Explaining Reviewing Effort: Existing Reviews as Potential Driver

Christoph Rode
Universität Innsbruck

Alexander Kupfer
Universität Innsbruck

Steffen Zimmermann
Universität Ulm

Explaining Reviewing Effort: Existing Reviews as Potential Driver

Christoph Rohde[1], Alexander Kupfer[1], and Steffen Zimmermann[2]

[1] University of Innsbruck, Department of Information Systems, Production and Logistics Management, Innsbruck, Austria
`{christoph.rohde,alexander.kupfer}@uibk.ac.at`
[2] Ulm University, Institute of Business Analytics, Ulm, Germany
`{steffen.zimmermann}@uni-ulm.de`

Abstract. Online reviews systems try to motivate users to invest effort in writing a review since their success crucially depends on the reviews' helpfulness. However, other factors might influence future reviewing effort as well. We analyze whether existing reviews matter for future reviewing effort. Analyzing a dataset from Google Maps which covers 40 sights across Europe with over 37,000 reviews, we find that reviewing effort – measured by the propensity to additionally write a textual review and (textual) review length – is negatively related to the number of existing reviews. Further, also the rating distribution of existing reviews matters: If there is a large discrepancy between the existing ratings and the own rating, we observe more additional textual reviews. Our findings provide important implications for review system designers regarding the presentation of review metrics: changing or omitting the display of review metrics for potential reviewers might increase their reviewing effort.

Keywords: Online Reviews, Reviewing Effort, Online Review Platform, Existing Reviews.

1 Introduction

Consumer reviews are strongly influencing the purchase decisions of other consumers. About 80% of consumers agree that reviews directly influence their consumption decisions [1], and over 70% consider them as the most credible information source [2]. Reviews are particularly important in online markets which do not allow tangible experiences before consumption and thus come with substantial information asymmetry [3]. For such markets, the reduction of information asymmetries by increasing review helpfulness has been shown to impact future sales performance [4], and reduce costs associated with product returns [5]. However, only a minority of consumers submit an actual review [6]. Even when consumers write a review, they are typically short and lack helpful information [7], [8]. While reviewing effort is important since it can be directly related to review helpfulness [9], Cao et al. [10] argue that consumers do not invest enough effort into writing a review. An online review platform's success, however, strongly depends on the helpfulness of its reviews and thus, review system designers want to create design features which increase the

reviewing effort of their users.[1] One aspect to ensure the effectiveness of design features is to understand users' underlying cognitive mechanisms when they observe the already existing reviews before they review by themselves. Thus, being aware of these mechanisms can help to create design features that increase the effort that users invest into writing reviews and thus make reviews more helpful. This would help to reduce information asymmetries in online markets and increase social welfare by reducing costs from mismatching products.

While research has already analyzed the effect of existing reviews on the propensity to review at all [11], and on the rating valence [12], [13], there is – to the best of our knowledge – no study that has analyzed the effect of existing reviews on reviewing effort, which is the amount of effort that users invest into reviewing an object. Since reviewing effort, however, is an important factor for review helpfulness, we aim to examine the following research question:

How do previously existing reviews influence reviewing effort?

We focus on two main metrics for existing reviews, namely the number of existing reviews and the rating distribution of existing reviews. For the number of existing reviews, we develop our hypotheses based on the collective effort model and for the rating distribution of existing reviews, we apply the expectation disconfirmation theory and balance theory, respectively. We empirically test our hypotheses with online reviews from Google Maps and analyze 37,370 reviews over a period of 12 months. The review system of Google Maps represents a well-suited research environment since textual reviews are not mandatory (a star rating is sufficient) and we can thus use the propensity to additionally write a textual review as a proxy for reviewing effort. We further use the length of textual reviews as a second proxy for reviewing effort. Our findings confirm the hypotheses and suggest that existing reviews matter for future reviewing effort. In more detail, the number of existing reviews is negatively associated with both proxies for reviewing effort. For the rating distribution of existing reviews, we observe a statistically significant association with one proxy for reviewing effort: If the discrepancy between the existing rating distribution (i.e., very positive ratings) and the user's own rating (i.e., very low rating) is large, the user is more likely to additionally write a textual review.

This study has important theoretical and practical implications: First, for theoretical implications, our findings suggest that reviewing can be seen as a collective task by the users of a review system and, as described by the collective effort model, the phenomenon of social loafing is relevant in the online review setting as well: A high number of existing reviews is related to a lower propensity to additionally write a textual review and to shorter textual reviews. Additionally, we provide evidence that the expectation disconfirmation theory in combination with the balance theory not only

---

[1] Note that for the remainder of this paper, we use the term "user" for persons who are actively contributing to the online review system (i.e., reviewing). On the contrary, we use the term "consumers" for those who read online reviews but do not necessarily write reviews.

helps to explain the tendency to submit a positive or negative rating (as shown by [13]) but also to explain reviewing effort.

