrument, which builds on reciprocity, shows the advantage that intrinsically motivated customers are not crowded out from participation in reputation systems. That is, high-quality customer reviews from intrinsically motivated participants would still be published and additional less detailed customer reviews could be elicited by this instrument to collect experience from other customer groups that would otherwise not report.

Our results indicate that indirect peer reciprocity (resulting from a feeling of moral obligation) is not effective in our experiment. Thereby, our experimental design abstracts from the various qualities of reviews as participants could only decide to rate or not to rate. Further research is therefore needed to validate the impact of our results on the overall efficiency of reputation systems. Only when our results with regard to the absence of crowding-out are supported, can reciprocal instruments be a substantial alternative to economic incentives. Evidence in the literature is ambiguous and calls for further research.

Acknowledgements

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.

References

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

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

Andreoni, J. and Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The economic journal*, 103(418):570–585.

Askalidis, G., Kim, S. J., and Malthouse, E. C. (2017). Understanding and overcoming biases in online review systems. *Decision Support Systems*, 97:23–30.

Batson, C. D. and Powell, A. A. (2003). Altruism and prosocial behavior. *Handbook of psychology*, pages 463–484.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142.

Burtch, G., Hong, Y., Bapna, R., and Griskevicius, V. (2017). Stimulating online reviews by combining financial incentives and social norms. *Management Science*, 64(5):2065–2082.

Cabral, L. and Li, L. (2015). A dollar for your thoughts: Feedback-conditional rebates on ebay. *Management Science*, 61(9):2052–2063.

Cerasoli, C. P., Nicklin, J. M., and Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological bulletin*, 140(4):980.

Chen, H., Duan, W., and Zhou, W. (2017). The interplay between free sampling and word of mouth in the online software market. *Decision Support Systems*, 95:82–90.

Chen, Y., Harper, F. M., Konstan, J., and Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review*, 100(4):1358–98.

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.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and economic behavior*, 46(2):260–281.

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6):627.

Duch, M. L., Grossmann, M. R., and 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.

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.

Falk, A. (2007). Gift exchange in the field. *Econometrica*, 75(5):1501–1511.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of economic perspectives*, 14(3):159–181.

Fehr, E. and Gächter, S. (2001). Do incentive contracts crowd out voluntary cooperation? CEPR discussion paper, Centre for Economic Policy Research.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *The quarterly journal of economics*, 108(2):437–459.

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

Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., and Freling, T. (2014). How online product reviews affect retail sales: A meta-analysis. *Journal of Retailing*, 90(2):217–232.

Frey, B. S. and Jegen, R. (2001). Motivation crowding theory. *Journal of economic surveys*, 15(5):589–611.

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

Greiner, B. and Levati, M. V. (2005). Indirect reciprocity in cyclical networks: An experimental study. *Journal of Economic Psychology*, 26(5):711–731.

Gutt, D., Schlangenotto, D., and Kundisch, D. (2017). You can't buy my rating! on the pivotal effect of an unconditional gift on rating behavior. In *Proceedings of the 13. Internationale Tagung Wirtschaftsinformatik (WI 2017)*.

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.

Hoyer, B. and van Straaten, D. (2021). Anonymity and Self-Expression in Online Rating Systems - An Experimental Analysis. Working Papers Dissertations 70, Paderborn University, Faculty of Business Administration and Economics.

Hu, N., Zhang, J., and Pavlou, P. A. (2009). Overcoming the j-shaped distribution of product reviews. *Communications of the ACM*, 52(10):144–147.

Khern-am nuai, W., Kannan, K., and Ghasemkhani, H. (2018). Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research*, 29(4):871–892.

Kube, S., Maréchal, M. A., and Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, 102(4):1644–62.

Lee, K. and Ashton, M. C. (2004). Psychometric properties of the hexaco personality inventory. *Multivariate behavioral research*, 39(2):329–358.

Li, L. and Xiao, E. (2014). Money talks: Rebate mechanisms in reputation system design. *Management Science*, 60(8):2054–2072.

Liu, X., Schuckert, M., and Law, R. (2016). Online incentive hierarchies, review extremity, and review quality: Empirical evidence from the hotel sector. *Journal of Travel & Tourism Marketing*, 33(3):279–292.

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

Neumann, J. and Gutt, D. (2019). He who pays the piper calls the tune: Online review elicitation by sellers and third-party platforms in b2b markets. In *Proceedings of the 27th European Conference on Information Systems (ECIS)*.

