

2013

What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions

Michael Siering

Goethe University Frankfurt, Germany, siering@wiwi.uni-frankfurt.de

Jan Muntermann

University of Göttingen, Germany, muntermann@wiwi.uni-goettingen.de

Follow this and additional works at: <http://aisel.aisnet.org/wi2013>

Recommended Citation

Siering, Michael and Muntermann, Jan, "What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions" (2013). *Wirtschaftsinformatik Proceedings 2013*. 7.

<http://aisel.aisnet.org/wi2013/7>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2013 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions

Michael Siering¹ and Jan Muntermann²

¹ Goethe University Frankfurt, Germany
siering@wiwi.uni-frankfurt.de

² University of Göttingen, Germany
muntermann@wiwi.uni-goettingen.de

Abstract. For consumers, online product reviews have become an important source for product-related information. Furthermore, they represent a beneficial addition to online retailers' websites. Due to the increasing amount of available product reviews, identifying the most helpful product reviews represents an important task in order to reduce information overload. Therefore, the factors influencing review helpfulness have to be identified. Thus, in order to explain review helpfulness, we build upon and extend review diagnosticity theory with concepts from marketing research and propose a research model that includes product quality, review sentiment and review uncertainty. Based on a sample of amazon.com product reviews, we evaluate our research model and find that statements about product quality positively influence review helpfulness. Furthermore, we identify that sentiment as well as uncertainty expressed in product reviews have an impact on review helpfulness. Finally, we confirm that the product category has a moderating effect on these relationships.

Keywords: Product Review, Review Diagnosticity, Helpfulness, Content Analysis

1 Introduction

In recent years, online product reviews have become important both for online consumers as well as online retailers. On the one hand, online product reviews represent an important source of information to consumers who base their purchase decisions on reviews provided by websites like amazon.com [1]. On the other hand, online product reviews are an important asset for online retailers since they attract online consumers who may also buy the related products [2].

Due to the enormous amount of reviews available for several products, customers may suffer from information overload, which reduces the beneficial aspects of product reviews [3], [4]. As a consequence, online retailers often display the most helpful product reviews first. In order to determine how helpful a product review is and to rank reviews according to their helpfulness, many online retailers offer their customers the possibility to vote whether a certain review is perceived to be helpful or not.

However, reviews receive votes over a long time period which hampers the ranking of recent ones. Furthermore, ranking of reviews that have not received any vote is not possible at all [3]. As follows, an appropriate understanding of what makes a review helpful is required for online retailers to assess the helpfulness of reviews that have not been evaluated by online consumers yet.

Next to the impact of online product reviews on sales that has been a field of research for several years [1], [5], [6], recent studies already provide first insights into the characteristics which determine online review helpfulness. Since Mudambi and Schuff (2010) provided evidence that review extremity, review depth and product type have an impact on review helpfulness [2], further studies have explored related characteristics such as specific emotions [4], [7] and review readability [3], [8]. However, previous studies neglect a main factor determining the readers' purchase decision and thus presumably affecting review helpfulness: product quality [9], [10]. As a consequence, it can be assumed that product reviews focusing on product quality will also provide diagnostic value and consequently, contribute to review helpfulness.

Furthermore, previous research does not provide deeper insights into the writing style of the product reviews under investigation since sentiment is often measured by the star rating rather at a textual level [7] or the focus lies on manual analyses of small sample sizes [11]. As already investigated within marketing research, consumers' purchase decisions are influenced by sentiment and the level of certainty expressed within advertising campaigns [12–14]. Consequently, we also investigate whether review sentiment and review uncertainty influence review helpfulness.

Overall, recent studies often neglect the differentiation between two important product categories: search goods and experience goods [15], [16]. Search goods are characterized as products for which information about their quality can be easily obtained before purchasing the product. In contrast, experience goods are defined as goods that often require purchase to evaluate product quality [2]. Thus, these product categories are supposed to influence the information required by consumers and consequently, may also impact perceived review helpfulness [2].

Consequently, our research aims at contributing to review diagnosticity theory by means of providing a better understanding of what contributes to the helpfulness of product reviews in the purchase decision process. Therefore, we build upon the initial model of review helpfulness by Mudambi and Schuff (2010) [2]. Our model extends the original factors *review extremity*, *review depth* and *product type* in order to incorporate product-related and stylistic aspects in the form of *product quality*, *review sentiment* and *review uncertainty*. To empirically validate our model, we acquire product reviews related to the 20 most popular products of 6 different amazon.com product categories. Thereafter, we analyze the reviews related to the variables under consideration.