Second, for practical implications, our results suggest that review system designers may reconsider the presentation of review metrics. For example, review system designers could specifically highlight the number of existing reviews if the number is low and not highlighting them if the number of existing reviews is high. Similarly, review system designers could segment users into subgroups (based on, e.g., age, language or purpose of purchase) to display the (lower) number of existing reviews for the subgroup than for the (higher) total number of existing reviews. Regarding the rating distribution of the existing reviews, review system designers could highlight the existing rating distribution in the case of a high discrepancy which could then motivate users to increase their reviewing effort.

2   Related Literature

Prior research on the reviewing behavior of users has mainly been focusing on different intrinsic motivations of users to write a review. Balasubramanian and Mahajan [14], for instance, provide a theoretical framework based on social interaction utility which postulates that users gain utility through the reviewing activity. The authors distinguish between different types of user utilities that can be obtained by writing online reviews. Hennig-Thurau et al. [15], extend this framework by including two additional utility types and derive particular motives for each user utility type.

While these frameworks are useful in explaining the initial motivation of a user to act as reviewer, the influence of the user environment is neglected. This includes, among other things, that users typically observe the already existing reviews, which might influence their own reviewing behavior. In this context, Dellarocas et al. [11], study the influence of the number of existing reviews on the propensity for users to review a product at all. Based on an archival dataset of online movie reviews, the authors find that users are more likely to contribute a review for products that are less available and less successful in the market (i.e., niche products). Muchnik et al. [12], analyze another aspect of the existing reviews which are the existing ratings. They find that prior ratings create a significant social influence bias in the form of asymmetric herding effects. Users are more likely to give a positive rating when prior ratings are positive but tend to correct prior negative ratings upwards. Similarly, Ho et al. [13] also study the influence of prior ratings on reviewing behavior. The authors use the expectation disconfirmation theory to hypothesize that the disconfirmation experienced (i.e., the discrepancy between the expectation built by existing reviews and the own experience) by a user influences whether to post a rating and what rating to post. Their results suggest that individuals are more likely to post a rating when the disconfirmation degree is higher and that disconfirmation amplifies the direction of rating.

All these studies have in common that they highlight the importance of the existing reviews for future reviewing behavior: Whether users submit a review at all [11], or whether users give a positive or negative rating [12], [13]. There is – to the best of our knowledge – no study, however, which analyzes the effect of existing reviews on

reviewing effort. Considering the enormous economic relevance of online reviews and reviewing effort as an important factor for review helpfulness, this study aims to develop an understanding of the underlying cognitive mechanisms of existing reviews that influence reviewing effort.[2]

3   Theoretical Background and Hypotheses Development

To study the influence of the number of existing reviews on the propensity to review a product at all, Dellarocas et al. [11], draw on the collective effort model by Karau and Williams [16], as a theoretical framework. The difference to our analysis is, however, that we analyze the effect of existing reviews on reviewing effort, measured by (i) the propensity to additionally write a textual review and (ii) the review length. These variables are appropriate to measure the reviewing effort because users have the option to only submit a star-rating. If a user decides to additionally write a textual review, this requires additional effort from the user. Writing lengthier reviews also requires additional effort from the user and prior research already measured reviewing effort through measuring the length of textual reviews [17].

Nonetheless, we build on the collective effort model by Karau and Williams [16], for our hypotheses regarding the effect of the number of existing reviews. Reviewing an object can be seen as a collective task that is accomplished by many different individuals who provide their unique perspectives. The aim of this collective task is to provide a complete and informative picture of an object that could not be accomplished by one individual alone. Therefore, we expect the collective effort model to fit well for a mechanism that influences reviewing effort. Importantly, the collective effort model describes the underlying psychological mechanisms that leads individuals to invest less effort when working collectively than when working individually. This phenomenon – called social loafing – is especially relevant in situations when individuals have the feeling that their individual effort will not make a major impact on the outcome of the collective task [18], and that the evaluation potential of an individual's effort is diminished for a collective task [19]. Translating this to online reviews, it means on the one hand that the individual effort is clearly distinguishable from the effort of others since the review is directly attributed to a user by her name and/or her photo next to the review. On the other hand, however, individuals would feel that their own review has less impact on the total evaluation of the reviewed object if the number of existing reviews is high. Thus, even though the user can see her name and photo linked to the review, the individual review is less discoverable due to the high number of existing reviews and the user might have the feeling that her individual effort will not make a major impact on the outcome of the collective task. We therefore expect that the number of previously existing reviews reduces the individual's reviewing effort. As outlined above, users have the possibility to submit a star rating in our research environment,

---

[2] Note that Burtch et al. [17], study the influence of actively providing information about the number of previously written reviews to users on reviewing effort. Their setting, however, is very different to our setting since they directly send users the information about how many reviews have been submitted by other users to deliberately provide a social norm.

and to additionally write a (non-mandatory) textual review. Since writing an additional review implies a higher reviewing effort, we formulate our hypothesis H1 as follows.