Peddibhotla, N. B. and Subramani, M. R. (2007). Contributing to public document repositories: A critical mass theory perspective. *Organization Studies*, 28(3):327–346.

Picazo-Vela, S., Chou, S. Y., Melcher, A. J., and Pearson, J. M. (2010). Why provide an online review? an extended theory of planned behavior and the role of big-five personality traits. *Computers in Human Behavior*, 26(4):685–696.

Reimer, T. and Benkenstein, M. (2016). Altruistic ewom marketing: More than an alternative to monetary incentives. *Journal of*

Retailing and Consumer Services, 31:323–333.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67.

Simpson, B., Harrell, A., Melamed, D., Heiserman, N., and Negraia, D. V. (2018). The roots of reciprocity: Gratitude and reputation in generalized exchange systems. *American Sociological Review*, 83(1):88–110.

Simpson, B. and Willer, R. (2008). Altruism and indirect reciprocity: The interaction of person and situation in prosocial behavior. *Social Psychology Quarterly*, 71(1):37–52.

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2):274–279.

Solow, J. L. and Kirkwood, N. (2002). Group identity and gender in public goods experiments. *Journal of Economic Behavior & Organization*, 48(4):403–412.

Suls, J., Witenberg, S., and Gutkin, D. (1981). Evaluating reciprocal and nonreciprocal prosocial behavior: Developmental changes. *Personality and Social Psychology Bulletin*, 7(1):25–31.

Sun, Y., Dong, X., and McIntyre, S. (2017). Motivation of user-generated content: Social connectedness moderates the effects of monetary rewards. *Marketing Science*, 36(3):329–337.

Vilnai-Yavetz, I. and Levina, O. (2018). Motivating social sharing of e-business content: Intrinsic motivation, extrinsic motivation, or crowding-out effect? *Computers in Human Behavior*, 79:181–191.

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

Appendix A.

Table A.8: Altruism construct built on HEXACO items 97 (“I have sympathy for people who are less fortunate than I am.”) and 98 (“I try to give generously to those in need.”). [n=220]

HEXACO	Mean	Std.dev.	Median	Altruists	Non-Altruists
Item 97	3.822	.835	4		
Item 98	3.393	.808	3		
Construct	3.607	.623	3.5	97	122

Table A.9: Subject sample of sessions. In session 7 one subject could not finish the experiment due to technical problems and was excluded from analysis.

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8	Total
n	24	28	28	28	28	28	27	28	219
Age	22.8 (2.3)	23 (2.5)	24.6 (4.6)	23.5 (2.5)	24.1 (5.8)	23.6 (2.9)	24.9 (4.3)	25.3 (3.6)	24.0 (3.8)
Male	20.8%	46.4%	28.6%	10.7%	42.9%	42.9%	40.7%	50.0%	35.6%
Studies: Economics	45.8%	57.1%	28.6%	60.7%	32.1%	53.6%	44.4%	57.1%	47.5%
Studies: Education	20.8%	25.0%	46.4%	32.1%	39.3%	17.9%	33.3%	28.6%	30.6%
Studies: Engineering	12.5%	7.1%	3.6%	3.6%	7.1%	17.9%	3.7%	7.1%	7.8%
Studies: Humanities	16.7%	10.7%	21.4%	3.6%	21.4%	10.7%	3.7%	7.1%	13.2%
Altruists	45.8%	42.9%	42.9%	39.3%	28.6%	57.1%	44.4%	53.6%	44.3%
Treatment 1	BL	IN	GI	BL	IN	GI	GI	BL	
Ratings	27.8%	92.9%	44.6%	29.8%	87.5%	31.0%	37.0%	32.1%	48.3%
Ratings per subject	1.67	5.57	2.68	1.79	5.25	1.86	2.22	1.93	2.90
Treatment 2	IN	BL	BL	GI	GI	IN	BL	GI	
Ratings	95.1%	14.3%	21.4%	29.8%	19.1%	85.1%	21.6%	33.9%	39.1%
Ratings per subject	5.71	0.86	1.29	1.79	1.14	5.11	1.30	2.04	2.35

Table A.10: Rating decisions in treatments

	Treatment	BL	GI	IN
Publishing decision	yes	239	326	583
	no	739	676	65
	yes (rel.)	24.4%	32.5%	90.0%
subjects' av. ratings		1.47	1.95	5.40

Table A.11: Classification of observations as More-Receiving, Median-Receiving, and Less-Receiving. Received median information is shown in parentheses. Note that the received information corresponds to ratings that actually reduce receivers' uncertainty about a specific seller. If uncertainty about a seller is already zero, the received information of its rating is zero. Background colors of cells correspond to the Baseline Treatment , Incentive Treatment , and Gift Treatment .