The remainder of this paper is structured as follows. Section 2 presents the background and the research model of our study, including our research hypotheses and the rationale behind them. Subsequently, section 3 provides an outline of our research methodology. The empirical results are presented in section 4. Finally, section 5 summarizes our results and concludes.

2 Background and Research Model

Our study aims to extend the existent research on review helpfulness by laying a focus on product-specific and stylistic aspects as actually expressed within product reviews. Building upon marketing research, we identify product quality, review sentiment and review uncertainty as important factors influencing customers' purchase decisions [9], [17], [18]. Accordingly, these factors may also affect the helpfulness of product reviews that are used to provide information during the purchase decision process [2]. Therefore, our research model includes these factors as independent variables with a hypothesized relationship to review helpfulness (H1a – H3a). Furthermore, we also hypothesize a moderating effect of product type, i.e. search good or experience good (H1b – H3b). To improve the robustness of our results and to be able to compare our content-related factors with existent research, our research model (Fig.) also includes the basic factors proposed by Mudambi and Schuff (2010) [2]. Therefore, review extremity and review depth are treated as control variables to check whether our analysis confirms the original findings. In the following sections, we elaborate on our research model.

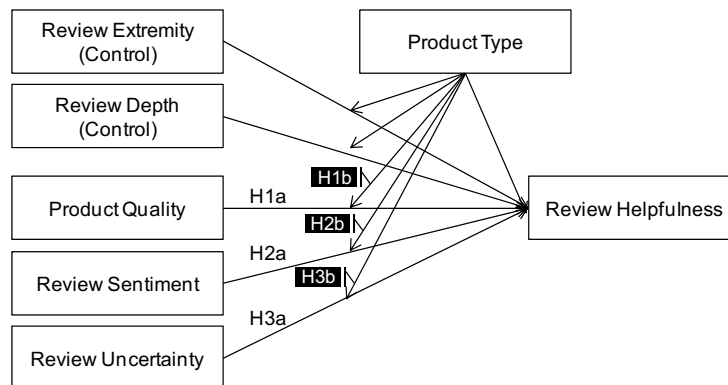


Fig. 1. Research Model on Review Helpfulness

2.1 Review Diagnosticity, Review Extremity and Review Depth

Review diagnosticity theory is closely related to the notion of information diagnosticity, which encompasses the question of whether a certain piece of information is helpful during the decision-making process or not [4], [19]. Analogously, review diagnosticity theory explains which factors make a review helpful during the different phases of a consumers' purchase decision-making process [2]. Thereby, online consumers can use online product reviews in order to evaluate a product and the available alternatives [1], [2].

Within the basic model of review diagnosticity, Mudambi and Schuff (2010) consider review depth and review extremity as relevant factors determining review helpfulness which is seen as a reflection of review diagnosticity [2]. While extremity ad-

dresses “whether the review is positive, negative, or neutral”, review depth has been defined as “the extensiveness of the reviewer comment” [2].

For product reviews that usually provide a one (worst) to five (best) star rating, review extremity “is the extent to which an individual's attitude deviates from the midpoint” [20]. In the literature, different studies have investigated the effect of review extremity, whereas mixed results have been found. On the one hand, extreme book reviews have been found to be more helpful compared to moderate reviews [5] and extreme evaluations have been shown to increase the likelihood of acceptance of advices given [21]. On the other hand, it is shown that extreme positions can negatively impact perceived source competence [22]. These presumably contradictory results guided Mudambi and Schuff (2010) to consider product type (i.e. experience vs. search good) as a moderating factor because the impact of review extremity could depend on the product type explored [2]. Based on their argument that, for experience goods, objective content should be preferred over extreme content, they hypothesize that for experience goods, extreme reviews are less helpful than moderate ones. To cover this aspect in our model, and to be able to compare the factors of the original model with the factors of our extended model, we include review extremity as well as the moderating effect of product type as control variables.

Furthermore, review depth represents another factor related to review helpfulness [2]. High review depth increases the amount of information available to the consumer which helps in the process of making a purchase decision. Consequently, increased review helpfulness has been found for reviews providing more detailed information [23] since increased review depth contributes to an increased amount of information that is available to consumers without any additional search costs [2]. Moreover, the effect of review depth on review helpfulness has also been argued to depend on product type. Following Mudambi and Schuff (2010), consumers of search goods are more likely to be interested in objective attributes and features, which are delivered by reviews providing more detailed information. Consequently, their hypothesis suggests that review depth in general has a positive effect on a review's helpfulness, which will be even more significant for search goods [2]. To control for this aspect, we also include review depth and the moderating effect of product type as control variables in our extended model.