H1: *The number of existing reviews decreases the propensity to write a textual review.*

In the same vein, we further hypothesize that this psychological mechanism also influences the effort invested in case a textual review is submitted. Users have the option to freely choose the length of their textual review. Even though users can simply copy a text, prior research has argued that lengthier reviews require more effort [17], and the fact that a user decided to write a textual review at all does not imply that the user will write a lengthy and informative review. In fact, most textual reviews are short and lack useful information [10]. Thus, additionally measuring the review length is important to determine how much effort users invest into a review once they have decided to write a textual review. Using the analogous explanation as above, a user may decrease her efforts for writing a textual review if the number of already existing reviews is large. Thus, the large number of existing reviews increases the probability of social loafing since the evaluation potential of the reviewer´s effort decreases as the review is perceived as less discoverable. Thus, the user might also believe that the review will have less of an impact because of the large number of existing reviews. Therefore, we formulate our hypothesis H2 accordingly.

H2: *The number of existing reviews decreases the length of a textual review.*

Further, we expect that not only the number of existing reviews is a relevant review metric for users' reviewing effort, but also the rating distribution of existing reviews. Users typically develop an expectation about the object to review based on the rating distribution of the existing reviews they observe and then experience the object.[3] Thus, there exists an interaction between the observed rating distribution of the existing reviews and the user's own experience which is described by the expectation disconfirmation theory [20], [21]. This theory is a cognitive theory which explains the satisfaction of individuals after experiencing and evaluating an object as a function of the disconfirmation of previously generated expectations. The expectation disconfirmation theory states that individuals are more likely to experience a high level of dissatisfaction when disconfirmation (the discrepancy between their expectations and their own evaluation of the object) is large. Vice versa individuals will experience a higher level of satisfaction if the individual's original expectation gets outperformed. Thus, the disconfirmation experienced has a moderating effect on the relationship between expectations and satisfaction after evaluation of the object. Prior research has shown that disconfirmation effects increase the propensity to submit a rating at all and that the user´s rating tends to be biased in the direction of the disconfirmation [13]. In other words, if the expectation of a user before consumption is high but she is then

---

[3] The outlined process of developing an expectation is not necessarily true for all users as they might also use other information sources. However, we expect that the sentiment of these other information sources correlates with the rating distribution of the existing reviews.

having a poor experience, disconfirmation degree is high which will increase the propensity to submit a rating and lead a more negative rating, respectively. We expect that similar cognitive mechanisms underlie for the case of reviewing effort: Users that experience a high level of dissatisfaction which arises from a high disconfirmation degree might want to vent their negative feelings. We thereby draw on the balance theory which states that individuals try to restore balance after they have become unbalanced [22], [23]. In the case of online reviews, Hennig-Thurau et al. [15], already mentioned this mechanism as one motive to review. Thus, we argue that users who experience a high disconfirmation degree are more motivated to review and will therefore invest more effort: A user who observed, for instance, a positively skewed rating distribution of existing reviews but has a poor own evaluation, experiences a high disconfirmation degree and is more likely to additionally write a (non-mandatory) textual review. Therefore, our hypothesis H3 reads as follows:

H3: *The disconfirmation degree increases the propensity to write a textual review.*

Analogous to the argumentation above, we expect that users who are writing a textual review will invest more effort into writing the review if the degree of disconfirmation experienced is high. Users who experience a high disconfirmation degree will be less satisfied and thus will invest more effort into writing the review to restore balance. Consequently, we expect that the degree of disconfirmation experienced increases the length of textual reviews. We formulate our H4 accordingly:

H4: *The disconfirmation degree increases the length of textual reviews.*

4   Empirical Analysis

4.1   Research Environment

We use the review system of Google Maps as research environment for our empirical analysis. As mentioned before, textual reviews are not mandatory on Google Maps, which allows us to analyze not only the length of reviews (i.e., H2 and H4) but also the propensity to write a textual review (i.e., H1 and H3). On Google Maps, users can virtually review every location ranging from restaurants and hotels, over shops to sights. Since restaurants, hotels and shops are subject to personal taste and depend on individual experiences (e.g., noisy room, unfriendly staff etc.), we focus on sights like bridges or fountains which are less sensitive to these subjective issues. In fact, due to the lack of interpersonal experiences, motivational aspects to writing reviews for sights are even more relevant. Further, sights are typically less sensitive to time variability than restaurants (e.g., a new chef) or hotels (e.g., renovated rooms). We, therefore, select 10 bridges, 10 squares, 10 fountains, and 10 monuments across Europe as relevant locations. We checked that these sights do not charge visitors, are accessible to the public, and are reviewed on Google Maps. We extract data from the Google Maps

website by scraping all existing reviews for each sight. Selected sights are located in 27 different cities across Europe and range from 25 de Abril Bridge in Lisbon to the Freedom Monument in Latvia.