		Periods Run 1						Periods Run 2						
Combination	Classification	1	2	3	4	5	6	1	2	3	4	5	6	Sum
BL-IN	More	0	0	12	8	8	4	0	0	8	12	8	8	68
	Median	24 (0)	16 (2)	0 (1.5)	12 (1)	16 (0)	20 (0)	24 (0)	20 (4)	12 (3)	0 (1.5)	16(0)	16(0)	176
	Less	0	8	12	4	0	0	0	4	4	12	0	0	44
IN-BL	More	0	0	0	12	8	12	0	12	8	4	12	4	72
	Median	28 (0)	20 (4)	20 (4)	12 (2)	12 (1)	16 (0)	28 (0)	16 (1)	20 (0)	16 (1)	16 (0)	24 (0)	228
	Less	0	8	8	4	8	0	0	0	0	8	0	0	36
GI-BL	More	0	27	12	8	27	16	0	16	16	12	27	20	181
	Median	55 (0)	20 (2)	20 (2)	20 (2)	16 (1)	28 (1)	55 (0)	16 (2)	20 (1)	28 (1)	28 (0)	35 (0)	341
	Less	0	8	23	27	12	11	0	23	19	15	0	0	138
BL-GI	More	0	12	20	16	16	20	0	24	12	24	16	8	168
	Median	56 (0)	24 (2)	32 (1)	36 (1)	24 (1)	36 (0)	56 (0)	16 (2)	20 (2)	28 (1)	40 (0)	48 (0)	416
	Less	0	20	4	4	16	0	0	16	24	4	0	0	88
IN-GI	More	0	0	12	0	12	8	0	4	8	12	8	4	68
	Median	28 (0)	16 (4)	12 (3)	16 (3)	4 (1)	20 (0)	28 (0)	12 (2)	8 (1)	16 (0)	20 (0)	24 (0)	204
	Less	0	12	4	12	12	0	0	12	12	0	0	0	64
GI-IN	More	0	4	12	0	0	12	0	0	8	8	12	4	60
	Median	28 (0)	12 (2)	12 (2)	16 (2)	28 (0)	16 (0)	28 (0)	16 (4)	8 (3)	8 (1)	16 (0)	24 (0)	212
	Less	0	12	4	12	0	0	0	12	12	12	0	0	64
	Overall Median	(0)	(2)	(2)	(2)	(1)	(0)	(0)	(2)	(1)	(1)	(0)	(0)	

Table A.12: Number of ratings and classification of observations into the groups of more-receiving, median-receiving, and less-receiving for each treatment.

	BL	IN	GI	
Ratings / More	54/247	122/132	75/238	251/617
Ratings / Median	152/598	353/392	185/587	690/1577
Ratings / Less	33/133	108/124	66/177	207/434
	239/978	583/648	326/1002	