2.2 Product Quality

Product quality can be defined as the “consumer's judgment about a product's overall excellence or superiority” [9] and covers the aspect of whether a product review provides detailed information on a products' core characteristics. In contrast to review depth, which is usually assessed on the basis of review word count [2], [7], product quality accounts for the level of detail with regard to relevant product performance characteristics, i.e. how much information on relevant quality characteristics is included within the review.

Product quality plays an important role for the product choice of consumers, but as compared to price information, product performance characteristics are much more difficult to obtain [15]. Furthermore, information on product or service quality is

highly relevant related to consumer attitudes and intentions [10], [24]. We therefore hypothesize: *More detailed statements on product quality have a positive effect on review helpfulness (H1a).*

As product features and characteristics of search goods, including product quality, can be evaluated more easily before purchase compared to experience goods [15], [25], it can be assumed that information related to product quality is more valuable to consumers when it is hard to obtain. As a consequence, we hypothesize: *The product type moderates the effect of statements related to product quality on review helpfulness, and the effect is greater for experience goods than for search goods (H1b).*

2.3 Review Sentiment

Sentiment represents an “attitude, thought, or judgment prompted by feeling” [26]. In the context of product reviews, sentiment covers the emotional statements expressed within the text. Thereby, review sentiment differs from other review aspects such as the star rating. Although the star rating is often used for sentiment classification [27], it rather focuses on the overall product evaluation than on the language used within the review. In recent years, the impact of sentiment expressed in textual sources on purchase decisions has been confirmed [28].

From a theoretical perspective, the potential relevance of review sentiment can be derived from the emotional value perceived by a customer related to the product. The emotional value affects the purchase decision and can be defined as “the utility derived from the feelings or affective states that a product generates” [17]. As follows, the more positive the review sentiment is, the higher may be the positive impact on the emotional value of the product and thus the review could be more helpful to consumers since it signals that the product is worth to consider [11]. In consequence, positive (negative) online review sentiment is supposed to lead to higher (lower) purchase intentions [29]. In this context, it can be assumed that an increased emotional value also leads to an increased satisfaction with the product review and thus increases review helpfulness. We therefore hypothesize: *More positive review sentiment has a positive effect on review helpfulness (H2a).*

However, to a certain extent, even negative sentiment can be helpful to consumers. For experience goods (first-run films in the UK), critical reviews have been found to have a significant effect on consumers as they positively affect revenues [30]. Thus, unfavorable product information may play a more important role for consumer behavior compared to favorable product information [31]. This phenomenon is widely discussed as negativity bias, after which, and compared to positive attitude, a negative attitude does have a stronger effect on behavior [32], [33]. As already discussed, experience goods are harder to evaluate before purchase than search goods. In this context, negative sentiment can be perceived to be more accurate and helpful [7]. As follows, online consumers will rather value negative sentiment representing other customers’ perceptions. We therefore hypothesize that product type has a moderating effect on the relationship between review sentiment and review helpfulness: *The product type moderates the effect of review sentiment on review helpfulness, and for experience goods, negative sentiment becomes more valuable (H2b).*

2.4 Review Uncertainty

In general, (un-)certainty can be defined as “the degree to which an individual is [not] confident that his or her attitude toward an object is correct” [20]. More specific, review uncertainty addresses a reviews’ ability to provide a clear judgment or evaluation of a certain products’ attributes that are relevant for the customers’ purchase decision [18]. An impact of higher certainty on perceptions of others has been found to be significant in the literature. As illustrated by Marks and Miller (1985), certainty about one’s opinion correctness positively affects its projection [12]. Further, and in the context of buyer behavior theory, having an overall confidence in a product brand positively affects the corresponding purchase decision [13]. In the case of a product review mainly conveying high uncertainty, it can be assumed that the product review as well as the information within the product review is perceived to be less helpful for the following purchase decision. We therefore hypothesize: *Review uncertainty has a negative effect on review helpfulness (H3a).*

With regard to search goods, consumers are more interested in specific and objective information on actual product characteristics [2]. In contrast, the evaluation of experience goods is more subjective and depends on each reviewer’s personal perceptions [15]. Thus, in the case of conflicting perceptions and attitudes between reviewers and online consumers, it can be assumed that product reviews which are connected with uncertainty are perceived to be less provoking. Consequently, it can be assumed that the general negative impact of review uncertainty is reduced for experience goods. As follows, we expect a moderating effect of the product type and hypothesize: *The product type moderates the effect of review uncertainty on review helpfulness, and the effect is greater for search goods than for experience goods (H3b).*

3 Research Methodology

3.1 Dataset Acquisition

To evaluate our research model empirically, we focus on product reviews that have been published on amazon.com. Thereby, we select reviews related to search and experience goods. For that purpose, we focus on the six product categories as applied by Mudambi and Schuff (2010) [2] and select the 20 best selling products for each category. A detailed definition of the different product categories covered can be found in Table 1. For each product, we acquire the corresponding product reviews. Therefore, we download review text, star rating, the number of people rating the review as helpful and the total number of people rating the review.