Reviews on Google Maps do unfortunately not include a timestamp but rather relative date information (like, e.g., one week ago) which restricts our period of analysis. More specifically, for all reviews written in the last year, Google Maps provides monthly relative dates (like, e.g., 11 months ago). For all reviews that are older than one year, only yearly relative dates are provided (like, e.g., two years ago). We, therefore, focus on the period with monthly reviews for our analysis. Since the review data was downloaded at the end of November 2017, our relevant period of analysis ranges from December 2016 to November 2017. For each review, we retrieved the review date, the star rating, the textual review (if available), and the number of reviews the user has already written. Table 1 presents descriptive statistics of all available reviews for each sight group individually and all sights in aggregate. Note that the number of reviews used in our regression analysis is lower than the number of all existing reviews since there already existed reviews before the beginning of our analysis (i.e., before December 2016).

Table 1. Summary statistics.

	All Sites	Monuments	Squares	Fountains	Bridges
Avg. Rating	4.43	4,45	4.35	4.47	4.49
Std. Deviation Ratings	0.83	0.81	0.86	0.83	0.80
Minimum Avg. Rating	4.15	4.36	4.31	4.15	4.30
Maximum Avg Rating	4.73	4.61	4.73	4.55	4.72
Number of Reviews	56,794	10,466	19,394	8,804	18,130
Minimum Number of Reviews	237	380	237	978	318
Maximum Number of Reviews	3053	2931	1835	3035	2013
Number of Textual Reviews	24,035	4,542	7,611	4,065	7,817
Avg. Number of Characters	65	66	64	69	62
Std. Deviation Nr. of Characters	67	63	68	68	69
Minimum Avg. Nr. of Characters	53	53	62	57	53
Maximum Avg. Nr. of Characters	83	71	82	77	83

Notes: Minimum/Maximum Avg. Rating and Minimum/Maximum of Avg. Number of Characters refer to the average sight-level minimum/maximum of the respective location category. Minimum/Maximum of Number of Reviews refers to the absolute sight-level minimum/maximum of the respective category.

4.2 Data Preparation and Method Specification

To test our hypotheses empirically, we have to appropriately aggregate existing reviews. Since we obtain monthly review data, we aggregate existing reviews on a monthly basis as follows: When starting in month 1 with our analysis, we need to add up all existing reviews until month 1. When analyzing month 2, all existing reviews are added up until month 2, and so on.[4] Table 2 abstractly shows how our data structure looks like for one sight.

Table 2. Abstracted data structure.

	Individual review specific data						Existing review data	
Review ID	Month	Rating	Text review	Photo	Review length	User experi-ence	Nr. of existing reviews	Rating skewness of existing reviews
1	12/2016	5	Yes	Yes	150	25	200	-1.60
2	12/2016	5	No	No	–	150	200	-1.60
3	01/2017	3	No	No	–	3	202	-1.65
4	01/2017	5	Yes	Yes	100	55	202	-1.65
5	01/2017	4	Yes	Yes	120	35	202	-1.65

Notes: We use the rating skewness rather than the rating average of existing reviews as "rating distribution of existing reviews" since this better captures whether the rating distribution is overall positive. E.g., while one 1-star and four 5-star ratings have an average rating of 4.2 and a skewness of -2.24, two 3-star and three 5-star ratings have the same average rating but a higher skewness (-0.61). Thus, the more negative the value for skewness, the more overall positive the rating distribution of existing reviews.

We then obtain specific existing review data for each sight and each month, respectively. For H1 and H2, we use the number of existing reviews as an independent variable in our regression analysis. The dependent variables – proxies for reviewing effort – are given by the individual reviews and represent whether an additional textual review has been submitted (H1) and the length of the textual review (H2). We use review-specific and user-specific data as control variables.

