Reaching for the Stars:
Consumers' Interpretations of Online Rating Distributions and Their Validity as an Indicator of Product Quality

Doctoral Dissertation

Submitted to the
Faculty of Business and Economics
TU Dortmund University

In partial fulfillment of the requirements
for the degree of
Doctor rerum politicarum (Dr. rer. pol.)

by

Sarah Köcher (née Küsgen), M.Sc.
Email: sarah.koecher@tu-dortmund.de

Dissertation Committee:

Prof. Dr. Hartmut H. Holzmüller
Department of Marketing
TU Dortmund University

Prof. Dr. Jenny van Doorn
Department of Marketing
University of Groningen

Prof. Dr. Jan Wieseke
Sales & Marketing Department
Ruhr-University Bochum

Dortmund, August 2018

To Sören

Acknowledgements

Along the path of completing my doctoral dissertation, I have met many great people who supported me and who I truly enjoyed working with. First and foremost, I am deeply grateful to my supervisor, Professor Dr. Hartmut H. Holzmüller, who motivated, inspired, and supported me in every aspect of my journey. I am thankful for his constant encouragement, guidance, and faith in my work as well as for providing me the opportunity to present my research around the world and to meet wonderful people who have become co-authors and friends. Furthermore, I want to express my sincere gratitude to Professor Dr. Jenny van Doorn for her readiness to be the second reviewer of my doctoral dissertation as well as to Professor Dr. Jan Wieseke who agreed to be the third member of the committee.

Without the support of my colleagues and friends at the Department of Marketing, the journey of doing doctoral research would not have been as much fun. We have shared great conversations, team events, as well as conference trips, and have mentally supported each other. I thank Nicole Ahl-Saelbstaedt, Thorsten Autmaring, Dr. Gerrit Cziehso, Sabrina Heix, Andreas Kessenbrock, Xenia Raufeisen, Svenja Rebsch, Stefan Ruffer, Prof. Dr. Tobias Schäfers, Dr. Moritz vom Hofe, and Dr. Linda Wulf as well as our student assistants. My thanks also go to my colleagues and co-authors from around the world, in particular, to Dr. Jay Kandampully, Dr. Linda Alkire, and Dr. Arne De Keyser for their support since the very beginning of my academic career.

Outside of academia, I am truly blessed to have wonderful friends who have been part of my life for many years and who have supported me throughout the journey. I am thankful for all the joy they bring to my life outside the office. Special thanks go to my dear friends Daniela, Annette, and Marius for countless motivational conversations and emotional support as well as for spending quality time together and many, many pizza dates.

I am also deeply grateful to my family who has been there for me every single day of my life. They always believed in me, cheered for me, and took care of me. I am thankful for their unconditional love and support. My parents, however, are my biggest role models in life and I owe them more thanks than I could ever express in words. I would not know what I would do without them.

Finally, I want to express my deepest gratitude to Sören, my best friend, my husband, my love. You are the most generous, thoughtful, and loving person I know. Over the last years we have travelled the world, presented and published our work, stepped outside our comfort zones, overcome challenges, created memories with our wonderful family and friends, and said yes to share our whole life together. I cannot imagine my life without you and I can’t wait for everything that life holds for us.

Dortmund, August 2018 Sarah Köcher

Table of Contents

List of Figures

List of Tables

A Introduction

“Remember to look up at the stars and not down at your feet.
Try to make sense of what you see […]. Be curious.”
Stephen Hawking

1 Motivation and Purpose

Over the past two decades, rapid advances in technology and the omnipresence of the Internet have led to a fundamental change in our shopping behavior. While purchase behavior in traditional bricks-and-mortar stores is constricted by, for instance, limited retail spaces and finite opening hours, the Internet enables customers to shop anything, anytime, and anywhere. Moreover, while in the past, consumers were reliant on the quality of sales people’s advice or recommendations from their friends, they can now share their experiences and opinions about products, services, companies, and brands on a variety of websites such as Amazon, TripAdvisor, and Google with anyone. As a consequence, customers can easily access numerous online reviews at the click of a mouse. For example, TripAdvisor’s website offers more than 600 million reviews covering about 7.5 million accommodations, airlines, attractions, and restaurants, to 455 million unique users each month (TripAdvisor 2018).

One of the key factors responsible for the enormous popularity of online consumer reviews is that they are deemed highly credible and trustworthy (e.g., de Langhe, Fernbach, and Lichtenstein 2016a; Jiménez and Mendoza 2013; Park and Kim 2008; Schlosser 2011; Sen and Lerman 2007); despite the fact that they mostly stem from unknown strangers. For instance, according to a Nielsen (2015) study, 66 percent of participants indicated that they would trust in consumer opinions posted online. This percentage exceeded respondents’ trust in any form of communication initiated by a company (e.g., branded websites or TV ads). As a result, when making purchase decisions, people heavily rely on consumer reviews to infer the quality of the available purchase options (e.g., Hu, Liu, and Zhang 2008; Li and Hitt 2008; Simonson and

Rosen 2014) such that these consumer-generated evaluations became highly influential in driving sales and other performance metrics (see e.g., Floyd et al. 2014 and Babić Rosario et al. 2016 for a meta-analysis). Being recognized as a powerful tool to attract and retain customers (Dellarocas 2003; Schlosser 2011), the world's ten leading online retailers (NRF 2017) have implemented online review systems on their shopping websites.

Given their great popularity on both sides, customers and companies, a substantial body of research has been devoted to examining online reviews from diverse perspectives[1]. Numerous studies have focused on factors affecting their credibility (e.g., Banerjee, Bhattacharyya, and Bose 2017; Cheung and Thadani 2012), helpfulness (e.g., Schlosser 2011; Singh et al. 2017; Yin, Bond, and Zhang 2014), and usefulness (e.g., Casaló et al. 2015; Cheng and Ho 2015; Mudambi and Schuff 2010). Investigating consumer reviews on a more aggregate level, a great deal of literature is centered toward consumers' response to different characteristics describing the distribution of rating scores, including, for instance, average product ratings (e.g., Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; Godes and Mayzlin 2004), the dispersion of rating scores (e.g., He and Bond 2015; Sun 2012; Zhang 2006), and rating volume (e.g., Liu 2006; Moe and Trusov 2011; Zhu and Zhang 2010). However, despite the abundance of research on consumers' reactions to different characteristics of rating distributions our knowledge of the effects caused by further distribution features (e.g., mode, median, skewness, etc.) is still limited.

Furthermore, several researchers have questioned whether the proliferation of online consumer reviews should be considered a positive development from a consumer welfare perspective by investigating if the evaluations posted online can actually reflect the 'true' quality of a product (e.g., de Langhe et al. 2016a; Hu, Pavlou, and Zhang 2006; Koh, Hu, and

[1] This development is also reflected in the research priorities announced by the Marketing Science Institute (MSI) classifying research related to how social media and digital technology change customer experiences and the consumer path to purchase as a tier one priority (MSI 2014) as well as calling for research on changing decision making processes in times when consumers are always connected (MSI 2016).

Clemons 2010). In this vein, de Langhe et al. (2016a) reported a substantial gap between the extent to which consumers trust in average ratings when making inferences about the quality of a product and the actual validity of such ratings as an indicator of a product's 'objective' performance. However, factors determining the relationship between average ratings and more objective measures of product quality (e.g., Consumer Reports and Stiftung Warentest quality scores) have remained unexplored.

Aimed at addressing these gaps in the literature, the purpose of this thesis is twofold. First, it intends to improve our knowledge regarding consumers' interpretations of different characteristics of online rating distributions by investigating the effects of the mode—a previously disregarded distribution feature—on consumers' inferences about product quality. Second, the current work aims to generate a better understanding of the validity and relevance of consumer ratings as an indicator of a product's quality by examining how the convergence between average ratings and objective measures of product quality alters over a product's life cycle as well as how both rated and objective quality jointly affect a product's sales performance.

The contribution of this doctoral thesis is of equal relevance from both perspectives, managerial as well as theoretical. Deemed as an issue of strategic importance, marketers need to understand the way consumers use online reviews as decision aids as well as their impacts on sales and other related performance figures (Kumar, Choi and Greene 2017; Wilson, Giebelhausen, and Brady 2017). Complementing extant knowledge about the consequences of different rating distribution characteristics, this research establishes the mode of rating distributions as an important parameter in consumers' product quality inferences and, thereby, offers marketers a new measure that should be involved when examining online review data. In addition, by shedding light on the convergence between online ratings and measures of objective product performance, this thesis gives advice when and why consumers should be rather reluctant in their use of online ratings as a quality indicator.

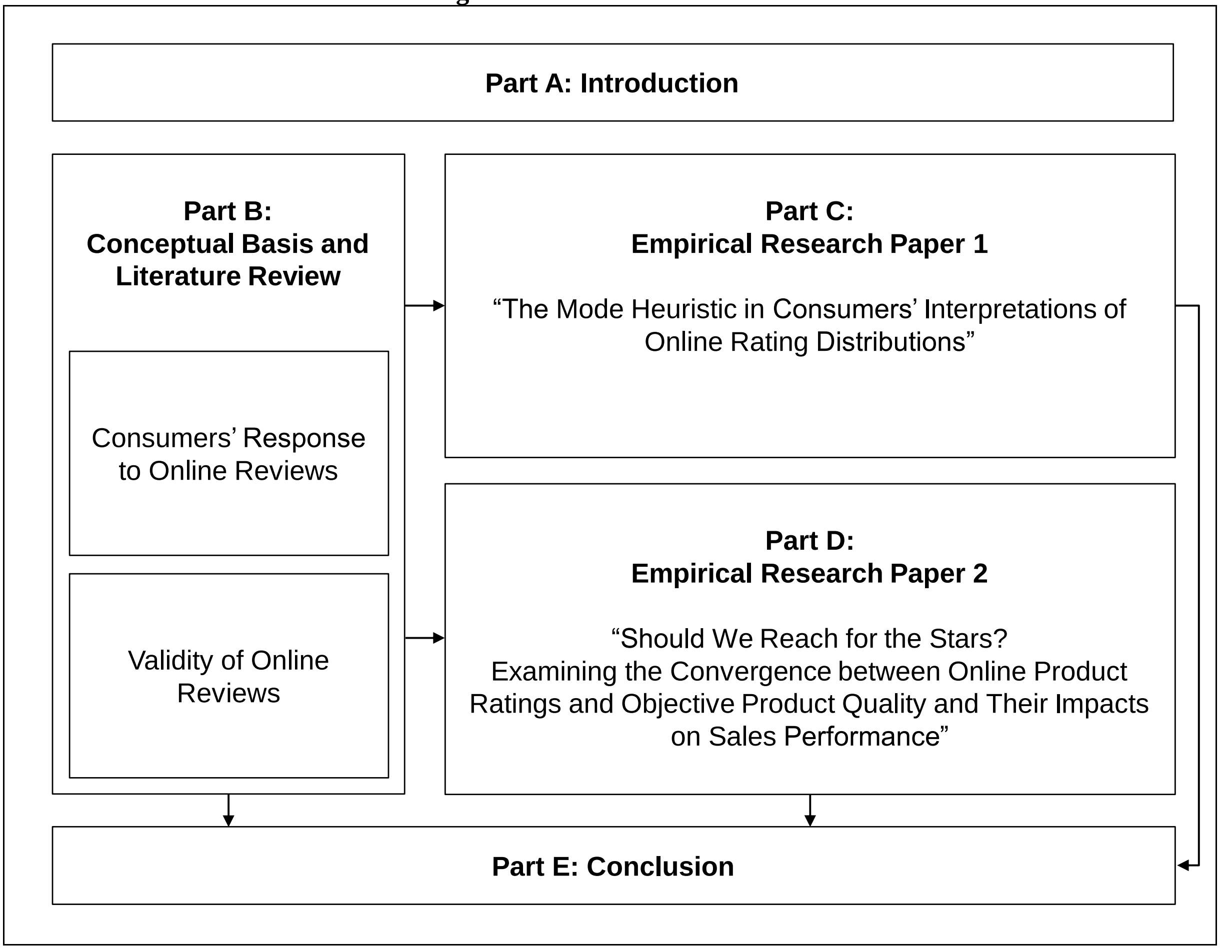
2 Outline of this Thesis

The current doctoral dissertation is subdivided into five parts organized as follows: Following the introduction in *part A*, *part B* provides the conceptual basis for this thesis. It comprises a definition of online consumer reviews and a literature review that structures prior studies in this field by identifying two major research streams; namely, research on (1) the effects of individual review and reviewer characteristics as well as studies on (2) the impacts of rating distribution characteristics. Additionally, several issues threating the validity of online reviews as a quality indicator will be discussed. This part concludes with a summarizing synthesis of the previous literature in the scholarly field and a description of the conceptual positioning of this thesis.

Part C and *part D* represent two empirical research papers—entitled (1) “The Mode Heuristic in Consumers’ Interpretations of Online Rating Distributions”, and (2) “Should We Reach for the Stars? Examining the Convergence between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance”. The focus of these manuscripts as well as their unique features will be described in the following two subchapters in more detail.

Finally, the concluding *part E* contains a summary of the major findings of the presented papers, a discussion of their theoretical contributions to different areas of research, as well as managerial implications for marketers and recommendations for consumers. A critical review of limitations and suggestions for further research conclude this thesis. Figure 1 summarizes the overall outline of this doctoral thesis.

Figure 1. Outline of this Thesis



2.1 Focus of Empirical Research Paper 1

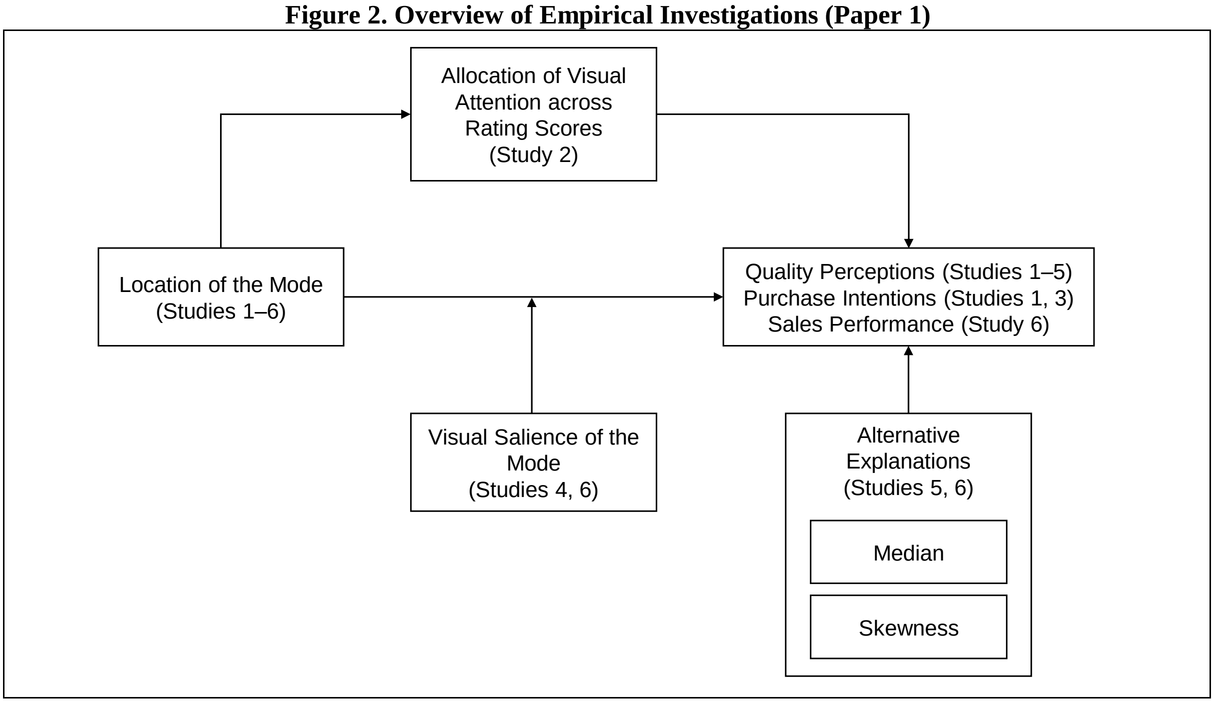
A very common practice to inform online shoppers about product evaluations from previous customers is to illustrate the distribution of rating scores through graphical formats; typically via bar charts, wherein each bar represents the number of votes a specific rating score has received. The first paper in this thesis entitled "The Mode Heuristic in Consumers' Interpretations of Online Rating Distributions" is concerned with an investigation of people's use of the mode—i.e., the rating score that has received the largest number of votes which is, thus, the most salient element of a bar chart describing the distribution of rating scores—when drawing quality inferences from such visualizations; thereby, this research adds to prior studies on consumers' response to different characteristics of rating distributions (e.g., average ratings, dispersion of rating scores, and rating volume).