Table 1. Product categories taken into account within this study as defined by [2]

Product Type	Product Categories
Search Good	Camera & Photo, Computer Printers, Cordless Telephones
Experience Good	MP3 Players, Music, PC-compatible Games

Overall, we focus on product reviews that received at least ten votes in total, which ensures reliability of results: In the case of a small amount of votes, a small number of online customers could have a large influence on the helpfulness score: in the case of one total vote, this vote determines whether the helpfulness score is 0% (not helpful) or 100% (which indicates that the review is very helpful).

3.2 Variable Operationalization

In order to be able to test our research hypotheses, we operationalized the different variables of interest by means of content analysis and direct extraction from the product review. Content analysis encompasses the process of “making inferences from a symbolic medium, usually texts” by classifying “textual material, reducing it to more relevant, manageable bits of data” [34]. Thereby, content analysis can consist of different techniques [35] and is often applied within psychology to make inferences about the writer of a message or about the communication between different individuals [36]. Overall, this process shall be conducted as objective as possible [37].

In general, different forms of content analysis exist [36]: On the one hand, manual coding of documents is used. On the other hand, different automated approaches for content analysis are available. In comparison to manual coding, no problems with inter-coder reliability prevail in the case of automated content analysis because this methodology is dictionary-based and can be repeated without a loss in quality [34]. Furthermore, automated content analysis has been proven to be reliable [35], [36] and, compared to manual coding, less time consuming [36]. Finally, since the dictionaries used are oftentimes publicly available, automated coding of documents is transparent and the results can be reproduced easily [38].

In the course of automated content analysis, dictionaries are used in order to map different words of a text to several pre-defined categories representing psychological concepts. As a next step, the frequencies of how often a certain category prevails within a document can be used for further analyses [34]. In this context, several dictionaries have been developed and evaluated that provide measurements for several psychological constructs, whereas in our study, we make use of the dictionary that is used within the General Inquirer (GI) [37], [39]. The GI represents a text analysis framework that has already been applied in several studies (see [37]). Building upon such well-established dictionaries is advantageous due to several reasons, including standardized classifications and the dictionary’s extensive previous validation [40].

We conduct an automated content analysis of the product reviews included in our dataset. To measure product quality, we determine the amount of terms related to the GI category “Quality” in relation to the total number of terms within the document (exemplary terms in this category are “secure”, “stable” or “weak”). In the case of review sentiment, we apply a sentiment polarity measure that takes into account the GI categories “Positiv” and “Negativ” (exemplary terms are “great” or “unhappy”). Finally, review uncertainty is determined on behalf of the “If” word list which covers feelings of uncertainty (exemplary terms are “almost” or “may”). Table provides an overview about the independent (IV) and dependent variables (DV) used within this study.

Table 2. Operationalization of Independent (IV) and Dependent Variables (DV)

Variable Type	Research Hypothesis	Variable	Operationalization
IV	H1: Product Quality	Quality	Ratio based on GI-Category “Quality” related to the total number of words.
	H2: Review Sentiment	Polarity	Ratio based on GI-Categories “Positiv” and “Negativ”: (Positiv - Negativ) / (Positiv + Negativ).
	H3: Review Uncertainty	Uncertainty	Ratio based on GI-Category “IF” related to the total number of words.
	Control Variables	ProductType	1 if experience good; 0 if search good.
		Extremity	Star Rating - Mean Rating.
Depth		Number of Words.	
	TotalVotes	Number of Votes.	
DV		Helpfulness	Number of Helpful Votes divided by TotalVotes.

Next to these variables that are measured by means of automated content analysis, we also extract several control variables from the review pages and operationalize these variables as defined in [2]: We include a dummy variable representing the product type, i.e. the question whether the product is a search or an experience good. Furthermore, we measure review depth as the number of words the review consists of. TotalVotes is measured as the number of people who participated at the voting, i.e. who voted the review to be either helpful or not helpful.

In the case of review extremity, Mudambi and Schuff (2010) only use the star rating as well as the squared star rating for operationalization [2]. However, since our study covers different products per category having different average ratings, we operationalize review extremity by the absolute difference between the star rating of the review and the average star rating of the product and are thus in line with the definition by [20]. This ensures that extremity measures how extreme a review is related to the average review concerning a certain product.

Finally, review helpfulness is measured as the number of helpful votes divided by the number of total votes. Due to the fact that we only consider reviews with at least ten votes, this measure can always be calculated and is not as volatile as in the case of a small number of votes taken into account.