We estimate a logit model for H1 as our dependent variable is binary and indicates whether the individual review includes a textual review or not:

$$Text_{i,t,j} = \alpha + \beta_1 \ln(Num\_Rev_{i,t-1}) + \beta_2 Controls_{i,t,j} + \epsilon_{i,t,j}, \tag{1}$$

where $Text_{i,t,j}$ represents a binary variable being 1 if a review includes a text and 0 otherwise. $\ln(Num\_Rev_{i,t-1})$ represents the natural log of the total number of existing reviews before month t (i.e., month $t-1$). $Controls_{i,t,j}$ depicts user-specific data for

---

[4] As noted above, data restricts us to rely on monthly observations of the review environment. This implies that we have to neglect reviews that are previously written in the same month in our analysis.

review *j* that might affect the likelihood to write a review. More specifically, our control variables are the natural log of user experience, the actual rating, and whether the review includes an additional picture which we expect to influence the reviewing effort as well: First, more experienced users (measured by the total number of reviews submitted by this user) might feel more related to the platform and therefore show more commitment to the collective reviewing task. Second, we expect that users, which have a specific positive or negative experience, might also include a (more detailed) explanation for this experience. Finally, when users add a picture to the review, they might also be likely to describe it.

For H2, we estimate an OLS regression with the number of characters as the dependent variable. For this analysis, the number of observations is lower as we obviously only include textual reviews. The estimation equation is defined by

$$\ln(Length_{i,t,j}) = \alpha + \beta_1 \ln(Num\_Rev_{i,t-1}) + \beta_2 Controls_{i,t,j} + \epsilon_{i,t,j} \qquad (2)$$

where $\ln(Length_{i,t,j})$ represents the natural log of number of characters for review *j*. We use the same independent variables including control variables as in Equation (1) above since we expect the same mechanism to work for writing longer reviews as for writing a review at all. Note that in both estimations, $\epsilon_{i,t,j}$ denotes the remaining error term. Robust standard errors are used in all models. For robustness, we additionally include month dummies to account for seasonal-specific like weather or holiday seasons.

To test H3 and H4, we additionally take the rating distribution of the existing reviews into account. For the underlying dataset, however, the rating distribution is highly skewed (i.e., most of the ratings have 4 stars or 5 stars). Thus, it is unlikely that a user experiences a high degree of disconfirmation in the way of an individual positive experience but a negative rating distribution of existing reviews. On the contrary, it is more likely to experience a high disconfirmation degree if the individual experience is negative since the previous ratings are outstandingly positive. Therefore, we focus our analysis on the latter case and proceed as follows: For each sight and month, we calculate the rating skewness for the existing review data as a measure for the rating distribution of existing reviews (see the note in Table 2 for an explanation why rating skewness is appropriate). We then sample each review in two groups based on the rating skewness of the existing review data. More specifically, one group represents an extremely positive rating (i.e., one third of the dataset) and the other group represents all other reviews with a less extreme but still positive rating (i.e., two third of the dataset). To finally test our hypotheses, we only pick 1-star ratings and 2-star ratings, since they imply a high degree of disconfirmation for the extremely positive rating group, and compare (i) the proportion of textual reviews for H3 and (ii) the length of textual reviews for H4 with the other group of reviews by using a one-sided t-test. Importantly, this focus on one direction of disconfirmation leads to the fact that we cannot analyze the entire scope (i.e., positive and negative disconfirmation) of Hypothesis 3 and Hypothesis 4, respectively.

4.3 Results

The results of our analysis for H1 and H2 are depicted in Table 3. For H1, Columns (i) and (ii) are relevant. The number of existing reviews is statistically significant and negative, implying that an increase in the number of reviews is negatively related to the propensity to write an additional text. This confirms our hypothesis H1 that the existing reviews in terms of the number of previously existing reviews are negatively associated with users' propensity to write a textual review. The result does not change after controlling for seasonal effects (i.e., Column (ii)). Interestingly, while the tendency is generally towards not submitting a textual review (i.e., negative and statistically significant constant), all control variables tend to increase the likelihood to submit a textual review. Since the marginal effects of a continuous and log-transformed variable are difficult to interpret, we created a new categorical variable which categorizes the number of reviews into ten bins. The first bin indicates the lowest number of existing reviews (i.e., an average of 240 reviews), the last bin indicates the highest number of existing reviews (i.e., an average of 2,325 reviews). We re-estimate Equation (1) above but replace $\ln(Num\_Rev_{i,t-1})$ with this categorical variable. This allows a convenient interpretation of the marginal effect for each bin, separately. Figure 1 below shows the average marginal effects for each bin including their 95% confidence intervals. Remarkably, we observe that the higher the number of existing reviews (i.e., bins with a higher number), the more negative the marginal effect on the propensity of writing an additional text. This observation is in line with our theoretical framework suggesting that the higher the number of existing reviews, the lower the reviewing effort.
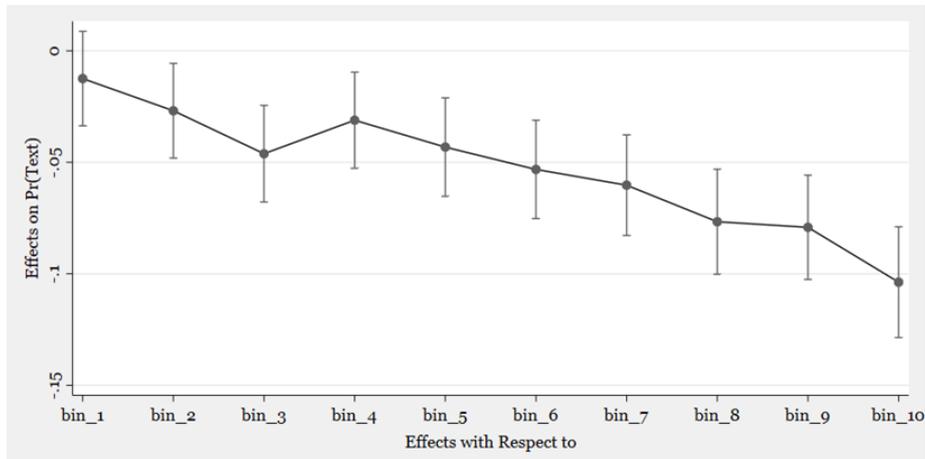
Figure 1. Average marginal effects with 95% CIs.