Across a series of six studies, this research demonstrates a tendency to use the mode as a heuristic basis when making product inferences from online rating distributions in such a way that product evaluations inferred from rating distributions with an equal average, standard deviation, and number of ratings systematically vary by the location of the mode; a phenomenon referred to as the mode heuristic. Specifically, using a mix of experimental and real-world data, this research provides strong empirical evidence for the existence of the mode heuristic, sheds light on this phenomenon at the process level, and demonstrates how product inferences based on the mode heuristic depend on the visual salience of the mode. Thereby, the first paper presented in this thesis answers the following research questions:

(1) *How are consumers' inferences about the quality of a product affected by the location of a rating distribution's mode?*
(2) *What is the process underlying the relationship between the location of the mode and quality inferences?*
(3) *Which factors determine the relationship between the location of the mode and quality inferences?*

Figure 2 provides an overview of the relationships put under scrutiny in this manuscript.

Figure 2. Overview of Empirical Investigations (Paper 1)



To ensure the robustness and generalizability of the generated insights, the studies reported in this paper cover a variety of different settings (i.e., products and services) and employ diverse methods of data collection; ranging from questionnaire-based online experiments (Studies 1, 3, 4, and 5) over an eye-tracking study (Study 2) to an extraction of consumer reviews from Amazon's website (Study 6). Overall, 911 subjects took part in the experimental studies, while the data set collected from Amazon contains review information about a total of 1,536 products. Table 1 summarizes the contexts, data collection methods, and sample sizes of the studies documented in Paper 1.

Table 1. Summary of Contexts, Methods, and Samples (Paper 1)

	Context	Method	Sample Size (N)
Study 1a Study 1b	Printers Fast Food Restaurants	Experiments	65 78
Study 2	Toasters	Eye-Tracking Experiment	54
Study 3a Study 3b	Printers Fast Food Restaurants	Experiments	67 92
Study 4a Study 4b	Fitness Tracker Movies	Experiments	140 129
Study 5a Study 5b	Electric Water Kettles Lecture Evaluations	Experiments	138 148
Study 6	Consumer Electronics	Amazon Data	1,536

2.2 Focus of Empirical Research Paper 2

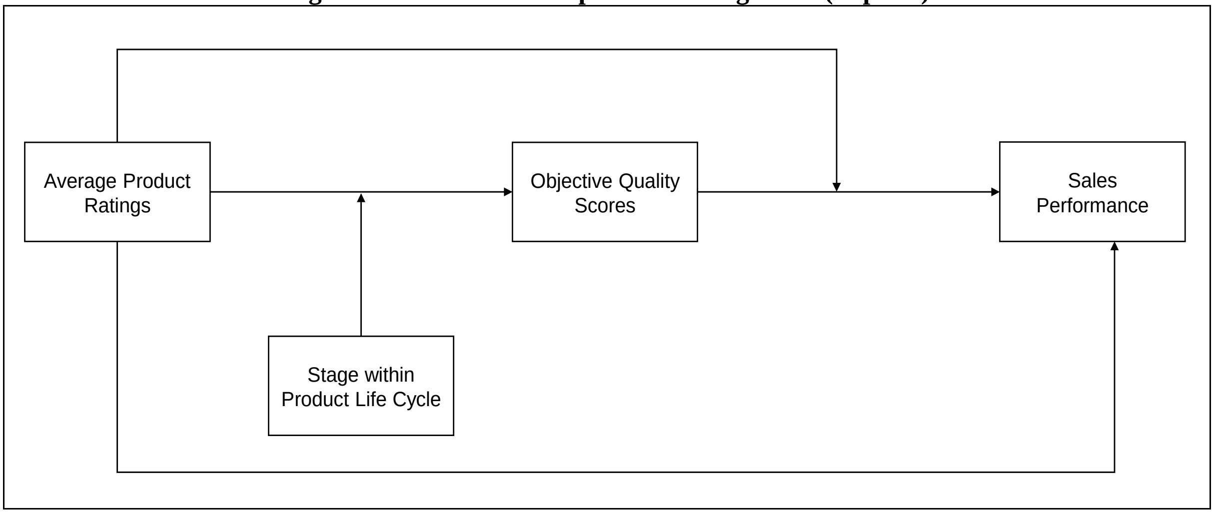
By demonstrating that average product ratings poorly correlate with quality scores provided by Consumer Reports—presumably a measure of 'objective' product quality—de Langhe et al. (2016a) found that consumers rely more heavily on such ratings when making quality inferences than they should. These findings have caused a lively discussion among several eminent researchers (de Langhe et al. 2016b; Kozinets 2016; Simonson 2016; Winer and Fader 2016) primarily questioning the reliability of Consumer Report scores as a measure of objective quality as well as the simplicity of analysis neglecting dynamic changes in consumer product ratings over time and, thereby, yielding a variety of worthwhile research opportunities.

Using a unique data set that unites all consumer electronic products that have been evaluated by Stiftung Warentest—the German equivalent of Consumer Reports—from the years 2014 to 2017 (i.e., 2,473 products) with review data of those items that were available on Amazon's German website (i.e., 1,833 products), the analyses documented in the second paper in this thesis entitled "Should We Reach for the Stars? Examining the Convergence between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance" replicate and extend de Langhe et al.'s (2016a) findings in several important ways. The obtained findings demonstrate that the convergence between average product ratings and objective quality scores varies over a product's life cycle and that the extent to which average ratings actually influence sales is surprisingly small when being benchmarked against the impact of objective performance. However, this paper also reveals that the relationship between objective quality and sales performance attenuates when average ratings increase; implying that high consumer ratings may be able to disguise a product's objective quality to some degree. In summary, the second paper responds to the following questions:

(1) *Is the average product rating an adequate indicator of a product's 'objective' performance?*
(2) *Does the convergence between rated and objective quality change over the product life cycle?*
(3) *What is the better predictor of sales performance, product ratings or objective quality scores?*

Figure 3 summarizes the investigated relationships in Paper 2.

Figure 3. Overview of Empirical Investigations (Paper 2)



B Conceptual Basis and Literature Review

The purpose of this section is to provide the conceptual foundation for the two empirical papers presented in this doctoral dissertation. Chapter 1 defines the concept of online consumer reviews. The subsequent chapter 2 reviews extant literature on the impacts of different facets and features of online reviews on decision making processes comprising (1) prior research on individual review and reviewer characteristics as well as (2) studies on consumers' response to characteristics describing the distribution of rating scores. Additionally, several aspects curtailing the validity of online reviews as a measure of the 'true' quality of a product or service will be discussed. This chapter concludes with a summarizing synthesis and the conceptual positioning of this thesis.

1 The Concept of Online Consumer Reviews

The tradition of engaging in word-of-mouth (WOM) probably dates back to the time that human beings started to communicate with each other (Simonson 2016). In this early phase, they may have already exchanged information and recommendations about necessities, basic needs, and threats. As it evolved, WOM has become a powerful force in consumers' shopping behavior (e.g., Herr, Kardes, and Kim 1991; Park and Kim 2008; Schlosser 2011). By sharing experiences, feedback, and opinions about products and services in person, consumers engaging in WOM provide an unpaid endorsement for the item under consideration which has been shown to be the most credible and trustworthy source of "advertisement" for companies (Nielsen 2015; Henricks 1998).

With the advent of the Internet, the traditional form of face-to-face WOM has been transferred to the online environment and electronic word-of-mouth (eWOM)—a less personal but still highly pervasive form of WOM—has come into vogue (e.g. Dellarocas 2003; Godes and Mayzlin 2004; Hennig-Thurau et al. 2004). From a conceptual perspective, eWOM refers to “*any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet”* (Hennig-Thurau et al. 2004, p. 39). In contrast to traditional WOM, eWOM is more accessible (Bakos and Dellarocas 2011; Duan, Gu, and Whinston 2008) and persistent (Dellarocas et al. 2007, Sen and Lerman 2007; van Doorn et al. 2010), and it usually does not involve a direct personal connection between sender and receiver; instead, it typically originates from unknown people and is read by an anonymous audience (Hennig-Thurau, Wiertz, and Feldhaus 2015).

Consumers can engage in eWOM in several manners including, for instance, through online reviews, tweets, online communities, or blog posts (Babić Rosario et al. 2016; Cheung and Thadani 2012). Online consumer reviews, often considered as one of the most widespread and prominent forms of eWOM (Gottschalk and Mafael 2017; Jiménez and Mendoza 2013; Kostyra et al. 2016), can be described as “*peer-generated product evaluations posted on company or third party websites*” (Mudambi and Schuff 2010, p. 186). They typically comprise textual elements, containing a written evaluation of customers’ usage experience, and a numerical rating representing an overall judgment of the reviewed item (e.g., Chatterjee 2001; Jiménez and Medonza 2013; Schlosser 2011; Villarroel Ordenes et al. 2017). Consumers predominantly consult these reviews to obtain quality information in order to reduce perceived purchase risks as well as to enhance decision confidence and precision (e.g., Mudambi and Schuff 2010; Zhu and Zhang 2010). In particular, the fact that review information is highly accessible and, thus, empirical data is relatively easy to retrieve, the study of online consumer reviews enjoys great popularity among academics. The subsequent sections give an overview of extant literature in this area.

2 The Impact of Online Consumer Reviews on Decision Making Processes

Over the last 15 years, a large body of research has advanced the understanding of the impact of online consumer reviews on decision making processes and other purchase-related behaviors. The purpose of this chapter is to provide an overview of extant literature and to depict the major research streams in more detail[2].

On closer examination, previous research in this area can be roughly classified into two streams depending on the level of abstraction: first, studies focusing on individual reviews and their characteristics and, second, those considering online reviews on a more aggregate level (Sridhar and Srinivasan 2012). Research on online reviews on an individual level is typically centered toward the effects of review features (e.g., review length, rating valence, or consistency between arguments and rating) and reviewer characteristics (e.g., experience and expertise) on consumers' perceptions of helpfulness, usefulness, and credibility as well as on information adoption and resulting behavioral intentions. In contrast, studies concentrating on online reviews from an aggregate perspective usually employ different measures summarizing and describing the distribution of ratings scores (e.g., average ratings, standard deviation and variance, or the number of ratings a product has received) and investigate consumers' response to such distribution characteristics.

Overall, although researchers have addressed various aspects of online consumer reviews, the primary focus of extant literature is on the impacts of aggregate measures of rating distributions rather than on individual review elements. Nonetheless, the following section aims

[2] The literature has been identified by conducting manual searches of the leading marketing journals (Journal of Marketing Research, Journal of Marketing, Journal of Consumer Research, Marketing Science, International Journal of Research in Marketing, Journal of the Academy of Marketing Science, Journal of Consumer Psychology, Journal of Retailing, and Journal of Service Research) for articles since 2005. This period has been chosen according to a recent classification of past research on digital and social media marketing by Lamberton and Stephen (2016). Furthermore, to also incorporate related research disciplines and prior work, relevant articles from current meta-analyses (e.g., Babić Rosario et al. 2016; Floyd et al. 2014; You, Vadakkepatt, and Joshi 2015) and literature overviews (e.g., Cheung and Thadani 2012; Kostyra et al. 2016) were included. In addition, keyword searches of electronic databases, such as Google Scholar and EBSCOhost Business Source Premier, have been applied.

to give a brief overview of prior research on several dimensions of individual reviews and their impacts on consumer behavior.

2.1 Research on Individual Review and Reviewer Characteristics

When posting an online review, consumers are typically completely free in describing and evaluating the purchased item (Jiménez and Mendoza 2013). These written evaluations comprise individual perceptions and preferences which may differ between reviewers of one and the same object (Zhu and Zhang 2010). For example, based on their consumption experiences, consumers tend to weight the importance of various product features differently and build their reviews accordingly resulting in diverse, sometimes contradicting, product descriptions and assessments. Furthermore, not only tastes but also acuity, expertise, and writing styles may vary across reviewers. As a result, each online review contains critical information about both the object under consideration as well as about the reviewer himself (Moon and Kamakura 2017). Hence, prior research studying online consumer reviews on an individual level has focused on a variety of different review and reviewer characteristics. A selection of the aspects that have been investigated is provided in Table 2.

Table 2. Previous Research on Review and Reviewer Characteristics

Review(er) Characteristics	Selected Outcomes	Exemplary Literature
Review Characteristics (e.g., review length, review valence, acuity, writing styles, language use, recency) Reviewer Characteristics (e.g., level of expertise, number of followers, experience)	Helpfulness Usefulness Credibility Behavioral Intentions Information Adoption Evaluation Sales Performance	Cheng and Ho (2015); Cheung and Thadani (2012); Filieri (2015); Jimenéz and Mendoza (2013); Jin, Hu, and He (2014); Moon and Kamakura (2017); Mudambi and Schuff (2010); Packard and Berger (2017); Schlosser (2011); Villarroel Ordenes et al. (2017)

Although quantitative rating scores might be the most salient component of an online review, past research has shown that consumers also pay attention to the textual elements rather than to solely consider these numerical judgments (Chevalier and Mayzlin 2006). Nonetheless, relatively few studies have concentrated on the written contents, presumably due to the high effort involved in measuring and analyzing it (Godes and Mayzlin 2004; Schlosser 2011). As illustrated in Table 2, the majority of studies in this context has examined the impact of review and reviewer characteristics on perceived credibility (e.g., Banerjee et al. 2017; Cheung and Thadani 2012), usefulness (e.g., Casaló et al. 2015; Cheng and Ho 2015; Mudambi and Schuff 2010; Sen and Lerman 2007), and helpfulness (e.g., Schlosser 2011; Singh et al. 2017) as well as on information adoption (e.g., Filieri 2015), product evaluations (e.g., Huang et al. 2016; Kim and Gupta 2012; Villarroel Ordenes et al. 2017), purchase decisions (e.g., Jimenéz and Mendoza 2013; Reimer and Benkenstein 2016), and sales performance (e.g., Fan, Che, and Chen 2017). Unsurprisingly, previous research has shown that review valence is positively related to consumers' purchase intentions (e.g., Purnawirawan et al. 2015; Reimer and Benkenstein 2016; Tsang and Prendergast 2009). This effect, however, has been found to depend on the trustworthiness of review information (Reimer and Benkenstein 2016). Interestingly, Wilson et al. (2017) demonstrated that even negative reviews can increase purchase intentions of consumers with a high self-brand connection; e.g., when the brand personalities are similar to the consumer's ones or when the products are purchased by a peer-group to which an individual aspires. Further research concerned with the valence of rating scores has indicated that extreme ratings are associated with lower levels of helpfulness (Mudambi and Schuff 2010), while other studies have shown that negative reviews are perceived as more helpful than positive reviews (e.g., Casaló et al. 2015; Chevalier and Mayzlin 2006; Yin et al. 2014).