3.3 Regression Analysis

To investigate the impact of the independent variables on review helpfulness, we follow previous research and use Tobit regression [2], [4], [8]. In comparison to OLS regression, Tobit regression is appropriate because it covers the following aspects. At first, the dependent variable is censored [2], which means that it has lower and upper bounds: if nobody considers a review as helpful, review helpfulness is zero and cannot be below this level. In contrast, if everybody considers the review as helpful, help-

fulness is 100%. At second, there exists a selection problem concerning the customers participating in the voting since not every reader also evaluates the reviews' helpfulness [2]. Equation 1 shows the regression run within our study. Thereby, we include the different hypothesized independent variables including the moderation effects as well as the control variables (ProductType, Extremity, Depth, TotalVotes) and, as proposed by Mudambi and Schuff (2010), the corresponding interaction effects (Extremity x ProductType, Depth x ProductType) [2].

$$\begin{aligned}
\text{Helpfulness} = & \text{Constant} + \beta_1\text{Quality} + \beta_2\text{Polarity} + \beta_3\text{Uncertainty} \\
& + \beta_4(\text{Quality} \times \text{ProductType}) \\
& + \beta_5(\text{Polarity} \times \text{ProductType}) \\
& + \beta_6(\text{Uncertainty} \times \text{ProductType}) + \beta_7\text{Controls} + \varepsilon
\end{aligned} \tag{1}$$

4 Empirical Study

4.1 Descriptive Statistics

In total, our dataset consists of 4,970 product reviews. As can be seen from Table 3, 1,245 product reviews deal with search goods and the remaining 3,725 product reviews deal with experience goods.

In order to provide first insights into the difference between search and experience goods, we test whether both product types differ. Therefore, the Wilcoxon-signed-rank test for equality of the different variables' medians is applied. As Table shows, product reviews related to search goods contain a significantly larger amount of statements related to product quality. Additionally, reviews related to experience goods also contain more uncertain statements as compared to search goods. This result can be explained by the fact that the evaluation of experience goods is oftentimes more subjective and thus, reviewers are less certain as compared to objective criteria. In the case of sentiment polarity, no significant difference can be detected. Related to the control variables, we confirm that the reviews related to search goods are perceived to be more helpful than reviews related to experience goods [2].

In order to ensure that our regression setup leads to satisfactory results, we check the corresponding variable correlations in order to avoid multicollinearity. As Table shows, the correlations between the different independent variables are very low. It has to be noted that there is a negative correlation between review helpfulness and review extremity: if the star rating of the review deviates from the average star rating, review helpfulness decreases. Additionally, these results provide evidence that the different categories of the General Inquirer used in this study are almost independent, i.e. terms contained in one word list are only rarely contained in another word list.

Table 3. Descriptive Statistics

Variable	Full Sample Mean (SD)	Search Goods Mean (SD)	Experience Goods Mean (SD)	p-value
Quality	0.0178 (0.0190)	0.0241 (0.0230)	0.0157 (0.0170)	0.0000
Polarity	0.2429 (0.3486)	0.2467 (0.3642)	0.2416 (0.3432)	0.4986
Uncertainty	0.0180 (0.0170)	0.0163 (0.0183)	0.0185 (0.0164)	0.0000
Extremity	1.5868 (0.9257)	1.4595 (1.0759)	1.6294 (0.8658)	0.0005
Depth	245.1535 (316.7760)	299.0233 (339.5894)	227.1487 (306.7206)	0.0000
TotalVotes	41.3388 (140.2850)	53.2185 (146.8246)	37.3683 (137.8224)	0.0001
Helpfulness	0.5989 (0.2779)	0.7443 (0.2884)	0.5503 (0.2565)	0.0000
n	4970	1245	3725	

(p-value indicates the results of a Wilcoxon rank-sum test for equality of medians related to search and experience goods)

Table 4. Variable Correlations

Variable	1	2	3	4	5	6	7
1 Quality	1						
2 Polarity	-0.03	1					
3 Uncertainty	-0.01	0.01	1				
4 Extremity	-0.05	-0.11	0.02	1			
5 Depth	-0.03	0.00	0.04	-0.21	1		
6 TotalVotes	0.02	0.01	0.02	-0.07	0.20	1	
7 Helpfulness	0.09	0.01	-0.05	-0.53	0.25	0.13	1

4.2 Empirical Results

Table shows the regression estimates of our research model. Related to the impact of statements related to product quality on review helpfulness, we confirm H1a at a 10% level of significance. As follows, if reviews focus on product quality, review helpfulness increases. However, the significance level of Quality is lower than the significance levels of other explanatory variables. This indicates that statements related to product quality influence review helpfulness, but the influence of other variables is more evident. In contrast, H1b has to be rejected since the moderating effect Quality x ProductType is not significant. Thus, product type does not moderate the positive relationship between product quality and review helpfulness.