For review length, Columns (iii) and (iv) of Table 3 are relevant. Similar as for a textual review, we observe a statistically significant negative relationship between review length and the number of existing reviews. The use of log-transformed data allows us to interpret the coefficients easily: A 1% increase in the number of existing reviews is

associated with a decrease of review length by 6.7% (see Column (iii)). Controlling for seasonal effects by adding month dummies does not change the results qualitatively meaning that we find support for H2 as well. The coefficients of the control variables, however, provide some interesting findings: While both, the submission of a picture and user experience, is positively related to review length, the submission of a 5-star rating is associated with a decrease in review length by 15.6% compared to the base level of a 3-star rating. Contrarily, the submission of a 2-star rating is related to an increase of 20% compared to the base level. The insignificant coefficient of a 1-star rating might be related to the low number of 1-star ratings submitted in our dataset (only 1.4% of all ratings are 1-star ratings). To account for this issue, we generate a less granular rating variable which ranges from 1 to 3, where 1 includes 1- and 2-star ratings, 2 includes all 3-star ratings and 3 includes 4- and 5-star ratings. While the result for the number of reviews does not change, the revised rating variable exhibits the expected coefficients: rating category 1 is significantly positive and rating category 3 is significantly negative related to review length (results not tabulated).

Table 3. Results for H1 and H2.

Dependent variable:	(i) Text (0/1)	(ii) Text (0/1)	(iii) ln(Length)	(iv) ln(Length)
ln(Nr. of reviews)	-0.123***	-0.215***	-0.0671***	-0.0829***
	(0.0183)	(0.0238)	(0.0114)	(0.0143)
Picture	0.549***	0.514***	0.284***	0.279***
	(0.112)	(0.113)	(0.0542)	(0.0543)
ln(User experience)	0.793***	0.800***	0.0512***	0.0526***
	(0.00950)	(0.00967)	(0.00580)	(0.00587)
1-star rating	0.880***	0.870***	0.120	0.119
	(0.131)	(0.131)	(0.0800)	(0.0801)
2-star rating	0.530***	0.533***	0.200***	0.202***
	(0.116)	(0.116)	(0.0719)	(0.0719)
4-star rating	0.0623	0.0622	-0.0236	-0.0226
	(0.0440)	(0.0441)	(0.0305)	(0.0306)
5-star rating	0.431***	0.427***	-0.156***	-0.157***
	(0.0410)	(0.0411)	(0.0284)	(0.0285)
Constant	-1.953***	-1.622***	3.997***	-0.0829***
	(0.137)	(0.159)	(0.0868)	(0.0143)
Month dummies	NO	YES	NO	YES
R-squared	0.21	0.21	0.02	0.02
Observations	37,370	37,370	17,058	17,058
Mean VIF	4.06	4.58	1.62	2.17

Notes: We use the centered 3-star rating is the base of the (indicator variable) individual rating. Column (i) and (ii) represent the logit model in Equation (1) above. The goodness of fit measure is therefore pseudo-R-squared. Column (iii) and (iv) represent the OLS regression in Equation (2). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The results of our analysis for H3 and H4 are depicted in Table 4. The first two columns indicate that users with a bad experience (i.e., 1-star rating or 2-star rating) are more likely to write a textual review if the existing review rating distribution is extremely positive. The difference is, based on a one-sided t-test, statistically significant and partially supports our hypothesis H3 for the case of negative disconfirmation An extremely positive existing rating distribution increases the likelihood to write a textual review by nearly 10 percentage points in the case of a 1-star rating. For review length (i.e., H4), however, the difference between the extremely positive rating distribution and the less extreme rating distribution is not statistically significant. The length of 1-star ratings and 2-star ratings in the case of an extremely positive rating distribution are not statistically significant different to the group of reviews which include a less extreme existing rating distribution.