Aside from review valence, previous research has revealed that review length (e.g., Cheng and Ho 2015; Mudambi and Schuff 2010) and the provision of images (Cheng and Ho 2015) can enhance helpfulness and usefulness perceptions. Interestingly, the positive effect of review depth is greater for search goods than for experience goods (Mudambi and Schuff 2010).

Similarly, Cheung and Thadani (2012) discovered that review quality has a positive impact on consumer’s information adoption decision. In a similar vein, Jiménez and Mendoza (2013) reported that a positive review increases purchase intentions more when it is more detailed. Examining the recency of online reviews, Jin et al. (2014) found that recent consumer reviews are more influential than out-dated reviews in near-future purchase decisions and the opposite was true when considering distant-future decisions. Finally, with regards to reviewer characteristics, among others, perceived expertise (e.g., Casaló et al. 2015; Cheng and Ho 2015) and number of followers (e.g., Cheng and Ho 2015) have been demonstrated to increase perceived usefulness. Likewise, perceptions of source credibility have been found to spill over to eWOM credibility (Cheung and Thadani 2012).

Against the background that the two manuscripts presented in this thesis both focus on characteristics describing the distribution of rating scores rather than on individual review features, the following chapter 2.2 is exclusively dedicated to review prior literature that has examined online consumer reviews from an aggregate perspective and structures previous studies in this specific research domain according to the rating distribution characteristics under investigation.

2.2 Research on Rating Distribution Characteristics

Just as any other distribution, distributions of consumer ratings can be summarized by a variety of descriptive statistics, such as frequencies of rating scores, measures of location (e.g., mean, mode, and median), measures of variability (e.g., standard deviation and variance), as well as measures of shape (e.g., skewness and kurtosis). As a consequence, rating valence (i.e., average ratings), rating volume (i.e., the number of ratings an object has received), and rating dispersion (i.e., the variation in ratings along the rating scale which represents the heterogeneity among consumers’ evaluations) have become central considerations in numerous studies. In this context, empirical studies have shown that these aggregate measures are meaningful predictors of sales and other relevant performance figures (Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Dellarocas, Awad, and Zhang 2004; Hu et al. 2008; see also Floyd et al. 2014 and Babić Rosario et al. 2016). However, when it comes to the extent and direction of the analyzed relationships, findings are often inconsistent and sometimes even contradictory. The following subsections aim to summarize the investigated effects organized around consumers' response to these three distribution characteristics.

2.2.1 Rating Valence

Most of the past research concentrates on the consequences of rating valence revealing that higher average product ratings are associated with favorable reactions reflected in, for instance, higher purchase intentions, better sales ranks, revenues, product choice probabilities, and even future ratings (e.g., Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Dellarocas et al. 2007; Godes and Mayzlin 2004; Kosytra et al. 2016; Li and Hitt 2008; Moe and Trusov 2011; Luca 2011; see Table 3 for an overview). The most prevalent, and intuitive, explanation for this finding is that customers' rely on average ratings as an indicator of product quality such that higher mean values imply greater quality and, thus, enhance attitudes toward the reviewed product or service which carry over to subsequent purchase decision making (e.g., Sun 2012; de Langhe et al. 2016a; Liu 2006). A few studies, however, did not find support for this relationship (Chen, Wang, and Xie 2011; Duan et al. 2008; Liu 2006). For instance, Liu (2006) argued that attitudes might not always predict behavior well (Ajzen and Fishbein 1980); instead, situational and contextual factors may influence behavior beyond what attitudes can explain.

Furthermore, research has also shown that negatively valenced online ratings are more influential than positive rating scores (Chen et al. 2011; Chevalier and Mayzlin 2006; Ludwig et al. 2013; see also Hennig-Thurau and Walsh 2003) such that the (negative) impact of one-star reviews on sales is significantly higher than the (positive) effect of five-star ratings. The reason why negative eWOM information is more influential may derive from prospect theory's loss aversion principle (Kahneman and Tversky 1979); suggesting that expected losses are weighted more heavily than anticipated gains in the same amount.

Table 3. Previous Research on the Impacts of Rating Valence

Type and Direction of Effects	Selected Argumentations	Supporting Literature
Main Effects		
Positive Effect	Average ratings serve as an indicator of quality such that higher average ratings suggest greater quality of a product or service.	Chen, Liu and Zhang (2011); Chevalier and Mayzlin (2006); Chintagunta et al. (2010); Dellarocas et al. (2007); Kostyra et al. (2016); Moe and Trusov (2011); Zhang and Dellarocas (2006); Zhu and Zhang (2010)
No Effect	Attitudes toward a product or service formed on the basis of average ratings may not always carry over to behavioral reactions and, thus, may not necessarily transform into sales.	Chen et al. (2011); Duan et al. (2008); Liu (2006)
Moderators		
Variance	Low (as compared to high) variance strengthens the quality signal emanating from review valence, rendering highly rated products even more attractive and low-rated products more unattractive.	Kostyra et al. (2016)
Volume	The positive effect of valence is stronger with an increasing review volume, since an increase in the number of ratings entails a greater persuasiveness, diagnosticity, and predictive power of average ratings.	Khare, Labrecque, and Asare (2011); Kostyra et al. (2016)
Brand Strength/ Brand Equity	The positive effect of valence is stronger for weak brands than for strong brands, because weak brands (as compared to strong brands) lack a credible quality signal.	Ho-Dac, Carson, and Moore (2013); Luca (2011)
Product Popularity	Online reviews are more influential for less popular products because consumers are more likely to consult them to attain quality information since other information sources are rare.	Berger, Sorensen, and Rasmussen (2010); Zhu and Zhang (2010)

Previous research has also attempted to identify factors that determine the strength of the positive effects of rating valence. In this context, prior literature suggests that the impacts of average product ratings depend on other distribution characteristics. For instance, Khare et al. (2011) proposed that the number of ratings increases diagnosticity and persuasiveness of the average. In other words, the more people express their opinion about an object the higher should be the correctness of this measure and, thus, its influence on consumers' preferences (see also Kostyra et al. 2016). In addition, Kostyra et al. (2016) demonstrated that the dispersion of ratings negatively moderates the impact of high and medium valenced ratings. Further research has revealed that the effect of rating valence is moderated by brand strength, in a way that average ratings affect sales only for weak brands but not for strong brands (Ho-Dac et al. 2013; Luca 2011). For instance, Luca (2011) found that average customer ratings of small non-chain restaurants on the recommendation website *Yelp!* had a positive impact on their revenues, but this effect was absent for large restaurant chains; resulting in a shift in revenue share toward independent restaurants, away from those restaurants with chain affiliation. Hence, average ratings seem to serve as a substitute for traditional information sources rather than a complement by, for example, curtailing the relevance of brands (Chen, Dhanasobhon, and Smith 2008; Luca 2011; Zhu and Zhang 2010); traditionally one of the key criteria when assessing the quality of a specific product or service (e.g., Jacoby, Olson, and Haddock 1971). Similarly, the effect of review valence is also known to vary depending upon the popularity of the product under consideration. In this vein, Zhu and Zhang (2010) found that average ratings are more influential for less popular products (see also Berger et al. 2010).

2.2.2 Rating Volume

As already mentioned, rating volume refers to the number of ratings an object has received. It has been argued that the direct effect of rating volume can be attributed to a higher likelihood that other online shoppers will become aware of the reviewed object as number of ratings increases (Dellarocas et al. 2007; Duan et al. 2008; Liu 2006) which, thus, generates greater sales (e.g., Godes and Mayzlin 2004; Liu 2006). However, the relationships between the number of ratings and sales might be more complex. In this regard, Duan et al. (2008) argued that rating volume and sales might be interdependent, such that the number of ratings may not only be an antecedent, but also an outcome of sales performance (see also Moe and Trusov 2011).

Although most extant studies account for the described endogeneity problem, findings on this distribution characteristic are still mixed. Whereas a large number of previous studies have documented a positive influence of rating volume on, for instance, sales performance (e.g., Chevalier and Mayzlin 2006; Duan et al. 2008; Moe and Trusov 2011; Sun 2012) and box office revenues (Liu 2006), others have not found a significant relationship (Chintagunta et al. 2010; Clemons et al. 2006; see Table 4).

Table 4. Previous Research on the Impacts of Rating Volume

Type and Direction of Effects	Selected Argumentations	Supporting Literature
Main Effects		
Positive Effect	As increasing number of ratings increases the likelihood other consumers become aware of the reviewed object.	Chevalier and Mayzlin (2006); Dellarocas et al. (2007); Duan et al. (2008); Li and Hitt (2008); Liu (2006); Moe and Trusov (2011); Sun (2012)
No Effect	-	Chintagunta et al. (2010); Clemons et al. (2006)

2.2.3 Rating Dispersion

Although average ratings and rating volume might be the most salient distribution characteristics, customers also attend to the degree of consensus among reviewers' evaluations. Therefore, several researchers have focused on the consequences of the variance and standard deviation of rating distributions as statistical measures for said heterogeneity.

In general, a low variability of ratings implies that reviewers strongly agree with each other, turning product inferences from slightly scattered rating distributions into a straightforward task; using the average rating as a cue should lead to a result at least close to the ‘real’ product quality. However, a high variability in ratings could be double edged. On the one hand, it has been argued that a high rating dispersion creates a higher degree of uncertainty as it entails a greater risk of misjudging a product’s actual performance evoking rather cautions consumer reactions (e.g., Hu et al. 2010; Langan, Besharat, and Varki 2017; Moon, Bergey, and Iacobucci 2010; Zhu and Zhang 2010). Consumers may also interpret such a high heterogeneity in ratings as an indicator that the product is a niche product delighting some people and disappointing others (Sun 2012). On the other hand, it has been proposed that a high variance of ratings may actually help customers to reduce the risk associated with purchase decisions because it draws people’s attention to both positive and negative facets of a product, which should entail even higher sales (Lu, Ye, and Law 2014). Although the two described explanatory approaches implicate contradicting predictions about consumers’ response to rating dispersion, they both find empirical support (see Table 5).

Table 5. Previous Research on the Impacts of Rating Dispersion

Type and Direction of Effects	Selected Argumentations	Supporting Literature
Main Effects		
Positive Effect	Highly dispersed product ratings help customers to reduce the risk associated with purchase decisions because they draw consumers’ attention to both positive and negative facets of a product, which entails a better sales performance.	Bao and Chang (2014); Clemons et al. (2006); Lu, Ye, and Law (2014); Moe and Trusov (2011)
Negative Effect	An increasing variation in ratings increases outcome uncertainty and the risk of misjudging the quality of a product.	Hu et al. (2010); Moon et al. (2010); Zhu and Zhang (2010)
No Effect	-	Chen et al. (2011); Chintagunta et al. (2010); Ye et al. (2011); Zhang (2006)

Table 5. *(continued)*

Type and Direction of Effects	Selected Argumentations	Supporting Literature
Moderators		
Average Rating	If the average rating is high, consumers are already confident of the product's quality. In this case, a high variance implies that some customers hold a contrary, negative opinion about the product, which harms its evaluation. In contrast, if the average rating is low, a higher dispersion improves consumers' perception of the product's quality.	Khare et al. (2011); Sun (2012); see also West and Broniarczyk (1998)
Volume	An increasing rating variance decreases preferences for products with high average ratings because a low consensus contrasts preferences away from positive evaluations. Because an increase in the number of ratings grants greater credibility, this effect is stronger when rating volume is high; the opposite holds true for unfavorable valence.	Khare et al. (2011)
Taste Similarity (Product Category)	The negative effect of dispersed rating distributions is attenuated in product domains where tastes are perceived to be dissimilar, since disagreement among reviewers can be attributed to heterogeneous preferences rather than to the product.	He and Bond (2015)
Product Nature (Hedonic vs. Utilitarian)	Relative to utilitarian products, hedonic products are more immune to the risks associated with higher levels of review variance due to variability in the subjective experiences inherent with the use and consumption of hedonic products.	Langan et al. (2017)
Product Type (Experience vs. Search Good) and Consumers' Prior Expectations	Consumers discount extreme product reviews that are not consistent with their prior expectation and prefer high variance to low variance product reviews when evaluating experience products (vs. search products).	Park and Park (2013)

A few studies, however, provide explanations for these inconsistent findings by identifying moderators of the effects of rating dispersion. In this context, prior literature suggests that the effects of rating dispersion might be reference-dependent, such that the consequences of an increasing variance could be essentially determined by a reference value represented by the average value of a rating distribution (Khare et al. 2011; Sun 2012; see also West and Broniarczyk 1998). It has been argued that if a product's average rating is high, consumers are already confident about the item's quality. In this case, however, a high variance implies that some customers hold a contrary, negative opinion about the product, which should harm people's evaluation of the reviewed product. In contrast, if the average rating is low, a higher dispersion can only enhance consumers' perception of the product under consideration (Sun 2012). In a similar vein, Khare et al. (2011) reported that an increasing variance diminishes preferences when the average rating is high, but only if rating volume is high as well; they argued that the effect of consensus among reviewers is conditioned by the credibility of assessments that comes along with higher rating volume. Furthermore, building on attribution theory, He and Bond (2015) proposed that consumers' response to dispersion in online ratings depends on their inferences about the causes for diverging product evaluations and suggested that perceptions of taste similarities within a product class determine whether rating dispersion is attributed to disagreement in reviewer preferences rather than to the product itself. Accordingly, the authors found that consumers were more tolerant to dispersion in taste-dissimilar product domains (e.g., paintings or music albums) than taste-similar product categories (e.g., desk lamps or flash drives). Langan et al. (2017) found another explanation for diverging impacts of the rating dispersion. They argued that the effect of the variance in online ratings on purchase intentions depends on product nature (i.e., hedonic vs. utilitarian products) proposing that a greater review variance entails greater purchase intentions for hedonic compared to utilitarian products. The authors suggested that hedonic products may be more immune to the risk of decision uncertainty associated with higher levels of variance in rating scores. Finally, Park and Park (2013) found that an increasing variance in reviews diminishes consumers' evaluations of products for which they have unfavorable prior expectations. However, considering high expectation products, the effect of rising variance is dependent on product category in such a manner that product judgments enhance for experience products and decrease for search products.

2.3 Validity of Online Reviews

The previous chapter gave a summary of the substantial evidence regarding consumers' use of online ratings when making purchase decisions in a variety of different contexts. Building on these findings, more recent research has started to question whether online ratings can actually depict the 'true' quality of a product (e.g., de Langhe et al. 2016a; Hu et al. 2006; Koh et al. 2010). For instance, de Langhe et al. (2016a) reported a considerable disconnect between the extent to which consumers trust in online consumer ratings when making inferences about the quality of a product and the actual validity of such ratings as an indicator of a product's 'objective' performance. Across a series of consumer studies, the authors found that people place enormous weight on average product ratings when assessing the quality of a product, while the convergence between average ratings and the quality scores provided by Consumer Reports—as a measure of objective quality—and, thus, their validity as a quality indicator, is evidentially weak. The following sections provide an overview of several reasons why online ratings, and, in particular, average ratings, might be a rather imprecise predictor of a product's quality; namely, statistical, sampling, and evaluation issues (de Langhe et al. 2016c).