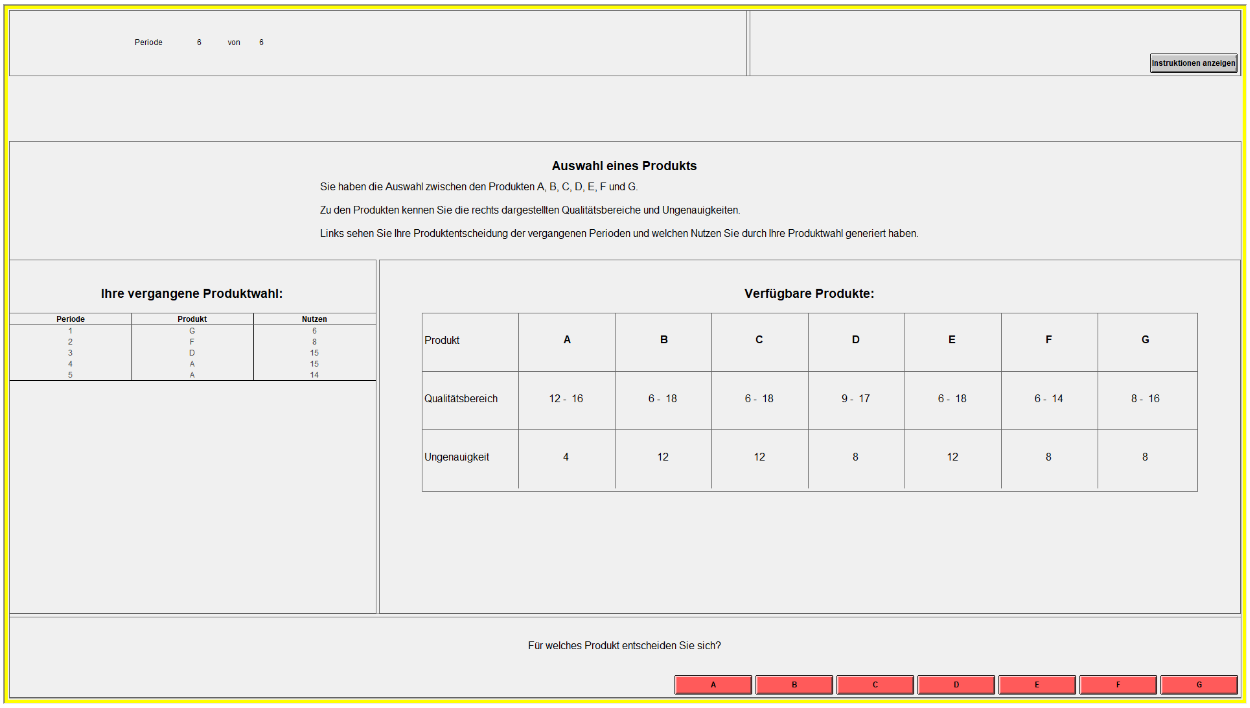

Figure A.6: Screenshot of first decision.

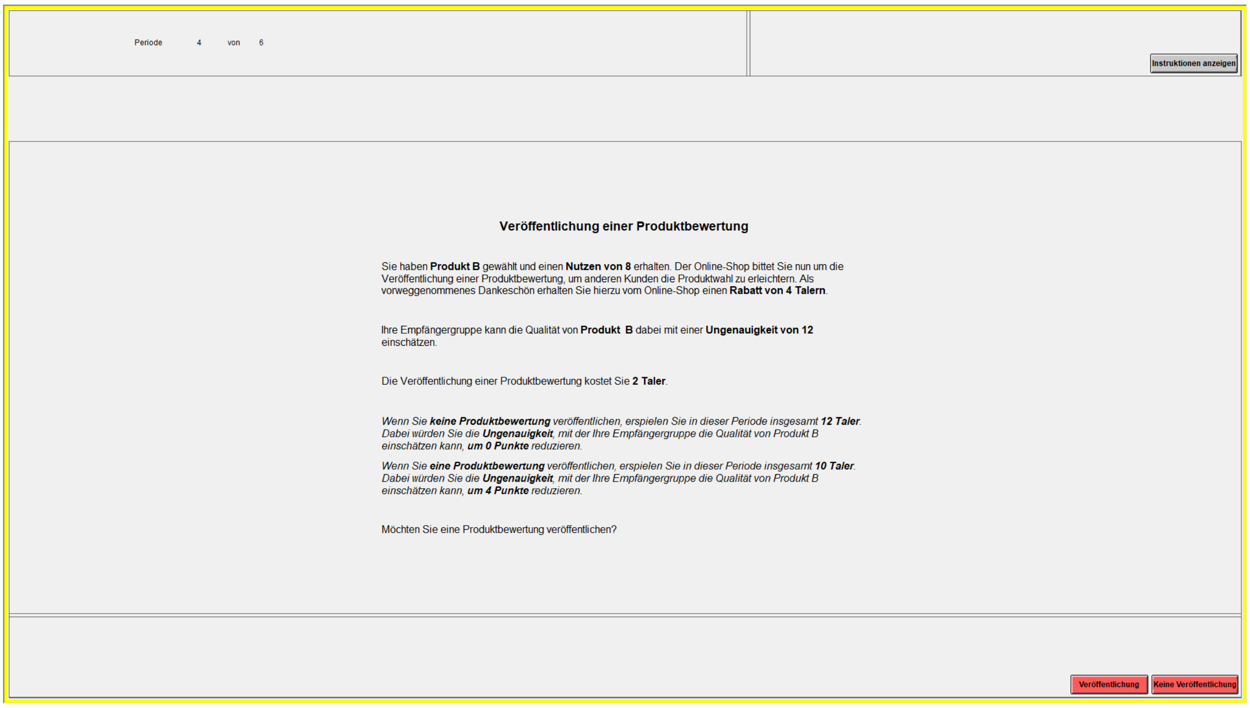

Figure A.7: Screenshot of second decision.