Table 4. Results for H3 and H4.

	Textual reviews (0/1)		**Length (in characters)**	
	1-star rating	2-star rating	1-star rating	2-star rating
Reviews with an extremely positive existing rating distribution	34,57% (n=188)	41,94% (n=155)	67 (n=65)	61 (n=65)
Reviews with a less extreme existing rating distribution	25,37% (n=406)	35,82% (n=455)	75 (n=103)	90 (n=163)
t-statistic	2,32**	1,36*	0,02	-1,12

*Notes: For all columns, a one-sided t-test is applied. For testing the statistical difference of review length, the variable was log-transformed. ** p<0.05, * p<0.1.*

5   Discussion and Conclusion

This study emphasizes the importance of existing reviews for future reviewing effort. First, we observe that the sole number of existing reviews matters for reviewing effort: A high number of existing reviews is associated with both, a lower number of additional textual reviews (H1) and a shorter length of textual reviews (H2). Thus, both hypotheses, which are developed from the collective effort model by Karau and Williams [16], can be confirmed. If a potential reviewer observes that there exists already a large number of reviews, her individual contribution will not make a big impact on the collective task of reviewing. Consequently, her reviewing effort will be rather low. On the contrary, if the existing number of reviews is low, the individual review is much more visible and impactful which makes the reviewer to put more effort into the task.

Second, our study highlights the importance of the rating distribution of existing reviews as well: If the rating distribution of existing reviews differs strongly to the own experience (measured by the individual rating), the high disconfirmation degree results in a higher propensity to additionally write a textual review. While this partially supports our hypothesis H3 developed from the expectation disconfirmation theory, we do not find support for hypothesis H4 – the actual review length is independent from the disconfirmation degree. Notably, we examined the content of reviews with a high disconfirmation degree and, based on anecdotical evidence, their topics do not differ from the content of reviews with a lower disconfirmation degree. We do not find any evidence that these reviews argue against the other existing reviews and their ratings. Thus, it seems that users with a high disconfirmation degree are more likely to additionally write a textual review but when they write the actual review, the content seems very similar to other reviews with the same rating. In fact, reviews with a low rating typically focus on outlining the own (bad) experience.

5.1   Theoretical Implications

From a theoretical perspective, our results contribute to understand the determinants of reviewing effort invested by the users. So far – to the best of our knowledge – no study has analyzed the relationship between the existing reviews and reviewing effort. Existing studies that use existing reviews as an independent variable focus on the propensity to submit a review at all or on the existing ratings. We provide additional evidence that reviewing can be seen as a collective task and the underlying cognitive mechanisms that individuals undergo when performing this task are accurately described by the collective effort model. In fact, we observe the phenomenon of social loafing in our online review setting as well. Further, our results indicate that the number of existing reviews as a typical metric in review systems does not serve as a social norm and thus social comparison theory is not adequately describing the underlying psychological influence of the existing reviews on reviewing effort (see [17], for a discussion). Similarly, our results also support the hypothesis by Hennig-Thurau et al. [15], that, among other motives, users submit a review to vent negative feelings: The expectation disconfirmation theory by Oliver [21], states that a high disconfirmation degree results in an increase level of dissatisfaction and this, in turn, makes users to restore the balance of the existing review ratings [22], [23]. Thus, our findings indicate that observing the number and the rating distribution of existing reviews triggers multiple cognitive mechanisms.

5.2   Practical Implications

The practical implications of our study are that review system designs need not only to consider consumers (who are observing the existing reviews) but also potential reviewers. Thus, while providing detailed information about existing reviews is helpful for consumers, it might also bring potential reviewers to invest less effort. Thus, review system designers face an important trade-off because helpful reviews mainly depend on the effort invested. They might incorporate adjusted design features like, for

instance, omitting information about the number of existing reviews or displaying only a few highlighted reviews at the first page to increase the users' feeling that they are might a significant contribution to the collective task of reviewing. With such a feature, the own review might be perceived as impactful and more discoverable. Similarly, review system designers could also segment users into subgroups so that each user only sees the existing reviews from other users of this subgroup, which is based on, e.g., age, language or purpose of purchase. Regarding the influence of the existing review rating distribution, review system designers could incorporate a design feature which highlights the existing rating distribution for early adopters of products to increase the disconfirmation experienced early on in a product lifecycle. Thus, the better a potential disconfirmation is observed, the more likely users will increase their reviewing effort.