From a statistical perspective, the representativeness and, thus, explanatory power of a mean, such as the average product rating, can be assessed using its standard error[3]. Consequently, the statistical precision of an average rating is a function of sample size and variability in rating scores. In other words, the accuracy of average ratings increases with the number of customers who have left a review and with their agreement in their evaluations. In general, average ratings should converge toward a 'true' value as the number of ratings ascends (Ho-Dac et al., 2013; Zhu and Zhang 2010). Unfortunately, typically not all customers who bought a product provide a review such that sample sizes are often not sufficiently large from a statistical standpoint (de Langhe et al. 2016c). As a consequence, the average rating from this sample does not perfectly

[3] Formally, the standard error of a distribution's mean is defined as

$$\text{standard error} = \text{standard deviation} / \sqrt{\text{sample size}}.$$

match with the mean value that would have been obtained if all customers had evaluated the item. Rating dispersion, on the other hand, tends to be high for a variety of reasons (see also sampling issues). For instance, some customers may adopt more extreme opinions in order to "correct" the mean rating to be closer to their own (Duan et al. 2008; Matakos and Tsaparas 2016). Similarly, 'fake' reviews—i.e., the phenomenon that companies incentivize people to post fake reviews praising the products they market or bad-mouthing those of their competitors (e.g., Dellarocas 2006; Mayzlin 2006; Mayzlin, Dover, and Chevalier 2014; Zhao et al. 2013)—tend to be more extreme (i.e., favorable or unfavorable; Luca and Zervas 2016; Malbon 2013) and, thus, enhance rating dispersion. Other reasons for a high variance in rating scores include taste-dissimilarities among reviewers (He and Bond 2015) and a random noise (de Langhe et al. 2016c). For example, reviewers may accidentally rate the wrong product or may leave a low rating to vent their anger about aspects of a transaction with no direct relationship to the product itself (e.g., speed of shipping, shipping damages, or invoicing). Finally, considering online ratings more formally, the rating scales that are typically used by marketers are ordinal scales; e.g., they range from excellent (over very good, average, poor) to terrible (Tripadvisor.com) or from five star to one star ratings. Strictly speaking, because the assumption of equal distances between categories may not hold for such ordinal scales[4], the use of mean values to describe rating distributions might not be appropriate. Instead, positional measures like the mode, median, and percentiles are recommended to be used (e.g., Hair et al. 2010). In other words, average ratings may simply not reflect a product's quality because the calculation of the mean is not a valid operation for ordinal data.

Second, sampling issues result from the fact that the subsample of customers who leave a review is usually not representative of the entire population of customers who have purchased the product (e.g., de Langhe et al. 2016c; Askalidis, Kim, and Malthouse 2017). In this vein, in

[4] In other words, the difference between „excellent" and „very good" may not be equivalent to the difference between „average" and „poor".

line with research on traditional WOM (e.g., Anderson 1998), it has been demonstrated that consumers are more likely to contribute a review when they are either very satisfied or very dissatisfied with the product they purchased; those with moderate satisfaction levels are more reluctant to make their experiences public (e.g., Dellarocas and Narayan 2006; Hu et al. 2006; Koh et al. 2010). As a result, rating distributions are often u-shaped with mostly 5-star ratings, some 1-star ratings, and only a small number of ratings in between. Hence, average ratings based on such distributions may not necessarily represent an accurate indicator for a product's quality and could lead to false conclusions. For instance, when rating distributions do not concentrate on the mean, the average rating may reflect a balance point of very different opinions—rather than a summarizing measure—pointing out that there is an equal number of people who evaluated the product better and worse than the mean value (Hu et al. 2006).

Third, evaluation issues may arise because precisely determining a product's quality typically requires sophisticated and often expensive measurements since many quality dimensions cannot be easily assessed (e.g., safety and reliability of a child car seat; de Langhe et al. 2016c). However, customers who write a review often do not have the knowledge, equipment, and time necessary to evaluate a product's performance in this way. In addition, consumers' quality judgments are not only based on their own perception of product performance but also on the evaluations of other customers. In this context, extant research on the social influence bias in consumers' product ratings—i.e., the tendency to conform to a majority opinion rather than to reveal an own uninfluenced evaluation (e.g., Askalidis et al. 2017)—has reported that consumers' rating behavior is affected by already existing ratings (e.g., Godes and Silva 2012; Moe and Schweidel 2012; Moe and Trusov 2011; Muchnik, Aral, and Taylor 2013; Sridhar and Srinivasan 2012). Moreover, it is well-known that consumers' quality assessments are heavily biased by variables other than sheer performance criteria, such as brand image (e.g., Grewal et al. 1998; Jacoby et al. 1971), price (e.g., Dodds, Monroe, and Grewal 1991; Monroe 1973; Rao and Monroe 1989; Zeithaml 1988), and physical appearance (e.g., Dawar and Parker 1994), as well as motivational aspects (e.g., Sundaram, Mitra, and Webster 1998, see also Hennig-Thurau et al. 2004 as well as Mathwick and Mosteller 2017 for an overview of diverse consumer motives to engage in eWOM).

2.4 Synthesis and Conceptual Positioning of this Thesis

Online consumer reviews and their implications for research and businesses have attracted considerable interest from marketing scholars and practitioners. The above presented literature review classifies the abundance of existing literature in this field into two major research streams—namely, studies focusing on the characteristics of individual reviews and reviewers, and, on a more aggregate level, research concentrating on rating distribution characteristics—and compiles the central empirical findings within these two areas.

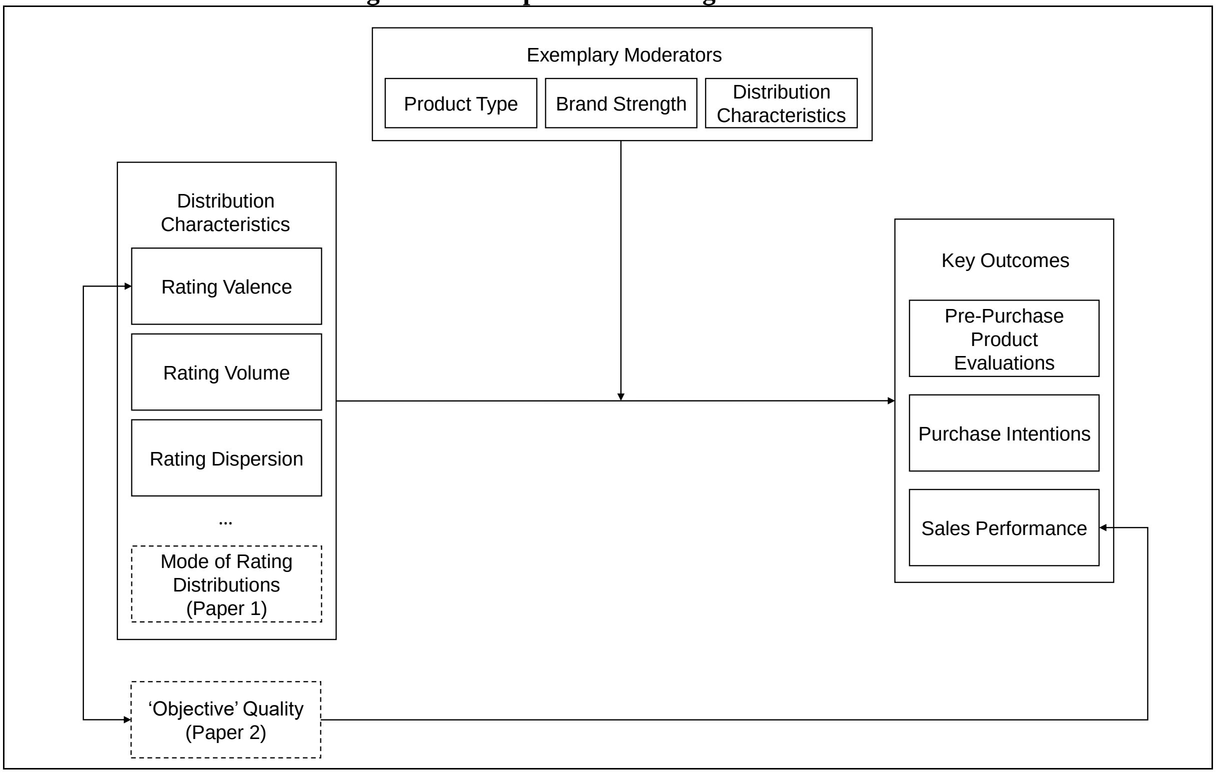
In summary, extant literature exerted a diverse set of methodological approaches, ranging from qualitative methods (e.g., content analyses, sentiment analyses, text mining, or verbal protocols; e.g., Cheng and Ho 2015; Gottschalk and Mafael 2017; Ludwig et al. 2013; Schlosser 2011; Villarroel Ordenes et al. 2017) over quantitative investigations (e.g., experimental studies or surveys; see for example Casaló et al. 2015; He and Bond 2015; Jiménez and Mendoza 2013; Khare et al. 2011; Kostyra et al. 2016; Kronrod and Danziger 2013; Langan et al. 2017; Moore 2015; Naylor, Lamberton, and Norton 2011; Purnawirawan, de Pelsmacker, and Dens 2012; Reimer and Benkenstein 2016; Sen and Lerman 2007; Shoham, Moldovan, and Steinhart 2017) to meta-analyses (e.g., Babić Rosario et al. 2016; Floyd et al. 2014; You et al. 2015). In addition, secondary real-world data used in the studies stems from marketers operating in a variety of different product and service domains; e.g., Amazon (e.g., Chevalier and Mayzlin 2006; de Langhe et al. 2016a; Ho-Dac et al. 2013; Ludwig et al. 2013; Moore 2015; Singh et al. 2017; Sun 2012), TripAdvisor (e.g., Sridhar and Srinivasan 2012; Villarroel Ordenes et al. 2017; Wilson et al. 2017), or Yahoo Movies (e.g., Chintagunta et al. 2010; Liu 2006; Moon et al.

2010; Schlosser 2011; Wang, Liu, and Fang 2015; see You et al. 2015 for an extensive overview).

Overall, the rich body of prior research has built a solid state of knowledge regarding the impacts of online reviews on consumer decision making and purchase behavior. However, several research and managerial questions still remain unanswered (see also King, Racherla and Bush 2014 for a research synthesis). On the one hand, the preceding literature review has summarized the manifold insights that have been gathered in terms of consumers' response to a variety of different characteristics of online rating distributions. In this regard, rating valence, volume, and dispersion, as well as their interactions and moderating effects evoked by several product-related or category-specific factors (e.g., product type, taste similarities and brand strength) have been found to be highly influential in affecting consumers' interpretations of and conclusions drawn from online reviews. However, despite the substantial body of research in this area our knowledge of the effects of distribution characteristics beyond average ratings, the number of ratings, and rating variance is still scarce. On the other hand, although several aspects that threaten the accuracy of consumer-generated product evaluations have been recognized and, thus, it should be considered doubtful whether online ratings actually constitute an adequate measure of a products' objective performance, only relatively few studies were concerned with the validity of consumer reviews as a quality indicator.

The two research papers presented subsequently address these gaps and, thereby, add to prior literature in the following manner: First, the studies documented in Paper 1 demonstrate that consumers use the mode of ratings distributions—a distribution feature that has been disregarded so far—when making predictions about product quality. Second, by examining the relationship between and the impacts of average product ratings and more objective measures of product quality, the second manuscript contributes to the literature on the convergence between rated and objective quality. Figure 4 illustrates the positioning of the aspects under investigation in this doctoral thesis within the field of research on the consequences of online consumer reviews.

Figure 4. Conceptual Positioning of this Thesis



Note: The main focus of this thesis is on the measures in interrupted boxes.

C Empirical Research Paper 1: The Mode Heuristic in Consumers' Interpretations of Online Rating Distributions

Abstract

This research demonstrates a tendency to use the mode as a heuristic basis when making product inferences from online rating distributions in such a way that product evaluations inferred from rating distributions with an equal average, standard deviation, and number of ratings systematically vary by the location of the mode; a phenomenon referred to as the mode heuristic. The results of a series of six studies, using a mix of experimental and real-world data, (1) provide strong empirical evidence for the existence of the mode heuristic in a variety of different contexts, (2) shed light on this phenomenon at the process level, and (3) demonstrate how product inferences based on the mode heuristic depend on the visual salience of the mode.

Additional note:
An extended version of this manuscript, co-authored by Sören Köcher (Köcher, Sarah, and Sören Köcher, "The Mode Heuristic in Consumers' Interpretations of Online Rating Distributions"), will be submitted to an A+ ranked journal (VHB-Jourqual3). Parts of this research have been presented and discussed at three consecutive AMA SERVSIG "Let's Talk About Service" Workshops in Namur (2015), New York (2016), and Antwerp (2017).

1 Introduction

The way people buy things has fundamentally changed. Enabled by modern technologies, consumers cannot only shop anything, anytime, anywhere but also share their opinions about products and services on a variety of websites such as Amazon, TripAdvisor, and Google with anyone. As a consequence, when making purchase decisions people increasingly rely on online ratings provided by previous customers as a credible information source to infer the quality of the available purchase options (e.g., Hu, Liu, and Zhang 2008; Li and Hitt 2008; Simonson and Rosen 2014); despite the fact that online reviews are deemed to be a rather imprecise indicator for a product's 'objective' quality (de Langhe, Fernbach, and Lichtenstein 2016a).

A very common practice to inform online shoppers about product evaluations from previous customers is to illustrate the distribution of rating scores through graphical formats; typically via bar charts, wherein each bar represents the number of votes a specific rating score has received. Although a broad body of literature has been devoted to acquiring insights into consumers' response to different characteristics of rating distributions (for an overview, see Babić Rosario et al. 2016) our knowledge of the effects of distribution characteristics beyond average ratings (e.g., Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; Godes and Mayzlin 2004), dispersion of rating scores (e.g., He and Bond 2015; Sun 2012; Zhang 2006), and rating volume (e.g., Liu 2006; Moe and Trusov 2011; Zhu and Zhang 2010) is still limited. Extending previous research on the impact of online ratings on pre-purchase product evaluations, the present work investigates how customers' interpretations of rating distributions are affected by the location of the mode; i.e., the rating score that has received the largest number of votes and, therefore, the most salient element in graphical visualizations of rating distributions. We argue that consumers tend to use the mode as a heuristic basis when making product inferences in such a way that product evaluations inferred from rating distributions with an equal average, standard deviation, and number of ratings systematically vary by the location of the mode; a phenomenon we refer to as the mode heuristic.

The contribution of this research is of equal relevance from both perspectives, theoretical as well as managerial. First, this article complements extant knowledge about the consequences of different rating distribution characteristics by placing a previously disregarded feature under scrutiny. Second, by establishing the mode of rating distributions as an important parameter in consumers’ product inferences, we provide marketers a new key figure which—aside from rating volume, average ratings, and rating dispersion—should be incorporated when monitoring, analyzing, and evaluating product review data. Third, by demonstrating that the mode of a rating distribution serves as a heuristic cue when inferring a product’s quality we also contribute to prior research that has reported systematic biases in the manner in which people process graphical illustrations of information (e.g., Cleveland and McGill 1984; Hutchinson, Alba, and Eisenstein 2010; Jarvenpaa 1990; Lewandowsky and Spence 1989; Raghubir and Das 2010) as well as to extant literature on the use of heuristics in judgement and decision making in general (see Gilovich, Griffin, and Kahneman 2002 for an overview).

The rest of the paper is organized as follows: We begin by reviewing previous research on consumers’ response to different characteristics of rating distributions. We then outline insights into how people process graphical visualizations of information like the bar charts that are typically used by online platforms to aggregate and summarize customer ratings (e.g., Amazon, TripAdvisor, Google Reviews) and apply them to the present research to derive the mode heuristic hypothesis. Thereafter, we report the results of a series of studies that demonstrate the existence of the mode heuristic using different survey designs and contexts. We conclude with a discussion of theoretical contributions, managerial implications, and future research directions.

