Journal of Retailing and Consumer Services 27 (2015) 113–120

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

Understanding online product ratings: A customer satisfaction model

Tobias H. Engler*, Patrick Winter, Michael Schulz

Philipps University at Marburg, School of Business & Economics, Universitätsstr. 24, 35032 Marburg, Germany

ARTICLE INFO

Article history:
Received 20 April 2015
Received in revised form
26 July 2015
Accepted 27 July 2015

Keywords:
Online ratings
Customer satisfaction
Electronic word of mouth
Expectation

ABSTRACT

Online product ratings have become a major information source for customers, retailers, and manufacturers. Both practitioners and researchers predominantly interpret them as a reflection of product quality. We argue that they in fact represent the customer's satisfaction with the product. Accordingly, we present a customer satisfaction model of online product ratings which incorporates the customer's pre-purchase expectations and actual product performance as determinants of ratings. We validate our model by applying it to two datasets collected at the German website of Amazon.com. The results indicate that both factors have a significant influence on online product ratings, supporting the proposed interpretation of ratings.

1. Introduction

Along with the growing diffusion of e-commerce, online product reviews have become a major information source for customers, retailers, and manufacturers. On the one hand, reviews and ratings contributed by online shop customers provide product information for prospective consumers, thereby reducing their uncertainty about the product (Chen and Xie, 2008). Consistently, research has shown that they affect sales in various contexts (e.g., Chevalier and Mayzlin, 2006; Lin et al., 2011; Park et al., 2007). On the other hand, online retailers and manufacturers increasingly rely on customer feedback to enrich their marketing strategy (Chen and Xie, 2008; Cui et al. 2012), to adjust product listings (e.g. via relevance sorting), and to create additional revenue streams (Mudambi and Schuff, 2010). For these reasons, it is not surprising that nearly all major online retailers such as Amazon.com or Ebay.com have implemented product rating functionalities.

Researchers, mainly from the fields of marketing and information systems, have adopted the topic and not only started to study the effects of online product ratings (e.g., on sales) but also their nature and determining factors. A common assumption of prior studies in the latter stream is that the baseline of a product's online ratings reflects its true quality. Various biases such as social dynamics or cultural influences were introduced to account for the unexplained part of the variance. However, empirical evidence suggests that online ratings do not accurately reflect a product's true quality (e.g., Hu et al., 2006; Koh et al., 2010). Since the influence of ratings on sales remains unaffected, retailers are left in an uncomfortable situation: it is difficult for them to adjust marketing strategies on the basis of online product ratings without knowing what they actually reflect.

Hence, the objective of this study is to find out what really builds the baseline of online product ratings and thereby refine their current interpretation. We argue that the weak explanatory power of product quality for online reviews is not only caused by actual biases: it is mainly caused by product ratings reflecting customer satisfaction than being a valid measure for product quality. This approach does not solely rely on product quality as the baseline for the rating but also integrates the customer's expectation of the product in the pre-purchase phase. Correspondingly, we present a customer satisfaction model of online product ratings based on the considerations of Fornell (1992) and Westbrook and Reilly (1983). We model the customer's pre-purchase expectation of the product and the actual performance as predictors of online ratings using structured equations. We validate our model by applying it to two datasets (digital cameras and televisions) collected from the German website of Amazon.com. The results indicate that both a customer's expectation of a product and the actual performance significantly influence the ratings customers assign to a product, supporting the proposed interpretation of online product ratings.

Several other observations in the datasets can help to get a more comprehensive view of online product ratings and are worth mentioning. First, we find that online ratings carry some percentage of unobservable information that cannot be predicted (using metrics from the website). Second, the data shows indications for confirmation, acquisition, and under-reporting biases.

* Corresponding author. Fax: +49 6421 2826554.
E-mail addresses: tobias.engler@wiwi.uni-marburg.de (T.H. Engler), patrick.winter@wiwi.uni-marburg.de (P. Winter), michael.schulz@wiwi.uni-marburg.de (M. Schulz).

http://dx.doi.org/10.1016/j.jretconser.2015.07.010
0969-6989/

Fig. 1. Word cloud of Amazon review (the bigger the word, the more frequent it was used in the reviews).

In the next section, we review the extant literature on product ratings. In Section 4.1.1, we elaborate on the theoretical background of customer satisfaction and present our research model. In Section 4.2.1, we discuss the measurement of the latent constructs and describe the collection of data. There, we also present the results of our analysis, which are discussed in Section 5. In Section 5, we conclude this paper by explaining its implications for practice and research, its limitations, and possible directions for future research.

2. Review of literature on product ratings

Online product ratings have been researched from various perspectives, which can be classified according to whether they investigate their antecedents or outcomes. The first category rather includes studies examining the influence of online product ratings on consumers' purchasing decisions. In most of these studies, a significant positive influence is found (e.g., Lin et al., 2011; Mauri and Minazzi, 2013; Park et al., 2007; Sun, 2011), although it may diminish over time (Hu et al., 2008). For a more detailed review we refer to Moe and Trusov, (2011) as this paper belongs to the second category that investigates the factors on which ratings depend.

Research on the antecedents of product ratings mainly focuses on different biases influencing the rating score. Wulff and Hardt, (2014) and Koh et al., (2010) find evidence of cultural differences among raters. The latter authors also find that product ratings do not reflect true (perceived) product quality, a result also obtained by Hu et al., (2006). Sridhar and Srinivasan, (2012) find that prior customer ratings weaken the effect of product experience but can either increase or turn around the effect of product failure, depending on how the latter is addressed. Moe and Schweidel, (2012) link a customer's rating to his decision to rate. They find that the influence of prior customers' ratings on a customer depends on how often this customer rates products.