Related to H2a, we find that positive review sentiment polarity increases review helpfulness, which is significant at a 1% level. Related to H2b, we confirm the negativity bias for experience goods: the moderating effect (Polarity x ProductType) is negative and significant at a 1% level of significance. If the regression coefficients are taken into account, it can be noted that for reviews with negative sentiment polarity, a positive impact on review helpfulness can be measured since the coefficient for the moderating effect has an absolute value that is larger than the coefficient of polarity. In contrast, positive sentiment polarity decreases review helpfulness because negative sentiment is perceived to be more accurate and helpful [7].

In order to test H3a, we consider whether statements related to uncertainty have a negative impact on review helpfulness. As can be seen from the negative coefficient that is significant at a 1% level of significance, H3a can be confirmed. Consequently, if a reviewer is not convinced of his review and uses statements expressing uncertainty, the review is perceived as less helpful. However, related to H3b, it can be noted that for experience goods, uncertain statements have a minor impact on review helpfulness: the moderating effect is positive and significant at a 5% level of significance. As a result, H3b can be accepted.

Finally, related to the control variables, we confirm that reviews related to search goods are generally perceived to be more helpful than reviews related to experience goods. This is evidenced by the significant negative impact of the ProductType dummy variable. Furthermore, product reviews that are connected with a star rating having a large deviation from the average star rating are also perceived to be less helpful. This effect is significant at a 1% level of significance. Additionally, the variables controlling for review depth and total votes as well as the corresponding moderating effects have a significant impact on review helpfulness.

Table 5. Tobit-Regression Estimates Explaining Review Helpfulness (n=4970, $p > \chi^2 = 0.000$)

Variable	Coefficient	Standard Error	t-value	p-value	
Constant	0.8763	0.0198	44.24	0.000	***
Quality	0.4774	0.2897	1.65	0.099	*
Polarity	0.0910	0.0197	4.63	0.000	***
Uncertainty	-1.2251	0.3846	-3.19	0.001	***
Quality x ProductType	-0.3463	0.3660	-0.95	0.344	
Polarity x ProductType	-0.1621	0.0226	-7.18	0.000	***
Uncertainty x ProductType	0.9658	0.4490	2.15	0.032	**
ProductType	-0.0727	0.0226	-3.22	0.001	***
Extremity	-0.1227	0.0069	-17.82	0.000	***
Depth	0.0002	0.0000	7.85	0.000	***
TotalVotes	0.0001	0.0000	3.85	0.000	***
Extremity x ProductType	-0.0356	0.0082	-4.34	0.000	***
Depth x ProductType	-0.0001	0.0000	-2.95	0.003	***

* / ** / *** indicate significance at a 10% / 5% / 1% level.

4.3 Discussion

Our study shows that, next to general text characteristics such as review length, a more fine-grained textual analysis enhances the understanding of review helpfulness. Thereby, the hypothesized impact of textual aspects related to product quality, review sentiment and review uncertainty on the perceived helpfulness of online product reviews is confirmed.

Furthermore, our results confirm and extend the knowledge about the impact of the product type on review helpfulness. In line with previous research, we find that reviews related to search goods have, compared to experience goods, a higher helpfulness rating [2]. In this context, we also show that variables can have a reversed impact on review helpfulness for different product categories: as positive review sentiment increases the helpfulness of product reviews dealing with search goods, negative review sentiment increases the helpfulness of reviews about experience goods.

Taking into account the control variables, we confirm previous results in terms of the positive impact of review depth and the negative impact of review extremity. In contrast to Mudambi and Schuff (2010), who find a negative impact of the number of total votes on review helpfulness [2], we measure a small positive impact. This result can be explained by the fact that we only take into account reviews that received at least 10 votings in order to improve the reliability of our results.

In line with Mudambi and Schuff (2010), we are aware of one limitation that is caused by the methodology used to assess review helpfulness [2]. Since review helpfulness is measured on the basis of the votes of the online retailers' clients who participate in the voting, the results may not fully cover the perceptions of clients who do not participate. From a methodological point of view, automated content analysis bears the limitation that the results depend on the dictionary used. If a term that characterizes a certain category is not included in the dictionary, the results may be biased [36]. However, we address this issue by applying standardized and well-established dictionaries as delivered by the General Inquirer [37], [39]. Furthermore, especially in the context of sentiment analysis, an approach that is based on term frequencies does not cover complex language constructs such as irony [41]. However, these concepts are often hard to identify by humans and related term-based approaches (as used within this study) have already been successfully applied for sentiment analysis in other disciplines [42].