2 Conceptual Background

2.1 Consumers' Response to Different Characteristics of Rating Distributions

Just as any other distribution, distributions of customer ratings can be summarized by a variety of descriptive statistics, such as frequencies of rating scores, measures of location (e.g., mean, mode, and median), measures of variability (e.g., standard deviation and variance), as well as measures of shape (e.g., skewness and kurtosis). Given the great popularity of online reviews, it is hardly surprising that a broad body of literature has been devoted to acquire insights into how consumers respond to different characteristics of rating distributions.

Most of this research concentrates on the effect of review valence revealing that higher average ratings are associated with favorable outcomes reflected in, for instance, higher purchase intentions, better sales ranks, revenues, and future ratings (e.g., Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Clemons, Gao, and Hitt 2006; Dellarocas et al. 2007; Godes and Mayzlin 2004; Liu 2006; Zhu and Zhang 2010). Interestingly, this relationship has been found to be dependent on brand strength, such that the positive effect associated with higher average ratings is more pronounced for weak brands rather than for strong brands (Ho-Dac, Carson, and Moore 2013; Luca 2011). In addition, a number of studies have examined the influence of review volume (i.e., the number of ratings) on product sales and related performance figures; albeit with mixed results. Several studies have revealed a positive effect of the number of ratings (e.g., Chevalier and Mayzlin 2006; Dellarocas et al. 2007; Liu 2006; Moe and Trusov 2011; Sun 2012; Zhu and Zhang 2010), while others could not support this relationship (Chintagunta et al. 2010; Clemons et al. 2006). In addition to review valence and volume, some studies have also focused on the impacts of rating dispersion in terms of the variance or standard deviation of rating distributions—reflecting the degree of consensus among reviewers' judgments—on consumers' product evaluations. However, findings on these distribution characteristics are notably ambiguous; they range from positive (e.g., Clemons et al. 2006; Lu, Ye and Law 2014; Moe and Trusov 2011) over non-significant (e.g., Chen, Liu, and Zhang 2011; Chintagunta et al. 2010; Zhang 2006) to negative effects (e.g., Bao and Chang 2014; Hu et al. 2010; Moon, Bergey, and Iacobucci 2010; Zhu and Zhang 2010; see also He and Bond 2015 for an overview). The effects of dispersion have been shown to vary by consumers' prior expectations (Park and Park 2013) and product type (He and Bond 2015; Langan, Besharat, and Varki 2017; Park and Park 2013). For instance, He and Bond (2015) found that consumers are more tolerant to dispersion in taste-dissimilar product domains (e.g., paintings or music albums) than in taste-similar product categories (e.g., desk lamps or flash drives). Finally, a few studies have documented interaction effects between the described distribution characteristics (Chintagunta et al., 2010; Khare, Labrecque, and Asare 2011; Kostyra et al. 2016; Sun, 2012). For example, Khare et al. (2011) reported that the positive effect of review valence is enhanced by rating volume, while Sun (2012) found that an increasing dispersion has a positive effect on sales if and only if the average rating is low (see also Khare et al. 2011; Kostyra et al. 2016).

In summary, although a great deal of research has studied consumers' response to different characteristics of rating distributions, insights beyond the effects of rating volume, valence, and dispersion as well as their interactions are still scarce. In the next section, we draw from extant research on how people process graphical formats like the bar charts used by marketers to display the distribution of product ratings in order to demonstrate that consumers' response to such illustrations can be crucially affected by their visual appearance.

2.2 People's Interpretations of Graphical Formats

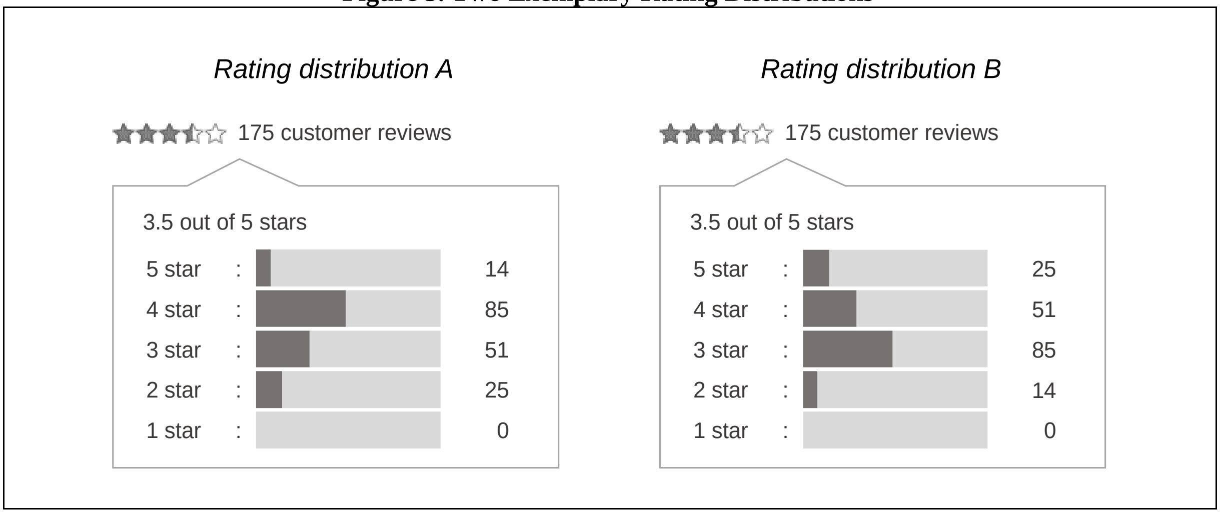
Similar to aggregated illustrations of individual rating scores by means of bar charts used by a variety of online retailers and review platforms, graphical visualizations of information are ubiquitous in our daily lives. For instance, graphics, such as bar and pie charts, as well as line graphs, are commonly used when reporting election results, visualizing weather forecasts, communicating health risks, or illustrating the development of stock prices. Unsurprisingly, research on people's interpretations of such visualizations has a long tradition; generally highlighting the usefulness of graphics as opposed to simple alpha-numeric representations (see, e.g., Lipkus 2007; Spiegelhalter, Pearson, and Short 2011; or Visschers et al. 2009 for a review). However, although graphical formats allow quick insights into the visualized data, the manner in which people process and interpret graphical information can be systematically biased (e.g., Cleveland and McGill 1984; Hutchinson et al. 2010; Lewandowsky and Spence 1989; Lurie and Mason 2007; Pinker 1990; Raghubir and Das 2010; Simkin and Hastie 1987). Extant literature suggests that when being confronted with bar charts, people tend to make comparisons between the magnitude of the bars and focus their attention to differences in physical length (e.g., Jarvenpaa 1990; Simkin and Hastie 1987; Spence 1990; Stone, Yates, and Parker 1997; Stone et al. 2003; see also Sun, Li, and Bonini 2010). Thus, when drawing conclusions from illustrations, the visual salience of each information provided may serve as a cue to its relative importance (Jarvenpaa 1990; Sanfrey and Hastie 1998). For instance, in one of their studies on potential differences in risk avoidance when communicating health risks via graphs or alpha-numeric displays, Stone et al. (1997) found that participants were willing to pay a significantly higher price for an improved toothpaste—with a reported likelihood of gum disease of 15 out of 5,000 people—relative to a standard toothpaste—30 out of 5,000 people affected by gum disease—when the chances of developing the disease for both alternatives were displayed as a bar chart; supporting their prediction that under graphic conditions, the extend of people's attention to information is determined by its visual salience. In a similar vein, a study conducted by Weber and Kirsner (1997) revealed that decisions between gambles can be biased toward the most salient elements of a bar chart representing possible payoffs. Furthermore, Ibrekk and Morgan (1987) demonstrated that people, when asked to estimate the mean value of a given distribution displayed as a bar chart, tend to anchor their estimates on the most salient bar within the graphic; i.e., the distribution's mode. Hence, we conclude that when distributions of customer ratings are communicated via bar charts—as is typically the case in business practices—people's product inferences from such graphs might be affected by the mode of a rating distribution in a similar manner. We further elaborate on this thought in the next section.

2.3 The Mode Heuristic

It is well known that individuals are typically not able to process and analyze all relevant information when forming judgments or arriving at decisions. Instead, they tend to base their judgments on simple cues or rules of thumb that facilitate the evaluation process (see Gilovich et al. 2002 for an extensive review). The use of such heuristics often leads to an inappropriate weighting of available informational cues. Prominent heuristics that exemplify such inadequate weighting include the anchoring heuristic (Tversky and Kahneman 1974), the availability heuristic (Tversky and Kahneman 1973; see also Reyes, Thompson, and Bower 1980) and the peak-end rule (Fredrickson and Kahneman 1993; Kahneman et al. 1993). The anchoring heuristic describes the tendency to heavily rely on the first piece of information acquired; even if this information is irrelevant for the judgmental task. The availability heuristic, in turn, refers to the phenomenon that people's judgments are strongly biased toward information that is easy to retrieve. Finally, according to the peak-end rule, overall evaluations of an affective experience are largely based on only certain salient moments—namely, its most intense (i.e., its 'peak') and final moment (i.e., its 'end')—rather than on an average of each single moment of the experience. In a similar vein, previous research on visual information processing has discovered the tendency to simplify judgmental tasks by drawing most attention to perceptually salient visual cues (e.g., Raghubir and Das 2010; see also Raghubir and Krishna 1999).

On the basis of this, we assume that the visual salience of the elements within graphical visualizations of rating distributions may serve as a heuristic basis when processing the provided information in a similar manner, such that people's inferences regarding a product's quality might be biased toward a distribution's mode; i.e., the most salient element within a graph. In order to illustrate the predicted mode heuristic we constructed two fictive rating distributions (see Figure 5). The two distributions share the same average rating (i.e., 3.5 out of 5 stars), standard deviation (i.e., SD = .84) and number of ratings (i.e., N = 175). However, they differ in terms of the location of the mode: In distribution A the mode is located above the average rating, while in distribution B the mode is situated below the distribution's mean value[5].

Figure 5. Two Exemplary Rating Distributions



According to our postulate, people might be prone to use the mode as a heuristic basis when drawing product inferences such that product evaluations inferred from the two illustrated distributions systematically diverge. More precisely, since the mode is the most salient bar it may attract people's attention more easily than the other bars and, thus, might be most influential when forming an overall impression of the reviewed product (e.g., Ibrekk and Morgan 1987; Weber and Kirsner 1997). Since the mode of distribution A—located above the average rating—directs consumers' attention to more favorable product evaluations (i.e., 4 out of 5 stars) than the mode of distribution B—located below the average rating (i.e., 3 out of 5

---

5 It should be mentioned that aside from having a different mode, the two distributions shown in Figure 1 also differ in terms of their direction of skew. Distribution A—wherein the distribution of values spreads from the mean further toward smaller values than toward larger values of the distribution—is left-skewed, while distribution B—wherein the distribution of values extends from the average value further toward larger values than toward smaller values—is skewed to the right. Although the skewness of a distribution and the location of its mode are typically strongly related, we empirically rule out that the skewness per se is instrumental in affecting consumers' interpretations of rating distributions (see Study 5).

stars)—product inferences derived from distribution A should be more favorable than those derived from distribution B. We refer to the tendency to interpret rating distributions predominantly on the basis of the location of their mode as the mode heuristic.

H1: *The mode of a rating distribution serves as a heuristic cue in consumers' product evaluations such that products will be judged more (less) favorably if the mode is located above (below) the average rating.*

Our hypothesis is also consistent with previous studies demonstrating that the visual properties of a stimulus that affect its visual salience (e.g., size, color, or shape) likewise guide people's attention to it (e.g., Janiszewski 1998; Mannan, Kennard, and Husain 2009; Milosavljevic et al. 2012; Parkhurst, Law, and Niebur 2002). The salience of a stimulus, in turn, has been shown to be influential in information processing, judgments, and decision making. For instance, salient product attributes have been demonstrated to be easier to remember (Ratneshwar et al. 1997) and to affect product evaluations and choice (e.g., MacKenzie 1986; Mandel and Johnson 2002; Shavitt and Fazio 1991).

3 Empirical Approach

In a series of six studies, using a mix of experimental and real-world data, we provide empirical evidence of the existence of the proposed mode heuristic. In Study 1, we demonstrate that consumers' inferences from rating distributions about the quality of a reviewed product as well as purchase intentions are affected by the location of the mode in the predicted manner. The subsequently reported Study 2 examines the mechanism behind this effect by investigating how the allocation of visual attention across individual rating scores (i.e., the bar of 5 star ratings, 4 star ratings, 3 star rating, and so forth) is determined by the mode of a rating distribution using an eye-tracking methodology. In Study 3, we replicate the findings from Study 1 in the context of u-shaped distributions. Study 4 examines the way in which the effect of the mode on product evaluations changes as a function of its visual salience. Then, in Study 5 we rule out that other distribution characteristics that are typically strongly related to a distribution's mode (i.e., skewness and median) can account for the observed mode heuristic. Finally, in Study 6 we provide evidence of external validity for the existence of mode heuristic using real-world customer review data from Amazon. Table 6 provides an overview of our empirical approach.

Table 6. Overview of Studies

	Study 1a and 1b	Study 2	Study 3a and 3b	Study 4a and 4b	Study 5a and 5b	Study 6
Method	Experiments	Eye-tracking experiment	Experiments	Experiments	Experiments	Amazon data
Study context and sample size	Printers (N = 65) and fast food restaurants (N = 78)	Toasters (N = 54)	Printers (N = 67) and fast food restaurants (N = 92)	Fitness trackers (N = 140) and movies (N = 129)	Electric water kettles (N = 138) and lecture evaluations (N = 148)	Top 100 products within 20 consumer electronics product categories (N = 1,536 usable observations)
Dependent variables	Perceived quality, purchase intentions	Allocation of visual attention to individual rating scores, perceived quality	Perceived quality, purchase intentions	Perceived quality	Perceived quality	Amazon bestseller ranks (as an indicator of sales performance)
Mediators	Perceived quality	Allocation of visual attention to individual rating scores	Perceived quality			
Manipulated (analyzed) distribution characteristics	Location of the mode (above vs. below the average rating)	Location of the mode (above vs. below the average rating)	Location of the mode (above vs. below the average rating)	Location (above vs. below the average rating) and visual salience of the mode (low vs. high)	Location of the mode (above vs. below the average rating), skewness (low vs. high), and median (constant)	Location and visual salience of the mode, skewness, median (controls: average rating, number of ratings, product price)
Purpose	Hypothesis test: Location of the mode (H1)	Hypotheses tests: H1 and mediating effect of the allocation of visual attention to individual rating scores (H2)	Robustness check of H1 for u-shaped rating distributions	Hypothesis test: Interplay between location and visual salience of the mode (H3)	Ruling out alternative explanations (i.e., skewness and median) for H1	External validity test

4 Study 1: First Evidence of the Mode Heuristic

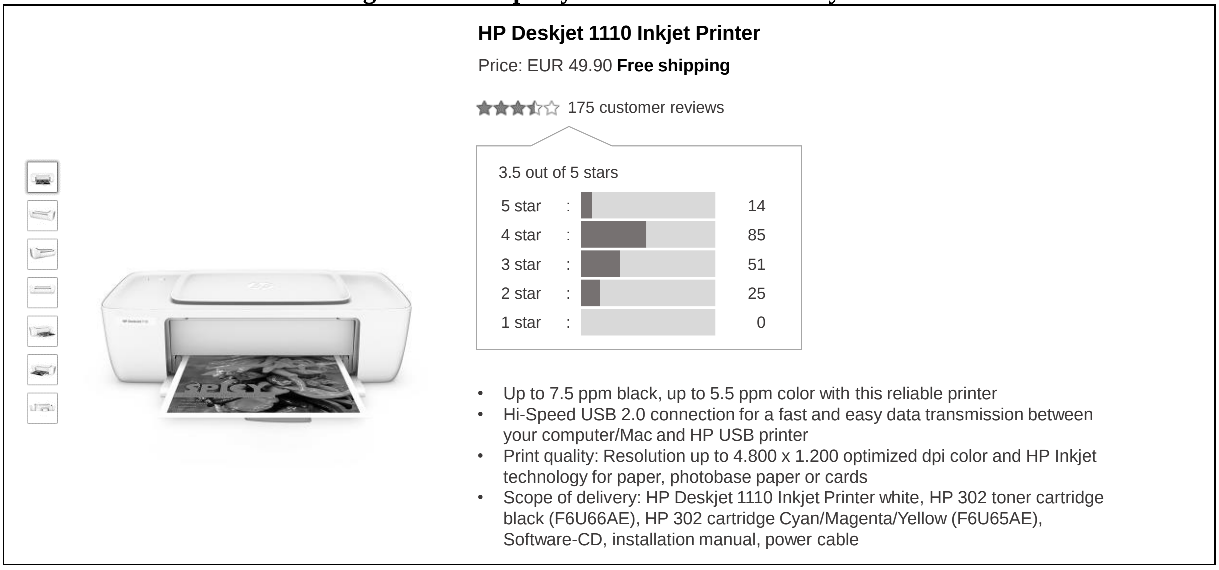
The purpose of Study 1 was to demonstrate the existence of the mode heuristic in consumers' interpretations of rating distributions by examining the effect of the location of the mode on consumers' product quality inferences (H1) and, by extension, on purchase intentions. In this study, we employed a single factor between-subjects design with two conditions where participants faced one of two rating distributions (i.e., location of the mode above vs. below the average rating). Hence, if the proposed mode heuristic did not exist, participants' reported quality perceptions and purchase intentions should be equivalent across the two conditions.