However, most of these studies suffer from two major shortcomings. First, the products which are analyzed are often books or movies. The individual ratings of these products can be assumed to depend strongly on personal taste and not on a common ground. Second, the true quality of a product is measured via surveys among non-experts. Therefore, they may not be well-suited to assess a quality effect in online product ratings, limiting the validity.

3. Hypothesis development

Previous research on online ratings has started from the premise that the baseline of online ratings represents the true quality of a product (cf. Hu et al., 2006; Moe and Trusov, 2011), including statements such as "each posted online review is an assessment of a single individual's perceived quality of a product" (Koh et al., 2010, p. 374). The conclusion of these studies is that online ratings do not accurately reflect product quality. They attribute this fact to various biases instead of questioning their underlying premise. Since sales heavily depend on online ratings, misinterpreting their baseline can be the root of misguided actions of retailers and manufacturers.

We argue that online product ratings are rather an expression of customer satisfaction than a pure quality assessment. The construct customer satisfaction is a central concept in marketing research measuring individual-level satisfaction with products and services (Yi, 1991). It is defined as a function of the customer's expectation and product performance (Fornell, 1992; Fornell et al., 1996). Expectation is conceptualized as the perceived probabilities of what the consequences of a purchase will be (Oliver, 1980). Performance refers to the perceived quality of the product after the purchase. Instead of being tied to a purely post-purchase quality-centered perspective (as online product ratings are assumed to be in prior research), the definition of customer satisfaction adds the perspective of pre-purchase expectations. In the following, we outline why online product ratings reflect customer satisfaction with the product and thereby include the consideration of both: the customers' expectation of the respective product and the performance they perceive after buying it.

A strong indication that online product ratings reflect customer satisfaction was found by searching through the full-text reviews on Amazon.com. For this, we used Amazon review data consisting of reviews from 1996 until 2014 (McAuley et al., 2015). The complete dataset contains 7,834,166 reviews of products in the category electronics. We created a word cloud (Fig. 1) using a randomly selected subset (10% of the complete dataset) to better understand the meaning behind online reviews. Stop words were removed and same words in plural/singular or different tenses were consolidated to condense the findings into a cloud depicting the top 100 words.

Besides words describing the general experience with the product ("great" was found in 36.6% of the reviews), product parts and product categories (e.g., cable (11.2%), camera (18.8%)), the cloud includes evidence for the proposed interpretation of online ratings (e.g., expect (5.3%), price (17.8%), and reviews (5.0%)). Additionally, we searched the full dataset for "expect" (as a word or part of the words expected and expectation) and found that 571,939 reviews (7.3%) include this word. When compared to the search result for the word "quality" (1,167,525 reviews, 14.9%), these figures strongly indicate that expectation is not a marginal phenomenon in review texts. Since the score of individual online ratings is consistent with the corresponding text (Ganu et al., 2009), it is very likely that they are based on more than a pure assessment of product quality. Instead, the baseline effect of online ratings includes both, pre-purchase expectations and the post-

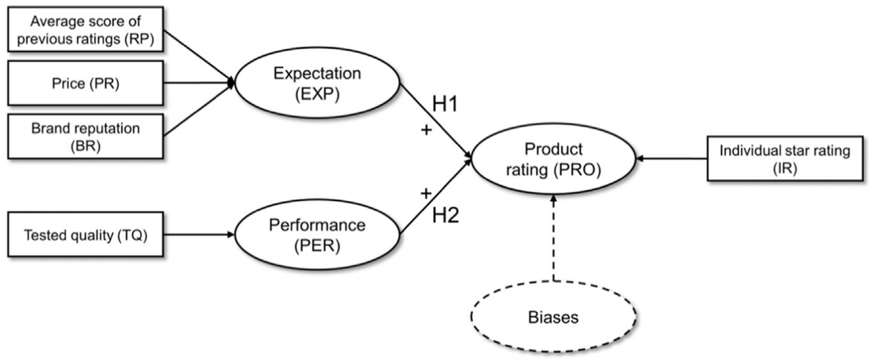

Fig. 2. Consumer satisfaction model of online product ratings.

purchase performance. Hence, we assume that customer satisfaction reflects the baseline of the rating score.

Biases may distort this baseline. Online ratings can, therefore, be expressed as a function of the baseline effect and biases. Because the latter stream has already been extensively researched as described in the previous section, we now elaborate in more detail on pre-purchase expectations, post-purchase performance, and the mechanisms behind their effect on the online product rating score. The resulting research model (including the measurement models discussed in the following section) is presented in Fig. 2.

An explanation for the underlying effect of pre-purchase expectations on online product rating scores is provided by the belief-adjustment model (Hogarth and Einhorn, 1992; Bolton, 1998). It describes the order of belief updating over time as a process of anchoring and adjustments. The central message of the belief-adjustment model is that individuals do not directly react to a new stimulus but rather adjust their prior expectations on the specific topic to the new stimulus while sustaining in the vicinity of the original anchor (cf. Oliver, 1980). Thus, pre-purchase expectations should have a positive impact on satisfaction. It was found to be applicable in various contexts. This leads us to assume that this process also takes place in the context of online shopping and the pre-purchase evaluation of products. First, customers form an expectation what the product might be like on the basis of information found on the product website. In a second step, they adjust this anchor within a reference frame set by the initial judgement when being confronted with the product's performance after the purchase and delivery. Hence, we hypothesize:

Hypothesis H1. : Pre-purchase expectations (EXP) have a positive impact on the score of online product ratings (PRO).