Finally, since product reviews have been found to influence sales, different product manufacturers have already started to publish very positive product reviews to increase the sales of their products or very negative product reviews in order to decrease the sales of their competitors [1]. Thus, analyzing product reviews published within the internet bears the risk that such manipulated reviews are part of the dataset and thus, the results of the study may be biased. However, this issue is addressed due to the fact that our results are based on an analysis of different products, so a manipulation related to a single product would only have a small impact on the overall results. Furthermore, since the products analyzed in our study are best sellers and connected with a large number of reviews, a manipulator would have to post a large amount of

manipulated reviews which makes a manipulation more time-consuming and less probable.

5 Summary and Conclusion

Product reviews have become important for online consumers and online retailers. In order to reduce information overload, previous research has started to identify the factors influencing review helpfulness in order to evaluate the helpfulness of unrated product reviews and to display the most helpful ones. Within this study, we extended the basic research model by Mudambi and Schuff (2010) [2] that only takes into account review extremity, review depth and product type and find that product quality, review sentiment and review uncertainty influence the level of review helpfulness. Thereby, product category moderates these relationships except from product quality.

With this study, we contribute to review diagnosticity theory by providing and evaluating a research model in order to explain which factors has an impact on review helpfulness. As follows, we confirm the impact of review extremity and review depth and extend review diagnosticity theory with the incorporation of product quality, review uncertainty and review sentiment. From a methodological point of view, we perform a deeper text analysis and also measure review sentiment based on textual content analysis rather than the number of stars.

From a practical point of view, we provide a basis for online retailers who want to develop review ranking systems for unrated reviews. In this context, they can apply the text analysis methodology which is part of this study in order to evaluate the different aspects of unrated online product reviews to identify the most helpful ones. Furthermore, review sentiment, product quality and review uncertainty could also be used as input variables for machine learning classifiers evaluating review helpfulness. Our study also provides helpful insights for online retailers to update their guidelines for online consumers on how reviews should be structured so that they are perceived as helpful. In this context, they should advise online consumers to basically focus on product-related characteristics, avoid uncertain statements and use positive sentiment for search goods and negative sentiment for experience goods.

Within future research, we plan to extend our study in multiple research directions. At first, we plan to investigate whether the cultural background of reviewers influences review style and the corresponding helpfulness assessments. Therefore, we want to examine product reviews related to products that are sold in different countries. Furthermore, our current study takes into account product reviews published on amazon.com which facilitates us to take a variety of products into account. Within further research, we want to extend our study on further data sources such as tripadvisor.com or imdb.com to validate whether the results remain robust also for reviews focusing on hotels and movies. Finally, since this study has the objective to explain what makes reviews helpful, as a next step, we plan to apply the understanding gained within this study in order to train and evaluate different machine learning classifiers to forecast the helpfulness of product reviews.

References

1. Chevalier, J.A., Mayzlin, D.: The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* 43, 345–354 (2006)
2. Mudambi, S.M., Schuff, D.: What Makes a Helpful Online Review? A Study of Customer Reviews on amazon.com. *MIS Quarterly* 34, 185–200 (2010)
3. Ghose, A., Ipeirotis, P.G.: Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *IEEE Transactions on Knowledge and Data Engineering* 23, 1498–1512 (2011)
4. Yin, D., Bond, S., Zhang, H.: Dreading and Ranting: The Distinct Effects of Anxiety and Anger in Online Seller Reviews. In: *ICIS 2011 Proceedings* (2011)
5. Forman, C., Ghose, A., Wiesenfeld, B.: Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research* 19, 291–313 (2008)
6. Zhu F, Z.X.: Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. *Journal of Marketing* 74, 133–148 (2010)
7. Wu, P., van der Heijden, H., Korfiatis, N.: The Influences of Negativity and Review Quality on the Helpfulness of Online Reviews. In: *ICIS 2011 Proceedings* (2011)
8. Korfiatis, N., García-Bariocanal, E., Sánchez-Alonso, S.: Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications* 11, 205–217 (2012)
9. Zeithaml, V.A.: Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. *Journal of Marketing* 52, 2–22 (1988)
10. Yang, Z., Fang, X.: Online service quality dimensions and their relationships with satisfaction: A content analysis of customer reviews of securities brokerage services. *International Journal of Service Industry Management* 15, 302–326 (2004)
11. Schindler, R.M., Bickart, B.: Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*. 11, 234–243 (2012)
12. Marks, G., Miller, N.: The Effect of Certainty on Consensus Judgments. *Personality and Social Psychology Bulletin* 11, 165–177 (1985)
13. Bennett, P.D., Harrell, G.D.: The Role of Confidence in Understanding and Predicting Buyers' Attitudes and Purchase Intentions. *Journal. of Consumer Research* 2, 110–117 (1975)
14. Mullainathan, S., Schwartzstein, J., Shleifer, A.: Coarse Thinking and Persuasion. *The Quarterly Journal of Economics* 123, 577–619 (2008)
15. Nelson, P.: Information and Consumer Behavior. *Journal of Political Economy* 78, 311–329 (1970)
16. Klein, L.R.: Evaluating the Potential of Interactive Media through a New Lens: Search versus Experience Goods. *Journal of Business Research* 41, 195–203 (1998)
17. Sweeney, J.C., Soutar, G.N.: Consumer perceived value: The development of a multiple item scale. *Journal of Retailing* 77, 203–220 (2001)
18. Howard, J.A., Sheth, J.N.: *The Theory of Buyer Behavior*. Wiley, New York (1969)
19. Feldman, J., Lynch, J.: Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology* 73, 421–435 (1988)
20. Krosnick, J.A., Boninger, D.S., Chuang, Y.C., Berent, M.K., Camot, C.G.: Attitude Strength: One Construct or Many Related Constructs? *Journal of Personality and Social Psychology* 65, 1132–1151 (1993)