4.1 Study 1a

4.1.1 Participants, Design, and Procedure

Sixty-five students (M_{age} = 22.2 years, 38.5% female) participated in this online study for partial course credit. At the beginning of the experiment, participants were asked to imagine that they needed a new printer and, thus, searched the Amazon website to get an overview of current offerings. Subjects then saw a constructed illustration of a printer on Amazon including several product information (e.g., price and performance characteristics) as well as an overview of customer ratings displayed in a horizontal bar chart; using the two rating distributions shown in Figure 5 (see also Figure 6 for an exemplary stimulus used in this study). Thus, in both experimental conditions, the printer had received ratings from 175 reviewers with an average of 3.5 out of 5 stars. The standard deviation in ratings was .84. We manipulated the location of the mode by condition: In the below-average condition, the mode was 3, and in the above-average condition, the mode was 4. Hence, the rating distribution in the below-average condition was right-skewed (γ = .34), while the distribution used in the above-average condition was skewed to the left (γ = –.34). Participants were randomly assigned to one of the two conditions.

Figure 6. Exemplary Stimulus used in Study 1a



Notes: Translated to English. The current average rating of the illustrated product on Amazon.de is very similar to the average rating used in the experiment (printer: 3.6). The product is listed among the 100 best selling items within its associated product category (i.e., 'inkjet printers').

After processing the provided information, we asked participants to indicate their quality perceptions of the illustrated printer ("The printer appears to perform satisfactory", "The quality of the printer seems to be better than average", "The printer appears to be better than most other printers", "I think the quality of the printer is bad/good"; adapted from Taylor and Bearden 2002; α = .84) and their purchase intentions ("Based on the information provided, how likely would you buy this printer?"; very unlikely/very likely; e.g., Bertini, Ofek, and Ariely 2009; Raghubir and Greenleaf 2006). We also included two items to measure the perceived realism of the applied scenarios ("It was easy to imagine myself in this situation", "The situation described was realistic"; Dabholkar 1996; r = .66, p < .01). All variables were assessed on seven-point scales.

4.1.2 Results

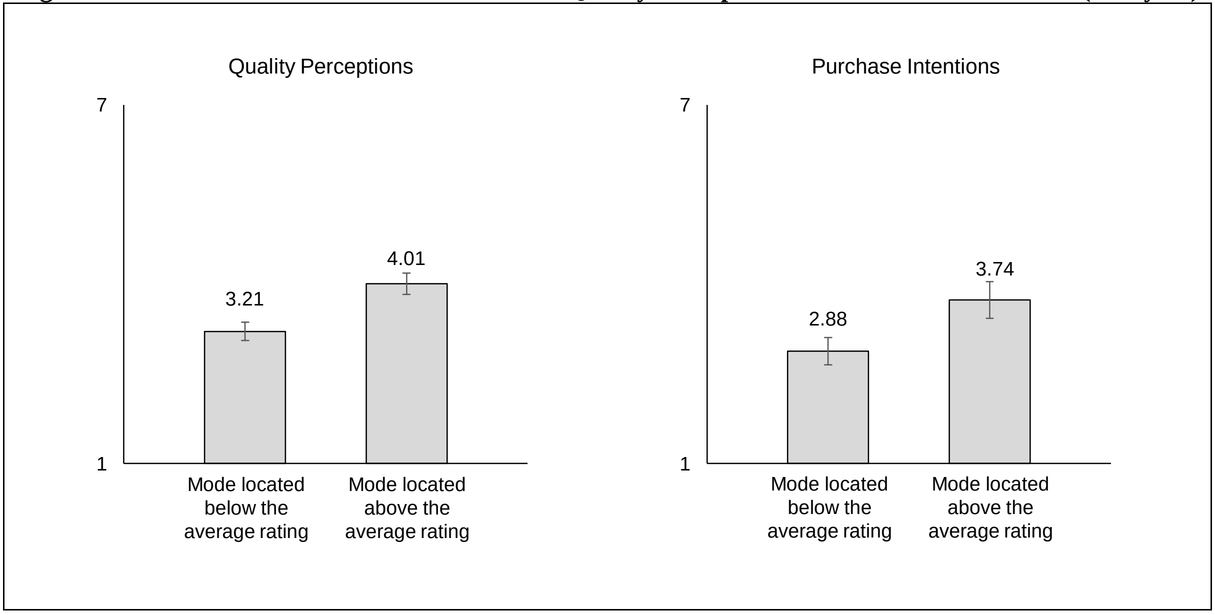
4.1.2.1 Realism Check

Answers to the realism check items (M = 6.06, SD = .99) indicated that respondents found the described scenarios to be highly realistic. Further analysis revealed that reported realism ratings were independent of the experimental conditions (t(63) = .10, p = .92).

4.1.2.2 Hypothesis Tests

Consistent with the proposed mode heuristic, participants who were confronted with the distribution wherein the mode was located above the average rating evaluated the quality of the presented printer significantly higher (M = 4.01, SD = .98) than those who were confronted with the distribution wherein the mode was located below the mean (M = 3.21, SD = .89, t(63) = 3.45, p < .01; see Figure 7).

Figure 7. The Effect of Location of the Mode on Quality Perceptions and Purchase Intentions (Study 1a)



Note: Error bars denote standard errors.

Findings on participants' reported purchase intentions showed the same pattern. More precisely, purchase intentions were significantly higher when the mode of the presented rating distribution was located above the average rating (M = 3.74, SD = 1.71) than when it was located below the mean (M = 2.88, SD = 1.32, t(63) = 2.28, p < .05).

4.1.2.3 Mediation Analysis

Lastly, we assessed whether quality perceptions mediated the detected effects of the location of the mode on purchase intentions using a process analysis (Hayes 2013; model 4). The estimated model included location of the mode (below vs. above the mean) as independent variable, purchase intentions as dependent variable, and quality perceptions as mediator of their relationship. We estimated the model with 95% bias-corrected confidence intervals (CIs) using 10,000 bootstrap samples. Our results showed a significant indirect effect of location of the mode on purchase intentions via quality perceptions (B = .88, SE = .25, CI_{95}: .46 to 1.43). Notably, the inclusion of quality perceptions in the model reduced the significant effect of location of the mode on purchase intentions uncovered in the above documented analysis to insignificance (B = −.02, SE = .30, t(62) = .08, p = .94). Thus, the detected variations in purchase intentions associated with different locations of the rating distributions' mode were completely explained by people's quality inferences derived from these different rating distributions.

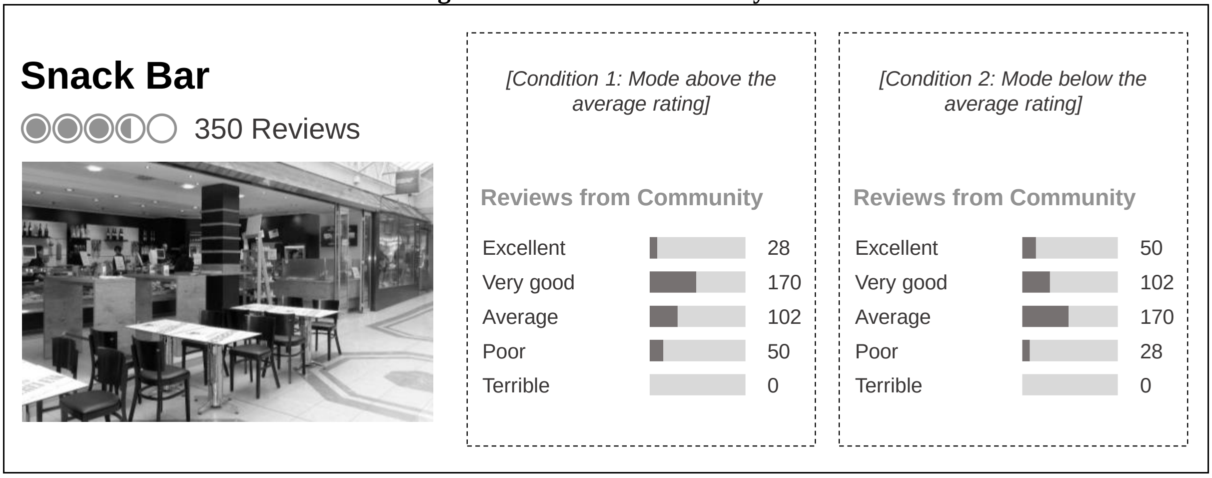
4.2 Study 1b

The purpose of Study 1b was to replicate the findings obtained from Study 1a in a service context and, thereby, to provide additional support for the existence of the mode heuristic.

4.2.1 Participants, Design, and Procedure

Seventy-eight students (M_{age} = 21.5 years, 46.2% female) participated in this study for partial course credit. Participants were asked to imagine that they went on a weekend sightseeing trip. After they had arrived at their destination they wanted to have a snack and, therefore, searched the Internet for a fast food restaurant nearby. Subjects then saw an illustration picturing a fast food restaurant on an online review website (see Figure 8). Basically, we used the same two rating distributions as in Study 1a but doubled the rating volume. Hence, in both experimental conditions, the restaurant had received ratings from 350 reviewers with an average of 3.5 out of 5 points. The standard deviation in ratings was .84. The location of the mode differed by condition: in the below-average condition, the mode was 3, and in the above-average condition, the mode was 4. Participants were randomly assigned to one of the two conditions.

Figure 8. Stimuli used in Study 1b



Note: Translated to English. The information on the conditions, provided in parenthesis, was not shown to participants.

After processing the provided information, participants were asked to indicate their perceptions of quality of the food at the illustrated fast food restaurant (“The food at this restaurant seems to have been good in the past”, “The quality of the restaurant’s food seems to be good”, “The food at this restaurant seems to be delicious”, “I think the quality of the food at this restaurant is bad/good”; adapted from Hess, Ganesan, and Klein 2003; $\alpha = .87$) and their purchase intentions (“Based on the information provided, how likely would you visit this restaurant?”; very unlikely/very likely). As in Study 1a, we also captured the perceived realism of the applied scenarios (“It was easy to imagine myself in this situation”, “The situation described was realistic”; $r = .61$, $p < .01$).

4.2.2 Results

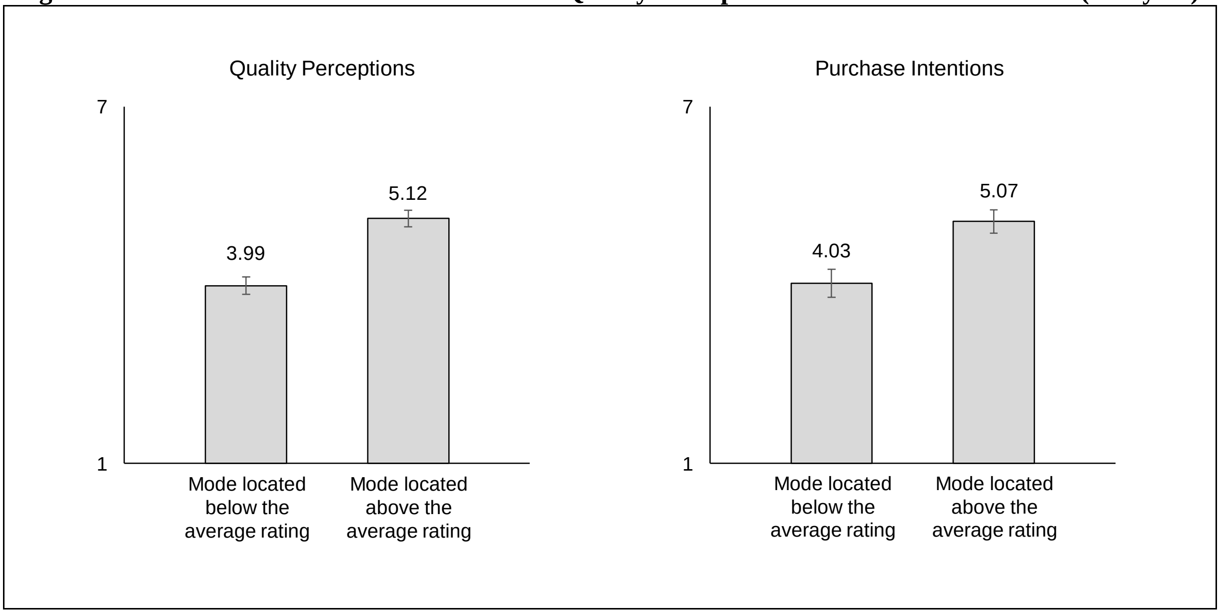
4.2.2.1 Realism Check

The calculated mean of the realism check items ($M = 5.90$, $SD = 1.05$) indicated that respondents found the described scenarios to be realistic. Further analysis revealed that reported realism ratings were independent of the experimental conditions ($t(76) = 1.62$, n.s.).

4.2.2.2 Hypothesis Tests

As in Study 1a, participants who were confronted with a distribution with a mode above the average rating perceived the quality of the presented restaurant significantly higher (M = 5.12, SD = .89) than those who were confronted with a distribution wherein the mode was located below the mean (M = 3.99, SD = .89, t(76) = 5.60, *p* < .01; see Figure 9). Analysis of participants' reported purchase intentions revealed similar results. More precisely, purchase intentions were significantly higher when the mode of the presented rating distribution was located above the average rating (M = 5.07, SD = 1.27) than when it was located below the mean (M = 4.03, SD = 1.44, t(76) = 3.40, *p* < .01).

Figure 9. The Effect of Location of the Mode on Quality Perceptions and Purchase Intentions (Study 1b)



Note: Error bars denote standard errors.