The direct effect of performance on satisfaction is supported by the value-percept disparity model developed by Westbrook and Reilly, (1983). They posit that satisfaction is a general perception based on the evaluation of customers' experiences with a product. A high satisfaction can, therefore, only be achieved if a product is able to fulfill the customer's needs. This mechanism is consistent with findings from Churchill and Suprenant (1982). The results of their study suggest that satisfaction with a durable good can be predicted by the product performance to a considerable extent. Further studies also support this direct effect of performance on satisfaction (Anderson and Sullivan, 1993; Fornell, 1992). Transferred to the online environment, this means that online product ratings are indeed influenced by the experienced quality of the product, as assumed by prior research (e.g., Koh et al., 2010). The product's performance should, therefore, have a positive effect on the score of online ratings. Thus,

Hypothesis H2. : A product's post-purchase performance (PER) has a positive impact on the score of online product ratings (PRO).

4. Research method and data analysis

4.1. Measurement and data collection

The research model was tested using crawled data of cameras and televisions to address the two major shortcomings of prior research as described above. Books and movies can be classified as experience goods while cameras and televisions are search goods (cf. Nelson, 1970, 1974). The ratings of experience goods heavily depend on personal feelings, cannot be evaluated on the basis of specific characteristics, and may vary across different individuals (Mudambi and Schuff, 2010; Weathers et al., 2007). Whereas the beauty of a book or a movie is in the eye of the beholder, it is pointless to argue about objective measures such as battery lifetime or viewing angel stability. Search goods such as cameras and televisions can be evaluated using a more systematic approach (Cui et al., 2012) including rather objective criteria such as technical functions (e.g., megapixels) into the evaluation process, hence, increasing rating reliability.

4.1.1. Expectation

The aim of this research is to identify factors that constitute the score of online ratings made by customers of an online shop. For this, we adopted the customer's perspective and focus on quantitative data that can be included in the evaluation by quickly overlooking the product's description on the website (see Fig. 3). Accordingly, expectation was captured using three indicators that can be evaluated by customers this way before buying the product: the average score of previous ratings, the product price, and brand reputation. While the score of previous ratings is the major source of information for online customers (Koh et al., 2010; Cui

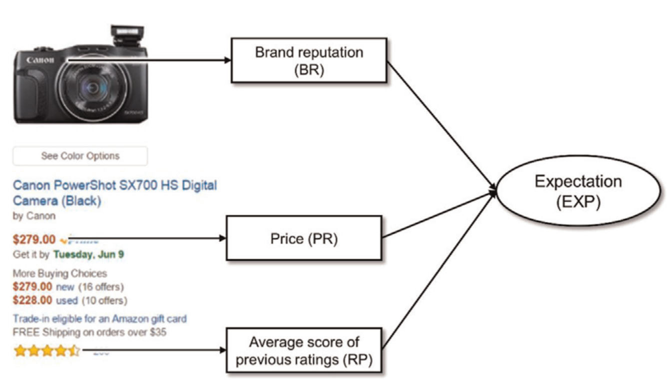

Fig. 3. Product description on amazon.com and measurement model of expectation.

et al., 2012), product price and brand reputation have been identified as the most important extrinsic (not product-inherent) quality indicators in the offline world (Zeithaml, 1988). The measures of the construct expectation are formative since a change in the indicators cause a change in the construct rather than reflecting it. Furthermore, there is no reason to expect that the indicators are necessarily highly correlated (Jarvis et al., 2003).

In online shops, usually two directly observable quantitative indicators of the customer feedback are accessible: the number and the score of previous ratings. While a high number of previous ratings may enhance the subjective weight of the score of previous ratings, the latter affect the customer's expectation directly. The influence of customer ratings on the customer's perception can be explained by different manifestations of social power. Five bases of social power have been identified: expert power, legitimate power, referent power, reward power, and coercive power (French and Raven, 1959). The customer's decision to rely on customer ratings can be attributed to the two mechanisms referent and expert power (Engler, 2014). Referent power describes the effect that individuals seek to hold similar opinions with their social environment to achieve personal satisfaction by conformity. The second phenomenon can occur even if conformity is not the root of social power. French and Raven, (1959) state that conformity with the group's opinion (here: the group of raters) can also be caused by expert power. For this, the customer regards the aggregated wisdom of previous ratings as an expression of expertise. We measure the score of the previous ratings by averaging all star ratings of the respective product up to the time of the individual rating.

Price is the second indicator forming expectation. It has been identified to influence the perceived quality of the product in offline and online shops (e.g., Dodds et al., 1991; Rao and Monroe, 1989; Chen and Dubinsky, 2003). Customers consider the product price as an indicator for product quality because they believe that the interplay of supply and demand leads to an order of competing products on a price scale in accordance with their quality (Scitovszky, 1944). The price information was collected in the same time period as the performance indicator.