5.3   Limitations and Future Research

Our research has some limitations which, however, offer fruitful possibilities for future research. First, the analysis is based on a dataset that consists of sights across Europe on Google Maps. It might be possible that the observed effects of the number of existing reviews and rating distribution are not generalizable across both, review platforms and review objects. Thus, future research could explore whether our findings can be confirmed, for instance, with other review objects on different review platforms. Furthermore, for sights that have no existing reviews yet, the underlying cognitive mechanism might again be different and could be addressed by future research.

Second, due to the nature of our data collection method, some contextual factors such as the average overall rating at the time, the most recent reviews or the top picture of the sight, might be missing. Thus, conducting a similar analysis with a different dataset may be appropriate.

Third, the rating distribution in our dataset is very positively skewed. This means that we could only test the effects of disconfirmation in one direction. Additionally, the ratings of users could be influenced by these positively skewed existing ratings and thus create significant herding effects. For future research, it is important to analyze whether the effects of disconfirmation influence the reviewing effort in both ways and whether herding effects are present.

Lastly, our findings suggest that the applied theories are accurately describing the underlying cognitive mechanisms that users undergo when reviewing an object and perceiving the existing reviews. Nonetheless, further research might solidify our insights regarding the applicability of these theories to this context. For example, experimental studies could be conducted to isolate the observed effects of the existing reviews on reviewing effort.

References

1. ChannelAdvisor. Consumer Survey: Global Consumer Shopping Habits, https://www.channeladvisor.com/blog/industry-trends/2011-global-consumer-shopping-habits-survey/ (Accessed: 27.04.2020).

2. Nielson. Word-of-mouth recommendations remain the most credible, https://www.nielsen.com/id/en/press-releases/2015/word-of-mouth-recommendations-remain-the-most-credible/ (Accessed: 27.04.2020).

3. Hong, Y., and Pavlou, P. A.: Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents. Information Systems Research. 25:2, 328–344 (2014)

4. Yu, X., Liu, Y., Huang, X., and An, A.: Mining online reviews for predicting sales performance: A case study in the movie domain, IEEE Transactions on Knowledge and Data engineering. 24:4, 720-734 (2010)

5. Sahoo, N., Dellarocas, C., and Srinivasan, S.: The impact of online product reviews on product returns. Information Systems Research. 29:3, 723-738 (2018)

6. Hu, N., Zhang, J., and Pavlou, P. A.: Overcoming the J-shaped distribution of product reviews. In: Communications of the ACM, pp. 144-147 (2009)

7. Mudambi, S., and Schuff, D.: What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Quarterly. 34:1, 185–200 (2010)

8. Askalidis, G., Kim, S. J., and Malthouse, E. C.: Understanding and Overcoming Biases in Online Review Systems. Decision Support Systems. 97:5, 23–30 (2017)

9. Wang, J., Ghoang, J., Ghose, A., and Ipeirotis, P.: Bonus, disclosure, and choice: What motivates the creation of high-quality paid reviews? In: Proceedings of the International Conference on Information Systems (ICIS 2012), Association for Information Systems (AIS), Atlanta. (2012)

10. Cao, Q., Duan, W., and Gan, Q.: Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach. Decision Support Systems. 50:2, 511–521 (2011)

11. Dellarocas, C., Gao, G., and Narayan, R.: Are consumers more likely to contribute online reviews for hit or niche products? Journal of Management Information Systems. 27:2, 127-158 (2010)

12. Muchnik, L., Aral, S., and Taylor, S. J.: Social influence bias: A randomized experiment. Science. 341:6146, 647-651 (2013)

13. Ho, Y. C., Wu, J., and Tan, Y.: Disconfirmation effect on online rating behavior: A structural model. Information Systems Research. 28:3, 626-642 (2017)

14. Balasubramanian, S., and Mahajan, V.: The Economic Leverage of the Virtual Community. International Journal of Electronic Commerce. 5 (Spring), 103–138 (2001)

15. Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D.: 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 (2004)

16. Karau, S. J., and Williams, K. D.: Understanding individual motivation in groups: The collective effort model. In: Groups at work: Theory and research. Ed. by M. E. Turner. Mahwah, NJ: Erlbaum, pp. 113–141 (2001)

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

18. Karau, S. J., and Williams, K. D.: Social loafing: A meta-analytic review and theoretical integration. Journal of personality and social psychology. 65:4, 681 (1993)

19. Harkins, S. G.: Social loafing and social facilitation. Journal of Experimental Social Psychology. 23, 1–18. (1987)

20. Bhattacherjee, A.: Understanding information systems continuance: an expectation-confirmation model. MIS quarterly, 351-370. (2001)

21. Oliver, R. L.: Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. Journal of applied psychology. 62:4, 480 (1977)

22. Heider, F.: Attitudes and Cognitive Organization. Journal of Psychology. 21, 107–112. (1946)
23. Newcomb, T.M.: An Approach to the Study of Communicative Acts. Psychological Review. 60, 393–404 (1953)