4.2.2.3 Mediation Analysis

Lastly, we assessed the mediating effect of quality perceptions within the relationship between the location of the mode and purchase intentions using bootstrapping analysis. Consistent with Study 1a, the mediation pathway from location of the mode to purchase intentions through quality perceptions was significant (indirect effect: B = .93, SE = .31, CI_{95}: .45 to 1.69), while the direct effect of location of the mode on purchase intentions turned out to be not statistically significant (B = .12, SE = .31, t(75) = .37, p = .71); indicating that the variance in purchase intentions induced by a varying location of the mode can be explained by diverging quality inferences drawn from the two presented rating distributions.

4.3 Discussion

Study 1 demonstrated consumers' use of the mode heuristic when inferring the quality of a product (Study 1a) or a service (Study 1b) from rating distributions. Specifically, as we predicted in H1, participants' reported quality perceptions were more favorable when the mode of a rating distribution was located above the average rating than when it was located below the mean. In addition, our mediation analysis confirmed that these effects can carry over to purchase intentions.

5 Study 2: The Mediating Role of Visual Attention

Consistent with our mode heuristic account, the results of Study 1 revealed that consumers' product evaluations inferred from rating distributions are affected by the location of the mode. The objective of Study 2 was to provide more direct evidence for our theorizing by examining the manner in which people process the graphical illustrations of the rating distributions used in Study 1. More precisely, we hypothesized that the mode—as the most salient element of a bar chart—attracts consumers' attention. Since the amount of attention directed to a specific piece of information has been shown to be positively related to its importance in judgment formation (e.g., MacKenzie 1986; Mandel and Johnson 2002; Shavitt and Fazio 1991), we assume that the focus on relatively positive (negative) product evaluations as induced by a location of the mode above (below) the average rating can be held responsible for the demonstrated effect on quality perceptions. In short, we expect that the amount of visual attention paid to the bars of each rating score (i.e., the bar of 5 star ratings, 4 star ratings, 3 star ratings, and so forth) varies by the location of the mode which, in turn, affects consumers' quality inferences. That is, we propose that the effect predicted by H1 and supported in Study 1 can be explained via changes in the allocation of visual attention to different rating scores. Formally stated,

H2: *The effect of the location of the mode on consumers' product evaluations is mediated by the allocation of visual attention to individual rating scores.*

To test the proposed mediation we conducted an eye-tracking experiment and examined participants' eye movements when processing graphical visualizations of rating distributions. In prior research, eye movements have often been used as a physiological measure to capture the allocation of visual attention in a variety of marketing-relevant contexts including, for instance, advertising effectiveness (e.g., Aribarg, Pieters, and Wedel 2010; Teixeira, Wedel, and Peiters 2012; Venkatraman et al. 2015; Zhang, Wedel, and Pieters 2009), assortment processing (e.g., Chandon et al. 2009; Deng et al. 2016; Townsend and Kahn 2014), and decision making (e.g., Atalay, Bodur, and Rasolofoarison 2012; Meißner, Musalem, and Huber 2016; Yang, Toubia, and De Jong 2015). The relationship between attention and eye movement processes has also been supported in neuroscientific studies (e.g., Corbetta et al. 1998; Kustov and Robinson 1996).

5.1 Participants, Design, and Procedure

Fifty-four students (M_{age} = 24.5, 55.6% female) from our university participated in this study for extra course credit. In this study, participants viewed the stimulus on a 24 inch computer screen. A Gazepoint GP3 eye-tracker—located below the screen—recorded the exact location of participants' eye fixations on the screen at any moment during the study. This eye-tracking device uses a 60 hertz machine-vision camera to track participants' eye gaze and allows head movements within a region of 25 centimeters × 11 centimeters × 15 centimeters. Since the device does not require headgear, participants were also able to wear reading glasses or contact lenses. To adapt the eye-tracker to each participant, we used a standard 9 point calibration and subsequent validation. After calibrating the eye-tracking device, we asked participants to imagine that they needed a new toaster and searched the Amazon website to get an overview of current offerings. Subjects then saw an illustration of a toaster on Amazon on the screen, including a product picture, product information (e.g., price and performance characteristics) as well as an overview of customer ratings through a bar chart. We used the same two rating distributions as in Study 1. Both rating distributions had an average rating of 3.5 out of 5 stars and a standard deviation of .84. We manipulated the location of the mode between-subjects: In the below-average condition, the mode was 3, and in the above-average condition, the mode was 4.

Participants were asked to carefully review the product on the screen as if they were indeed intending to buy a new toaster. We constrained the viewing time to 30 seconds for each participants. A pretest confirmed that this was enough time to unhurriedly read all information provided on the screen and to get a first impression of the illustrated product. After processing the provided information, we asked participants to estimate the quality of the presented toaster on a seven-point scale ranging from "bad" (1) to "good" (7).

5.2 Results

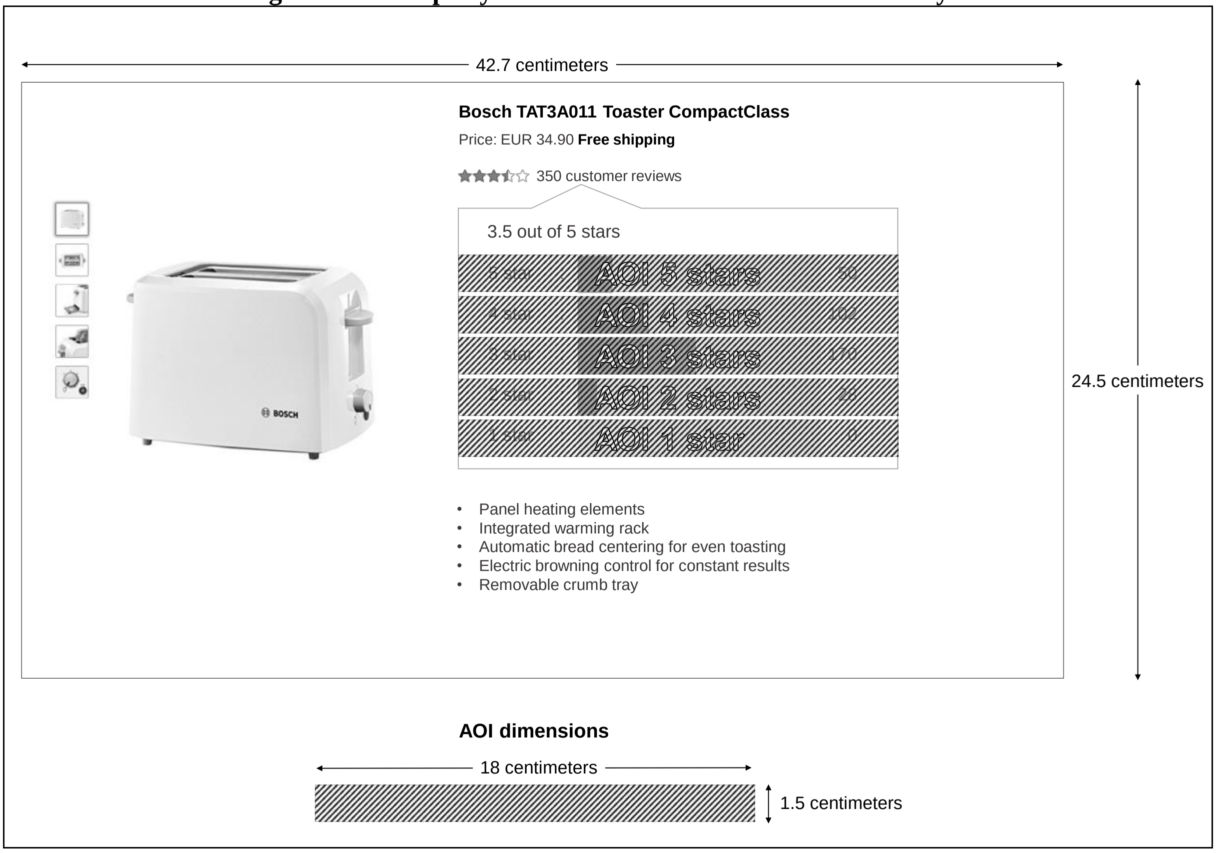
5.2.1 Quality Perceptions

As in Study 1, participants evaluated the quality of the presented toaster higher when the mode of the rating distribution they saw was located above the average rating (M = 4.81, SD = .96) than when it was situated below the mean (M = 4.22, SD = .97, $t(52) = 2.25$, $p < .05$).

5.2.2 Allocation of Visual Attention

In a next step, we examined how often and for how long participants looked at each element of the presented bar chart—i.e., the bars representing the number of 5 star ratings, 4 star ratings, 3 star ratings, and so forth, including their labeling—by assigning participants’ fixations to areas of interest (e.g., Pieters, Rosbergen, and Wedel 1999; see Figure 10 for an exemplary stimulus used in this study including the defined areas of interest).

Figure 10. Exemplary Stimulus and Areas of Interest in Study 2



Note: AOI = Area of interest; translated to English.

On average, participants fixated 6.23 seconds (SD = 2.73) on the bars within the chart; the average number of fixations on the bar chart was 32.11 (SD = 11.45). Neither fixation duration (t(52) = 1.29, p = .20) nor the number of fixations (t(52) = .59, p = .56) was significantly different between the two conditions.

We then analyzed the allocation of visual attention to each of the five bars within the graph. To account for differences in the overall time spent on the bar chart across participants, we calculated the relative number of fixations on each area of interest (i.e., each bar of the chart) as well as relative fixation durations per participant (see Table 7)[6]. Analysis of these relative measures of visual attention revealed that, indeed, most of the time spent on viewing the bar chart was devoted to the bar of the distribution's mode. More precisely, when the distribution's mode was 3, 37 percent of the total time spent on the bar chart was devoted to the mode's bar which was significantly higher than the percentage of time spent on any other bar (all t's > 2.02, all *p*'s < .05). Similarly, when the distribution's mode was 4, 42 percent of the total time spent on the bar chart was devoted to the mode's bar which was also significantly higher than the percentage of time spent on any other bar (all t's > 3.46, all *p*'s < .01). Accordingly, participants in the below-average condition (i.e., mode = 3) devoted a significantly higher proportion of the time spent on the rating distribution to the bar representing the number of 3 star ratings (M = .37, SD = .18) than those in the above-average condition (i.e., mode = 4; M = .20, SD = .10, t(52) = 4.33, *p* < .01). Vice versa, participants in the above-average condition devoted a significantly higher proportion of the time spent on the rating distribution to the bar of the 4 star ratings (M = .42, SD = .14) than those who were in the below-average condition (M = .27, SD = .13, t(52) = 4.01, *p* < .01). Interestingly, participants' attention to the other bars of the chart (i.e., the bars of 1 star, 2 star, and 5 star ratings) was not significantly different between the two experimental conditions (all t's < 1.20, all *p*'s > .23). As shown in Table 7, similar effects emerged when considering the relative number of fixations.

[6] We calculated the relative number of fixations for each of the five bars by dividing the number of fixations on a specific bar (e.g., the bar of 5 star ratings) by the sum of number of fixations on all of the five bars. Analogously, we calculated relative fixation durations for each bar by dividing the fixation time on a specific bar (e.g., the bar of 5 star ratings) by the sum of the time spent looking at all of the five bars.

Table 7. The Effect of the Location of the Mode on the Allocation of Visual Attention (Study 2)

	Number of Fixations						Fixation Durations (in seconds)					
	Distribution A (Mode = 4 stars)		Distribution B (Mode = 3 stars)		Differences between Distribution A and B		Distribution A (Mode = 4 stars)		Distribution B (Mode = 3 stars)		Differences between Distribution A and B	
Areas of Interest	Absolute	Relative	Absolute	Relative	Absolute (t-value)	Relative (t-value)	Absolute	Relative	Absolute	Relative	Absolute (t-value)	Relative (t-value)
5 Star Ratings	8.48 (3.71)	.27 (.08)	6.81 (3.05)	.24 (.13)	1.80*	.94	1.75 (1.17)	.27 (.14)	1.11 (.67)	.22 (.19)	2.46**	1.01
4 Star Ratings	12.81 (5.97)	.39 (.08)	9.44 (5.79)	.28 (.10)	2.11**	4.25***	2.88 (1.72)	.42 (.14)	1.71 (1.28)	.27 (.13)	2.82***	4.01***
3 Star Ratings	7.04 (4.01)	.21 (.08)	9.78 (4.55)	.31 (.11)	2.35**	3.84***	1.37 (.93)	.20 (.10)	2.25 (1.47)	.37 (.18)	2.62**	4.33***
2 Star Ratings	3.15 (2.25)	.09 (.06)	3.56 (2.19)	.12 (.08)	.68	1.41	.50 (.40)	.07 (.06)	.48 (.35)	.10 (.09)	.12	1.20
1 Star Ratings	1.56 (1.50)	.05 (.05)	1.59 (1.85)	.06 (.06)	.08	.31	.21 (.23)	.04 (.04)	.19 (.32)	.03 (.06)	.24	.21
Attention weighted Mean	3.73 (.29)		3.53 (.39)		2.16**		3.81 (.31)		3.55 (.46)		2.40**	

Notes: SD in parentheses; $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$.

In a next step, we calculated two aggregated measures of participants' allocation of visual attention by multiplying each rating score by its associated (1) relative number of fixations or (2) its relative fixation time, respectively[7]. Hence, higher (lower) values on these measures indicated that participants devoted more attention to the bars representing the frequencies of higher (lower) rating scores. Analysis of the attention weighted mean values based on the number of fixations across the two experimental conditions revealed that this measure was significantly higher for participants in the above-average condition (M = 3.73, SD = .29) than for those in the below-average condition (M = 3.53, SD = .39, t(52) = 2.16, *p* < .05). A similar effect emerged when considering the attention weighted mean values based on fixation durations ($M_{\text{below-average}}$ = 3.55, SD = .46 vs. $M_{\text{above-average}}$ = 3.81, SD = .31, t(52) = 2.40, *p* < .05); indicating that participants in the above-average condition allocated more attention to higher rating scores than those in the below-average condition.

5.2.3 Mediation Analysis

Finally, we tested whether the identified changes in participants' allocation of visual attention caused by variations of a rating distribution's mode mediated the effect of the location of the mode on quality perceptions (H2) using bootstrapping analysis. The estimated model included location of the mode (below vs. above the mean) as independent variable, quality perceptions as dependent variable, and the calculated attention weighted mean based on fixation durations

---

7 Formally, we calculated the attention weighted mean for each participant using the following formula:

$$\text{attention weighted mean} = \sum_{i=1}^{5} \text{rating score}_i \times \text{relative number of fixations}_i$$

where we weighted each rating score_i (ranging from 1 to 5) by its associated relative number of fixations. For instance, assuming a participant directed 50% of all eye fixations toward the bar representing the number of 4 star ratings and the other 50% toward the bar of 3 star ratings, then the attention weighted mean would be 3.5 (i.e., 4 × .5 + 3 × .5) for this participant. As a second measure of participants' allocation of visual attention across the elements of the bar chart, we calculated the attention weighted mean based on fixation durations by replacing the relative number of fixations in the above formula with relative fixation durations.

as mediator of their relationship. This analysis revealed a significant indirect effect of location of the mode on quality perceptions via the attention weighted mean (B = .19, SE = .10, CI_{95}: .04 to .46). Notably, the inclusion of the attention weighted mean in the model reduced the significant effect of location of the mode on quality perceptions uncovered in the above documented analysis to insignificance (B = .40, SE = .27, t(51) = 1.51, *p* = .14); indicating that the attention weighted mean was a meaningful predictor of the detected variations in quality perceptions associated with different locations of the rating distributions’ mode. Using the attention weighted mean based on the number of fixations instead of fixation durations as a mediator produced similar results. We found a significant indirect effect of the location of the mode on quality perceptions via the attention weighted mean (B = .16, SE = .09, CI_{95}: .02 to .41), while the direct effect turned out to be not statistically significant (B = .43, SE = .27, t(51) = 1.63, *p* = .11).