A vast body of research (e.g., Dodds et al., 1991; Jacoby et al., 1971; Zeithaml, 1988) has found that not only price but also brand reputation also influences the expected performance of a product. Similarly to the effect of price on the expected performance, the brand name can add information to the product that can otherwise not be accessed in the pre-purchase phase (Zeithaml, 1988). Customers assume that companies do not want to threaten a positive reputation by selling poor quality products (e.g., Nguyen and Leblanc, 2001; Yoon et al., 1993). Therefore, we model brand reputation as the third indicator constituting expectation. A well-proven measure for brand reputation is RepTrak® (Ponzi et al., 2011). RepTrak® is measured by the Reputation Institute and is based on an emotion-based measure of corporate reputation (Reputation Institute, 2014). We used the Global RepTrak® 100 score which is based on data collected in 15 countries (including Germany) to calculate the model. National differences are not as important as customer differences for high-tech goods such as consumer electronics and customers of these products are globally similar (Domzal and Unger, 1987). Hence, we assume that using the Global RepTrak® would not lead to a considerable bias in this study where we use data from the German website of Amazon.com. We used the brand specific RepTrak® scores that were up-to-date the time performance was measured.

4.1.2. *Performance*

Performance is the construct that product quality relates to. Prior research made a distinction between an objective and a perceived concept of product quality (e.g., Garvin, 1983; Holbrook and Corfman, 1985). While objective quality is defined as the "actual technical superiority or excellence of the products" (Zeithaml, 1988, p. 4), perceived quality reflects consumers' judgment about the products' features. However, it soon was recognized that objective quality can hardly be measured because the criteria which are used to do so and their weights are chosen subjectively (Zeithaml, 1988). Still, a distinction should be made between quality assessments that are mainly based on subjective feelings and experiences and those that are based on scientific and repeatable measurement methods. We refer to the latter as tested quality and use a correspondingly named indicator to measure performance.

More precisely, performance is measured using the grades of "Stiftung Warentest" (SW), a German customer magazine similar to "Consumer Reports" in the US. SW is a neutral organization founded by the German government in 1964. It is financed by selling test results online and in paper form and supported by the state (Stiftung Warentest, 2015). The organization's main objective is to test products and services using scientific methods. The test results of SW can be downloaded from a fee-based website. The grades represent an objective and mechanistic approach to evaluate products. The measurement is reliable since the outcomes of repeated product tests (even if carried out by different persons) would lead to the same results. The data for the single indicator *tested quality* was collected on the website of SW. We included all digital camera and television tests during a five year period from 2009 to 2014 to achieve a large overlap between the tested products and currently sold products on the German website of Amazon.com. The evaluation scheme of SW ranges from 1 = "very good" to 5 = "inadequate" and has been inverted before calculation. Overall product grades are the result of averaged sub category grades that can lie anywhere between the extremes, and are rounded to one decimal place. The distribution of SW grades in our sample is shown in Fig. 4.

4.1.3. *Product rating*

As discussed above, the customer feedback in online shops can be seen as the expression of their satisfaction with the product. Therefore, the dependent variable product rating was measured using the score of the individual star rating that a customer has

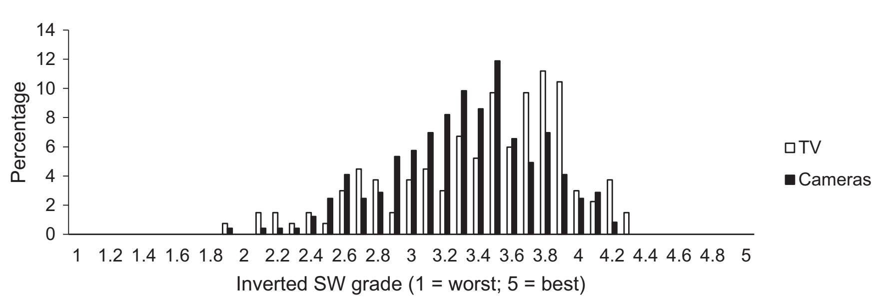

Fig. 4. Distribution of SW grades per product in the studies on cameras and televisions.

assigned to a specific product. We gathered the online rating data from the German website of Amazon.com, which is by far the largest online store in Germany (Statista, 2015). Other online stores were not taken into account to avoid biases caused by different qualities of retailers. Customers of Amazon.com are able to rate products on a five-star rating scale ranging from 1 = "I hate it" to 5 = "I love it" and additionally write customer reviews. We used a crawler to identify those products on Amazon.com that were tested from SW and downloaded all ratings and their timestamps. The camera data was collected on September 13, 2014 and the television data was collected on November 05, 2014.

Of the 1423 products tested by SW between 2009 and 2014, 56 percent have been identified on the Amazon.com website. 31 products were removed from the analysis because no online product ratings were available. Some reviews on Amazon.com are not uniquely associated with one product, but with several product types (e.g., a television model that is available with a 42 in. and 55 in. screen). In this case, one of the products was randomly selected following the procedure of Lim et al., (2010). Because of this, 62 items were removed so that all ratings are assigned to only one product. Additionally, we included only manufacturers of digital cameras and televisions that are in the RepTrak® 100. Overall, 378 products and 28,873 ratings were used for the calculation. Table 1 presents a detailed overview of the dataset and Figs. 5 and 6 illustrate the distribution of individual star ratings per customer and the average star rating per product on Amazon.com.

4.2. Data analysis and results

Structural equations were used to model the research model. Structural equation modelling (SEM) is a family of techniques that allow to model relationships between one or more independent variables and one or more dependent variables. Both independent and dependent variables can either be measured directly or indirectly (latent variables) (Ullman and Bentler, 2003). SEM differentiate between measurement models of (latent) variables and the relationships between the variables – the so-called structural model. Within the set of SEM techniques we chose the partial least squares (PLS) algorithm (cf. Chin, 1998) because it allowed us to handle single item measures (performance and product rating) and formatively measured latent constructs (expectation) simultaneously (Hair et al., 2013). Distributions of the indicators building a satisfaction construct are often heavily skewed (Fornell, 1992). The distribution of star ratings in the presented studies show a high skewness towards the higher ratings as well (see Fig. 5). PLS offers the advantage that non-normal distributions can be computed without manipulating the original data. Therefore, we used SmartPLS 3.2 (Ringle et al., 2015) to calculate the data. In a first step we evaluate the measurement model of the construct expectation and then assess the relationships of the research model.