21. Gershoff, A.D., Mukherjee, A., Mukhopadhyay, A.: Consumer Acceptance of Online Agent Advice: Extremity and Positivity Effects. *Journal of Consumer Psychology* 13, 161–170 (2003)
22. Li, J., Zhan, L.: Online Persuasion: How the Written Word Drives WOM. *Journal of Advertising Research* 51, 239–257 (2011)
23. Cheung, C.M., Lee, M.K., Rabjohn, N.: The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet Research* 18, 229–247 (2008)
24. Wen, I.: Factors affecting the online travel buying decision: a review. *International Journal of Contemporary Hospitality Management* 21, 752–765 (2009)
25. McCluskey, J.J.: A Game Theoretic Approach To Organic Foods: An Analysis Of Asymmetric Information And Policy. *Agricultural and Resource Economics Review* 29, 1–9 (2000)
26. Merriam-Webster: Merriam-Webster Online Dictionary: Sentiment. <http://www.merriam-webster.com/dictionary/sentiment>
27. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 79–86, Philadelphia (2002)
28. Uhl, M.W.: Explaining U.S. consumer behavior with news sentiment. *ACM Transactions on Management Information Systems* 2, 1–18 (2011)
29. Xia, L., Nasr N. I.: Word of Mouse: The Role of Cognitive Personalization in Online Consumer Reviews. *Journal of Interactive Advertising Research* 9, 3–13 (2008)
30. Elberse, A., Eliashberg, J.: Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science* 22, 329–354 (2003)
31. Mahajan, V., Muller, E., Kerin, R.A.: Introduction Strategy for New Products with Positive and Negative Word-of-Mouth. *Management Science* 30, 1389–1404 (1984)
32. Jordan, N.: The "Asymmetry" of "Liking" and "Disliking": A Phenomenon Meriting Further Reflection and Research. *The Public Opinion Quarterly* 29, 315–322 (1965)
33. Rozin, P., Royzman, E.B.: Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review* 5, 296–320 (2001)
34. Weber, R.P.: Measurement models for content analysis. *Quality & Quantity* 17, 127–149 (1983)
35. Schnurr, P.P., Rosenberg, S.D., Oxman, T.E., Tucker, G.: A Methodological Note on Content Analysis: Estimates of Reliability. *Journal of Personality Assessment* 50, 601–609 (1986)
36. Rosenberg, S.D., Schnurr, P.P., and Oxman, T. E.: Content Analysis: A Comparison of Manual and Computerized Systems. *Journal of Personality Assessment* 54, 298–310 (1990)
37. Stone, P.J., Hunt, E.B.: A Computer Approach to Content Analysis: Studies Using the General Inquirer System. *Proceedings of the AFIPS Spring Joint Computer Conference*, 241–256 (1963)
38. Weber, R.P.: Computer-aided content analysis: A short primer. *Qualitative Sociology* 7, 126–147 (1984)
39. Stone, P.J., Bales, R.F., Namenwirth, J.Z., Ogilvie, D.M.: The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science* 7, 484–498 (1962)
40. Weber, R.P.: *Basic Content Analysis* Sage, Newbury Park (1990)

41. Baccianella, S., Esuli, A., Sebastiani, F.: Multi-facet Rating of Product Reviews Proceedings of the ECIR 2009, LNCS 5478, 461–472 (2009)
42. Tetlock, P.C.: Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* 62, 1139–1168 (2007)