Table 1
Dataset details.

Criterion	Cameras	Televisions	Total
Products tested by SW since 2009	885	538	1423
Products found on the German Amazon.com website	571	222	793
Products that were removed because they had no ratings	15	16	31
Products that were removed because of duplicate ratings	0	62	62
Products that were removed because of missing brand indices	312	10	322
Products used in the analysis	244	134	378
Product ratings used in the analysis	12,563	16,310	28,873

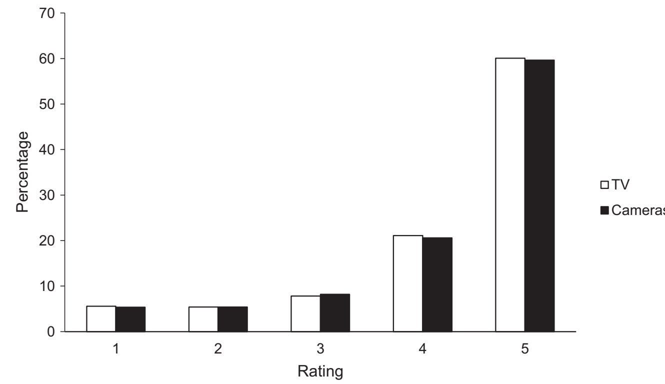

Fig. 5. Distribution of individual Amazon.com ratings in the studies on cameras and televisions.

4.2.1. Measurement models

The formative measurement model of the latent construct expectation was evaluated by looking at the indicator weights, their significance, and an assessment of multicollinearity (Hair et al., 2012). An overview of the results is given in Table 2. In the study on cameras all indicators significantly affect expectation in the theorized way. The variance inflation factor (VIF) score of 1.5 is well below the recommend upper limit of 5 and indicates a non-critical level of collinearity (Hair et al., 2013). The study on televisions shows mixed results. Only the indicator previous ratings has a positive and significant weight while the outer weight of price is insignificant and brand reputation is significant but negative. Nevertheless, we kept the indicators in the research model for two reasons. First, an elimination of insignificant indicators would affect the definition of the construct (Diamantopoulos and Winklhofer, 2001) and would lead to an incomparability between the two studies. Second, negatively weighted items should remain in the model if they are collinear and do not show reversed signs across studies (Cenfetelli and Bassellier, 2009).

4.2.2. Structural model

As shown in Table 3 and Table 4, we examined the path coefficients (β) and the level of significance for every hypothesized relationship as well as the explained variance (R^2) of the dependent variable for both studies. The path coefficients between expectation (CA: β=0.133; TV: β=0.202) and performance (CA: β=0.044; TV: β=0.024) on the one hand and satisfaction on the other hand are significant at $p < 0.001$ confirming H1 and H2 in both studies. Expectation consistently affects satisfaction considerably higher than performance. Although both hypotheses are strongly confirmed for cameras and television, expectation and performance explain 2.6% and 4.2% of the satisfaction variance respectively. We discuss the implications of these results in the next section.

5. Discussion

Comparing the distributions of ratings given to a product by SW (Fig. 4) and by customers (on average) (Fig. 6), we first note that product ratings do not reflect pure product quality since both distributions clearly differ from each other. This is consistent with prior research (Hu et al., 2006; Koh et al., 2010) and results, inter alia, from neglecting users' expectations, as elaborated on earlier.

When taking into account users' expectations an ambiguous conclusion can be drawn: On the one hand, H1 and H2 are confirmed by our data; that is, we find our research model to be suited for explaining online product ratings. As indicated by the text reviews, the rating score reflects the customer's expectation of the

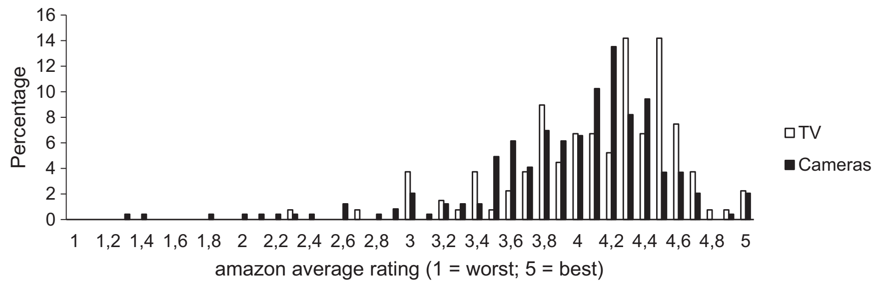

Fig. 6. Distribution of average Amazon.com ratings per product in the studies on cameras and televisions.

Table 2
Measurement model of the construct expectation.

	BR	PR	RP	EXP
Study CA (Camera)	0.154**	0.423***	0.745***	1.429
Indicator weight				
VIF				
Study TV (Television)	−0.093*	−0.059n.s.	1.003***	1.008
Indicator weight				
VIF				

* $p < 0.05$.
** $p < 0.01$
*** $p < 0.001$

Table 3
Statistics of dataset from study on digital cameras.

	r^2	Path coefficient (β)	*T*-value	*p*-value	Hypothesis confirmed
PRO	0.026				
EXP→PRO		0.133***	13.121	$p < 0.0001$	H1: Yes
PER→PRO		0.044***	4.412	$p < 0.0001$	H2: Yes

*** $p < 0.001$

Table 4
Statistics of dataset from study on televisions.

	r^2	Path coefficient (β)	*T*-value	*p*-value	Hypothesis confirmed
PRO	0.042				
EXP→PRO		0.202***	24.700	$p < 0.0001$	H1: Yes
PER→PRO		0.024***	3.806	$p < 0.0001$	H2: Yes

*** $p < 0.001$

product and an assessment of product quality. In contrast to the prevailing opinion, we find ratings to be even more influenced by expectation than product quality.

On the other hand, however, the explained variance is relatively low in both studies. This can be assumed to have three causes: First, a particular customer's satisfaction with the product he has purchased is likely to depend on the expectation and performance of his specific needs (e.g., a long lasting battery). These needs are not observed in our study. The drawback of measuring performance by the product's quality (i.e., on product-level) is that we cannot break down the performance to each of its characteristics. Second, the presence of high fake ratings significantly diminishes explained variances: since a fake rating does neither depend on customer expectation nor on product performance, the corresponding observation cannot be explained by our model. Thus, the percentage of variance explained can be expected to be much higher if no fake ratings are present. In contrast, it should be noted that the general results are not affected by fake ratings because their frequency distribution can be assumed to be uncorrelated with the indicators used. Third, we argue that reviews exhibit a high degree of "randomness" by nature. This result, which might seem intuitive at a first glance, has an important implication: if individual ratings could be explained by any model to a high degree, they would become superfluous. A rating that can be accurately predicted cannot contain any new information.

The same applies if ratings are rather determined by observable factors than by raters' experiences. Indeed, we find them to be significantly influenced by a product's price and the reputation of its manufacturer consistent with results of prior research (Dodds et al., 1991). Furthermore, our results provide evidence for social dynamics as described in Moe and Trusov (2011). Customers base their evaluations rather on previous ratings than on their individual experience. The weight of the previous ratings' score is even greater than the weights of the other indicators, suggesting that social dynamics have a stronger influence on customers than price or brand effects.

We also find signs for biases during the rating process. First, the product rating distribution is highly skewed. This is often attributed to under-reporting bias (Anderson, 1998): customers with extreme values of satisfaction (very low or very high) are more likely to review a product than customers with mean levels. Interestingly, however, the distribution is negatively skewed, that is, high ratings are much more prominent than low ratings. This may have two reasons. First, it could result from the so-called acquisition bias (Hu et al., 2006): only users who have a sufficiently high expectation of a product will consider purchasing it. Second, it is known that a certain amount of ratings are fake (e.g., ca. 16% at yelp.com, (Luca and Zervas, 2013)). They are created by or on behalf of manufacturers and retailers to increase the average ratings and, hence, the sales of their products. This effect spans a stream of research of its own (e.g., Malbon, 2013; Lappas et al., 2012; Mukherjee et al., 2012). We find no indications for the reverse effect, that is, fake reviews given to products by competitors in order to decrease their average rating.

Finally, we find that customer satisfaction is more affected by expectation than by performance. In addition to the hypothesized belief-adjustment mechanism underlying the relationship between expectation and rating, this might also indicate a confirmation bias (cf. Nickerson, 1998). Customers tend to interpret evidence in favor of their prior expectations about the product instead of evaluating the product they have purchased objectively – they see what they like to see. This also relates to the theory of cognitive dissonances (Festinger, 1962). If the product does not meet their expectations, a cognitive dissonance between expectation and performance occurs. Our results suggest that customers rather resolve this dissonance by mitigating the product's deficiencies than by revising their expectations.

6. Implications, limitations, and conclusion

In this study, we have shown that the customer satisfaction model of online product ratings is better suited to explain the score of ratings than traditional quality-centered explanations. This means that customers' ratings of products depend on their expectation about these products and their performance. This finding has rich and concrete implications for both research and practice.

The development and empirical test of this model advances theoretical knowledge by introducing the customers' expectation as a determinant of online ratings. Thereby, we refine the current understanding of the baseline of online ratings. The empirical results suggest that the model provides a valuable tool to analyze online ratings and is a valid starting point to elaborate on biases more accurately.

Without the insights of this study, practitioners in retailing and manufacturing may draw erroneous conclusions for marketing decisions based on existing reviews if they rely on the invalid assumption that online product ratings reflect true quality. To counteract this, rating mechanisms have to be optimized. We recommend establishing a rating system that allows users to input their individual expectations of specific products. This way, products can be ranked according to a rating based on the confirmation of expectations. For example, a user who wants to take a few snapshots has other expectations towards a digital camera than a professional photographer. When considering solely the rating score, both camera types might look like they were recommended for both user types but the camera for beginners will most certainly not meet the expectations of professional photographers and vice versa. The problem is that the different expectations are not accounted for by current rating systems based on a single rating value. Even current multi-criteria rating systems (as used, e.g., on ebay.com), which allow ratings for several criteria (e.g., robustness) of a product are not suited for this approach. This is because expectations may relate to several criteria simultaneously (e.g., quality and support). By assessing expectations and the degree of fulfilment separately, manufacturers can learn about users' expectations of their own and competing products which enables them to develop better marketing strategies. On the other hand, they can deduce how satisfied their customers are with the degree to which these expectations are met which enables better product designs. Furthermore, the accuracy rate of recommender systems can be improved this way. If retailers know the customer's expectation of a product, they can suggest further potentially interesting products with similar expectation values.

As every empirical work, our paper is not free of limitations. First, we have analyzed online ratings without considering the textual reviews accompanying them to test the theory. Prior research (Chevalier and Mayzlin, 2006) has shown that these reviews carry information that adds up to the information carried by the ratings. Furthermore, some websites offer their users the possibility to rate customer reviews in order to determine their helpfulness (Mudambi and Schuff, 2010). These second-order ratings were not considered in this study to emphasize the theorized hypotheses. Future research can validate our results by including these additional data. Second, crawled data were chosen to evaluate this exploratory research model. On the hand side, this approach increases the external validity but one the other hand, it limits internal validity. To contradict this issue, survey-based research measurement of expectation should be employed in addition to crawled data in future research. Third, we focused a single marketplace and two groups of tech products to avoid biases caused by different shopping experiences. Future research can, therefore, examine different marketplaces or products from different categories such as experience goods.

References

Anderson, E.W., 1998. Customer satisfaction and word of mouth. J. Serv. Res. 1 (1), 5–17.
Anderson, E.W., Sullivan, M.W., 1993. The antecedents and consequences of customer satisfaction for firms. Mark. Sci. 12 (2), 125–143.
Bolton, R.N., 1998. A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction. Mark. Sci. 17 (1), 45–65.
Cenfetelli, R.T., Bassellier, G., 2009. Interpretation of formative measurement in information systems research. MIS Q. 33 (4), 689–707.
Chen, Y., Xie, J., 2008. Online consumer review: word-of-mouth as a new element of marketing communication mix. Manag. Sci. 54 (3), 477–491.
Chen, Z., Dubinsky, A.J., 2003. A conceptual model of perceived customer value in E-commerce: a preliminary investigation. Psychol. Mark. 20 (4), 323–347.
Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. J. Mark. Res. 43 (3), 345–354.
Chin, W.W., 1998. The partial least squares approach to structural equation modeling. In: Marcoulides, G.A. (Ed.), Modern Methods for Business Research. Mahwah: Lawrence Erlbaum Associates, pp. 295–358.
Churchill Jr., G.A., Surprenant, C., 1982. An investigation into the determinants of customer satisfaction. J. Mark. Res. 19 (4), 491–504.
Cui, G., Lui, H.-K., Guo, X., 2012. The effect of online consumer reviews on new product sales. Int. J. Electron. Commer. 17 (1), 39–58.
Diamantopoulos, A., Winklhofer, H.M., 2001. Index construction with formative indicators: an alternative to scale development. J. Mark. Res. 38 (2), 269–277.
Dodds, W.B., Monroe, K.B., Grewal, D., 1991. Effects of price, brand, and store information on buyers' product evaluations. J. Mark. Res. 28 (3), 307–319.
Domzal, T., Unger, L., 1987. Emerging positioning strategies in global marketing. J. Consum. Mark. 4 (4), 23–37.
Engler, T.H., 2014. Trusting review mechanisms in knowledge management systems: antecedents, outcomes, and the role of perceived risk. In: ECIS 2014 Proceedings. Tel-Aviv.
Festinger, L., 1962. A Theory of Cognitive Dissonance. Stanford University Press, Stanford.
Fornell, C., 1992. A national customer satisfaction barometer: the Swedish experience. J. Mark. 56 (1), 6–21.
Fornell, C., et al., 1996. The American customer satisfaction index: nature, purpose, and findings. J. Mark. 60 (4), 7–18.
French, J.R.P., Raven, B., 1959. The Bases of Social Power. In: Cartwright, D. (Ed.), 1959. MI: Institute for Social Research, pp. 150–167, pp..
Ganu, G., Elhadad, N., Marian, A., 2009. Beyond the stars: improving rating predictions using review text content. In WebDB. Citeseer, pp. 1–6.
Garvin, D.A., 1983. Quality on the Line. Harv. Bus. Rev. 61 (5), 64–75.
Hair, J.F., et al., 2012. An assessment of the use of partial least squares structural equation modeling in marketing research. J. Acad. Mark. Sci. 40 (3), 414–433.
Hair, J.F., et al., 2013. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Thousand Oaks: SAGE Publications, Thousand Oaks, California.
Hogarth, R.M., Einhorn, H.J., 1992. Order effects in belief updating: the belief-adjustment model. Cogn. Psychol. 24 (1), 1–55.
Holbrook, M.B., Corfman, K.P., 1985. Quality and value in the consumption experience: phaedrus rides again. Perceived Qual. 31 (2), 31–57.
Hu, N., Liu, L., Zhang, J., 2008. Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. Inf. Technol. Manag. 9 (3), 201–214.
Hu, N,. Pavlou, P.A., Zhang, J., 2006. Can online reviews reveal a product's true quality? In: Proceedings of the 7th ACM Conference on Electronic Commerce. New York, pp. 324–330.
Institute, R., 2014. Global RepTrak® 100. Available at: ⟨http://www.reputationinstitute.com/thought-leadership/global-reptrak-100⟩ (Accessed: 01.04.15).
Jacoby, J., Olson, J.C., Haddock, R.A., 1971. Price, brand name, and product composition characteristics as determinants of perceived quality. J. Appl. Psychol. 55 (6), 570–579.
Jarvis, C.B., et al., 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J. Consum. Res. 30 (2), 199–218.
Koh, N.S., Hu, N., Clemons, E.K., 2010. Do online reviews reflect a product's true perceived quality? an investigation of online movie reviews across cultures. Electron. Commer. Res. Appl. 9 (5), 374–385.
Lappas, T., 2012. Fake reviews: the malicious perspective. In: Bouma, G., et al. (Eds.), Natural Language Processing and Information Systems. Springer, Berlin Heidelberg, pp. 23–34.
Lim, E.-P. et al., 2010. Detecting Product Review Spammers Using Rating Behaviors. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management. New York, pp. 939–948.
Lin, C.-L., Lee, S.-H., Horng, D.-J., 2011. The effects of online reviews on purchasing intention: the moderating role of need for cognition. Soc. Behav. Personal.: Int. J. 39 (1), 71–81.
Luca, M., Zervas, G., 2013. Fake it till you make it: reputation, competition, and yelp review fraud. Harvard Business School NOM Unit Working Paper, (14-006).
Malbon, J., 2013. Taking fake online consumer reviews seriously. J. Consum. Policy 36 (2), 139–157.
Mauri, A.G., Minazzi, R., 2013. Web reviews influence on expectations and purchasing intentions of hotel potential customers. Int. J. Hosp. Manag. 34 (0), 99–107.

McAuley, J. et al., 2015. Image-based Recommendations on Styles and Substitutes. In SIGIR 2015 Proceedings. Santiago de Chile.
Moe, W.W., Trusov, M., 2011. The value of social dynamics in online product ratings forums. J. Mark. Res. 48 (3), 444–456.
Moe, W.W., Schweidel, D.A., 2012. Online product opinions: incidence, evaluation, and evolution. Mark. Sci. 31 (3), 372–386.
Mudambi, D., Schuff, S.M., 2010. What makes a helpful online review? a study of customer reviews on Amazon.com. MIS Q. 34 (1), 185–200.
Mukherjee, A., Liu, B., Glance, N., 2012. Spotting Fake Reviewer Groups in Consumer Reviews. In: Proceedings of the 21st International Conference on World Wide Web. New York, pp. 191–200.
Nelson, P., 1970. Information and consumer behavior. J. Political Econ. 78 (2), 311–329.
Nelson, P., 1974. Advertising as information. J. Political Econ. 82 (4), 729–754.
Nguyen, N., Leblanc, G., 2001. Corporate image and corporate reputation in customers' retention decisions in services. J. Retail. Consum. Serv. 8 (4), 227–236.
Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gen. Psychol. 2 (2), 175–220.
Oliver, R.L., 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. J. Mark. Res. 17 (4), 460–469.
Park, D.-H., Lee, J., Han, I., 2007. The effect of on-line consumer reviews on consumer purchasing intention: the moderating role of involvement. Int. J. Electron. Commer. 11 (4), 125–148.
Ponzi, L.J., Fombrun, C.J., Gardberg, N.A., 2011. RepTrak™ pulse: conceptualizing and validating a short-form measure of corporate reputation. Corp. Reput. Rev. 14 (1), 15–35.
Rao, A.R., Monroe, K.B., 1989. The effect of price, brand name, and store name on buyers' perceptions of product quality: an integrative review. J. Mark. Res., 351–357.
Ringle, C.M., Wende, S. & Becker, J.-M., 2015. SmartPLS 3.2.
Scitovszky, T., 1944. Some consequences of the habit of judging quality by price. Rev. Econ. Stud. 12 (2), 100–105.
Sridhar, S., Srinivasan, R., 2012. Social influence effects in online product ratings. J. Mark. 76 (5), 70–88.
Statista, 2015. Umsatz der 100 größten Online-Shops in Deutschland im Jahr 2013 (in Millionen Euro). Available at: ⟨http://de.statista.com/statistik/daten/studie/170530/umfrage/umsatz-der-groessten-online-shops-in-deutschland/⟩.
Sun, M., 2011. How does the variance of product ratings matter? Manag. Sci. 58 (4), 696–707.
Ullman, J.B., Bentler, P.M., 2003. Structural equation modeling. In: Schinka, J.A., Velicer, W.F. (Eds.), Handbook of Psychology: Research Methods in Psychology Vol. 2. Hoboken: Wiley Online Library, Hoboken, New Jearsey, pp. 607–634.
Warentest, S., 2015. Über uns. Available at: ⟨https://www.test.de/unternehmen/ueberuns⟩ (Accessed 01.04.15).
Weathers, D., Sharma, S., Wood, S.L., 2007. Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. J. Retail. 83 (4), 393–401.
Westbrook, R.A., Reilly, M.D., 1983. Value-percept disparity: an alternative to the disconfirmation of expectations theory of consumer satisfaction. In: Bagozzi, R. P., Tybout, A.M. (Eds.), NA – Advances in Consumer Research Volume 10. MI: Association for Consumer Research, Ann Arbor, pp. 256–261.
Wulff, J. & Hardt, D., 2014. Can You Trust Online Ratings? Evidence of Systematic Differences in User Populations. In ECIS 2014 Proceedings.
Yi, Y., 1991. A critical review of customer satisfaction. In: Zeithaml, V.A. (Ed.), Review of Marketing 1990. American Marketing Association, Chicago, pp. 68–123.
Yoon, E., Guffey, H.J., Kijewski, V., 1993. The effects of information and company reputation on intentions to buy a business service. J. Bus. Res. 27 (3), 215–228.
Zeithaml, V.A., 1988. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. J. Mark. 52 (3), 2–22.