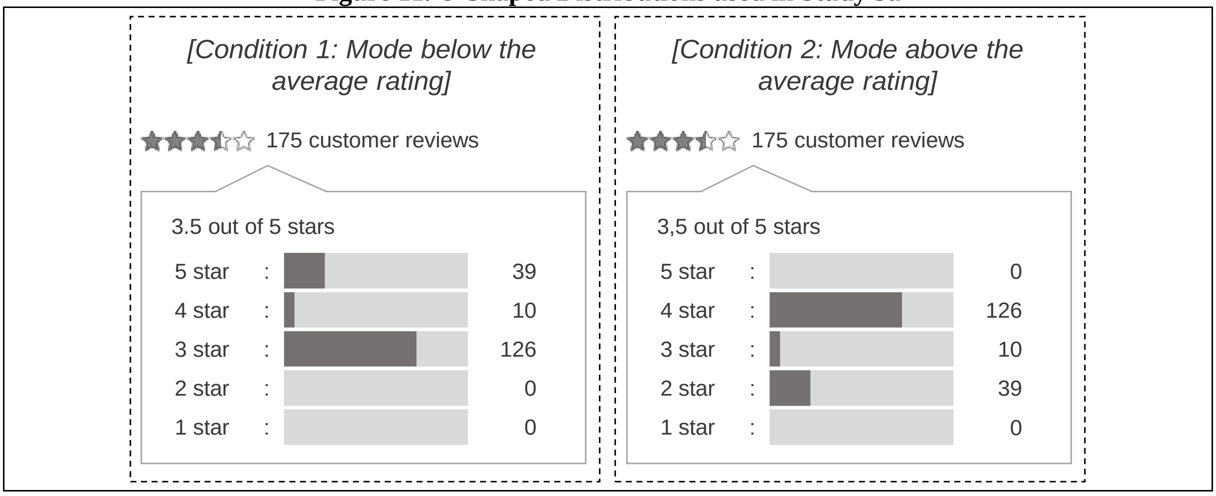
5.3 Discussion

In sum, Study 2 sheds light on the underlying mechanism of the mode heuristic by investigating the manner in which graphical displays of rating distributions are processed. As discussed earlier, we anticipated that the allocation of visual attention to the bars of different rating scores is determined by the location of the mode. Consistent with this expectation, we found that participants’ attention was directed toward more favorable (unfavorable) product ratings when the mode was located above (below) the average rating and that this shift in the allocation of visual attention prompted more favorable (unfavorable) product evaluations. This mediation pathway was robust when considering different measures of visual attention (i.e., number of fixations and fixation durations).

6 Study 3: The Mode Heuristic in U-Shaped Rating Distributions

The purpose of Study 3 was to test the robustness of the mode heuristic when considering u-shaped rating distributions. In order to do so, we replicated Study 1a and 1b. The only difference was that we replaced the unimodal distributions used in the first study with u-shaped rating distributions, while still keeping the average rating (M = 3.5), standard deviation (SD = .84), and rating volume (N = 175) at the same level. As in Study 1, the location of the mode differed by condition: In the below-average condition, the mode was 3, and in the above-average condition, the mode was 4 (see Figure 11).

Figure 11. U-Shaped Distributions used in Study 3a



Note: Translated to English. The information on the conditions, provided in parenthesis, was not shown to participants.

6.1 Study 3a

6.1.1 Participants, Design, and Procedure

Sixty-seven students (M_{age} = 21.1 years, 40.3% female) participated in this study for partial course credit. Using the same cover story as in Study 1a, subjects were asked to imagine that they needed a new printer. They were then exposed to a constructed illustration of a printer on Amazon including a product picture, product information (e.g., price and performance characteristics), as well as an overview of customer ratings through a bar chart. Participants were randomly assigned to one of the two conditions. After participants had gone through the scenario, we asked them to indicate their perceptions of quality of the illustrated printer (α =

.83), purchase intentions, as well as the perceived realism of the described situation (r = .77, *p* < .01) using the same measures as in Study 1a.

6.1.2 Results

6.1.2.1 Realism Check

The calculated mean of the realism check items (M = 5.50, SD = 1.28) indicated that respondents judged the described scenarios as realistic. Further analysis revealed that reported realism ratings were independent of the two conditions (t(65) = .76, *p* = .45).

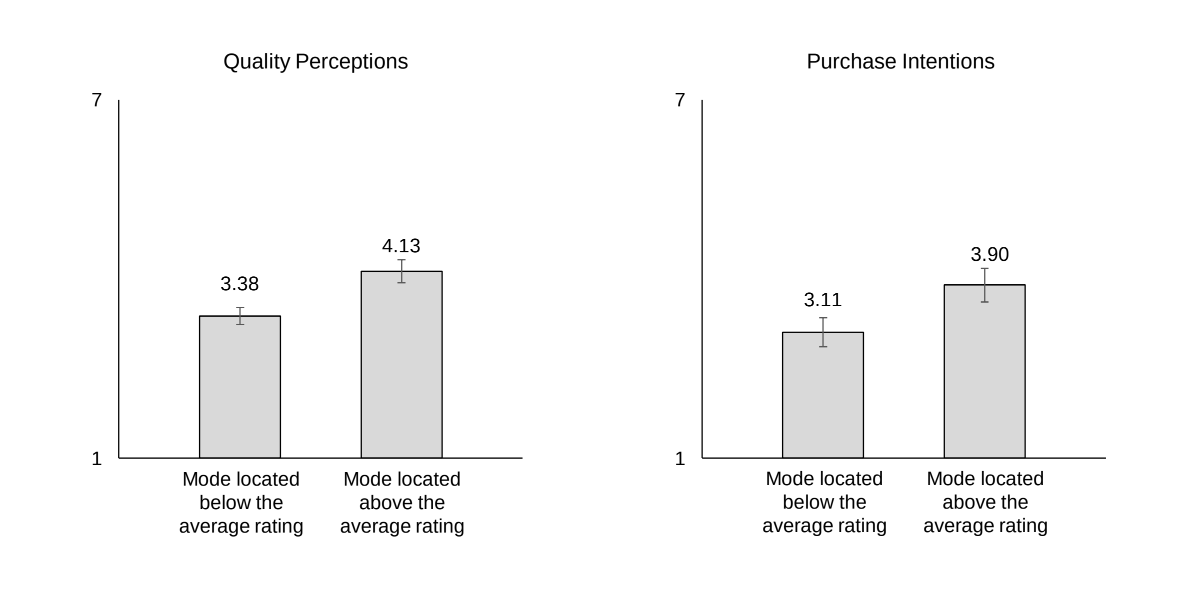
6.1.2.2 Hypothesis Tests

In accordance with the findings obtained from Study 1a, participants who were confronted with the rating distribution wherein the mode was located above the average rating evaluated the quality of the presented printer significantly higher (M = 4.13, SD = 1.06) than those who were confronted with the distribution wherein the mode was situated below the mean (M = 3.38, SD = .85, t(65) = 3.20, *p* < .01; see Figure 12). In addition, participants' reported purchase intentions were significantly higher in the above-average condition (M = 3.90, SD = 1.58) than in the below-average condition (M = 3.11, SD = 1.45, t(65) = 2.14, *p* < .05).

6.1.2.3 Mediation Analysis

Analogous to Study 1a, we also assessed whether quality perceptions mediated the detected effects of the location of the mode on purchase intentions using bootstrapping analysis. Our results showed a significant indirect effect of location of the mode on purchase intentions via quality perceptions (B = .93, SE = .30, CI_{95}: .36 to 1.54). As in Study 1a, the direct effect of location of the mode on purchase intentions turned out to be not statistically significant (B = −.13, SE = .25, t(64) = .54, *p* = .59).

Figure 12. The Effect of Location of the Mode on Quality Perceptions and Purchase Intentions (Study 3a)



Note: Error bars denote standard errors.

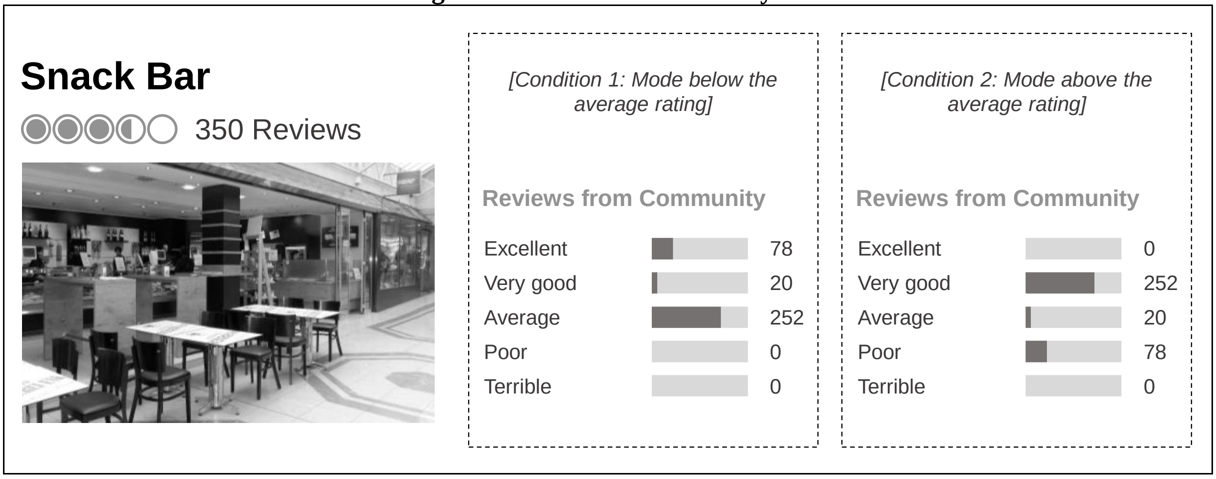
6.2 Study 3b

Similar to Study 1b, the purpose of Study 3b was to replicate the findings obtained from Study 3a in another context and, thereby, to provide further support for the robustness of the mode heuristic when considering u-shaped rating distributions.

6.2.1 Participants, Design, and Procedure

Ninety-two students (M_{age} = 21.4 years, 32.6% female) participated in this study for partial course credit. Using the same cover story as in Study 1b, participants saw an illustration picturing a fast food restaurant on an online review website. Basically, we used the same two rating distributions as in Study 3a but doubled the rating volume. Hence, in both experimental conditions, the restaurant had received ratings from 350 reviewers with an average of 3.5 out of 5 points. The standard deviation in ratings was .84. The location of the mode differed by condition: in the below-average condition, the mode was 3, and in the above-average condition, the mode was 4. Participants were randomly assigned to one of the two conditions (see Figure 13). After processing the provided information, participants answered to the same scales as in Study 1b to measure their perceptions of quality of the food (α = .87) at the illustrated fast food restaurant, their purchase intentions, and the perceived realism of the described scenario (r = .64, $p < .01$).

Figure 13. Stimuli used in Study 3b



Note: Translated to English. The information on the conditions, provided in parenthesis, was not shown to participants.

6.2.2 Results

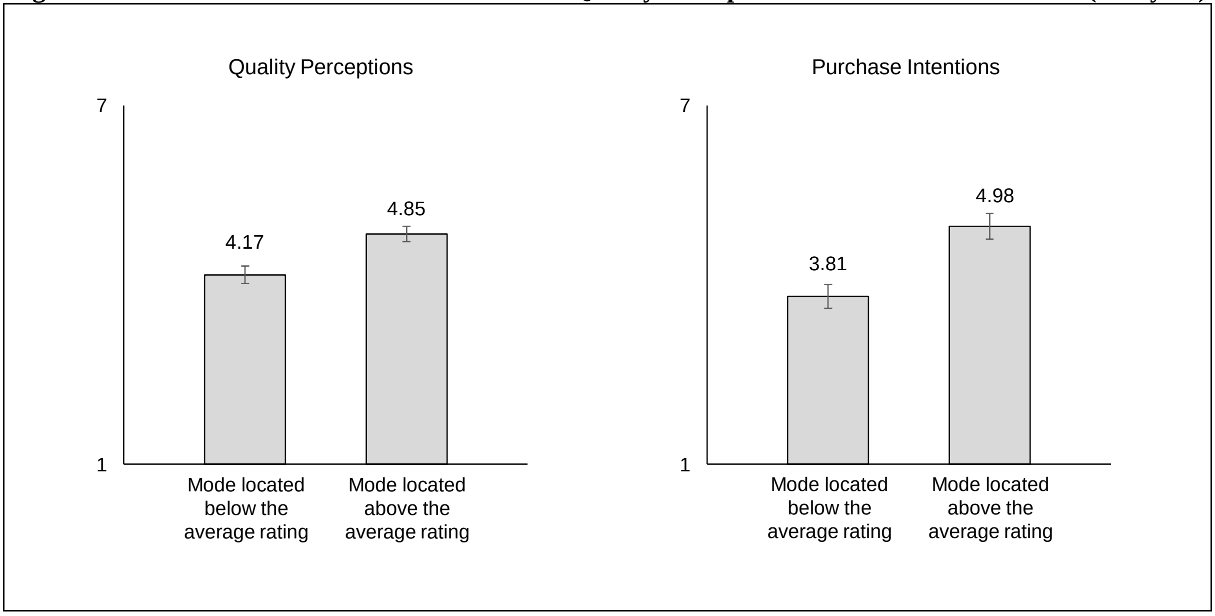
6.2.2.1 Realism Check

The calculated mean of the realism check items (M = 5.43, SD = 1.34) indicated that respondents found the described situation to be realistic. Further analysis revealed that reported realism ratings were independent of the experimental conditions (t(90) = .88, $p = .38$).

6.2.2.2 Hypothesis Tests

In accordance with the findings obtained from Study 3a, participants who were confronted with a distribution wherein the mode was located above the average rating perceived the quality of the presented restaurant significantly higher (M = 4.85, SD = .85) than those who faced a distribution with a mode below the mean (M = 4.17, SD = 1.00, t(90) = 3.50, $p < .01$; see Figure 14). In addition, participants' reported purchase intentions were significantly higher when the mode of the presented rating distribution was located above the average rating (M = 4.98, SD = 1.42) than when it was located below the mean (M = 3.81, SD = 1.38, t(90) = 3.99, $p < .01$).

Figure 14. The Effect of Location of the Mode on Quality Perceptions and Purchase Intentions (Study 3b)



Note: Error bars denote standard errors.

6.2.2.3 Mediation Analysis

Finally, we assessed the mediating effect of quality perceptions within the relationship between the location of the mode and purchase intentions. Consistent with Study 3a, we found a significant indirect effect of the location of the mode on purchase intentions via quality perceptions (B = .68, SE = .21, CI_{95}: .32 to 1.14). However, the effect of the location of the mode on purchase intentions was only partially mediated by quality perceptions; the direct effect remained significant (B = .48, SE = .23, t(89) = 2.07, $p < .05$).

6.3 Discussion

In Study 3a we found support for our primary hypothesis (H1) when considering u-shaped rating distributions; successfully replicating the findings obtained in Study 1a. In addition, Study 3b confirms these results in a different context (i.e., restaurant evaluations). In sum, these findings (together with those of Study 1a and 1b) provide evidence for the robustness of the effect of the location of the mode on people's quality inferences across differently shaped rating distributions; providing additional support for the robustness of our primary hypothesis (H1).

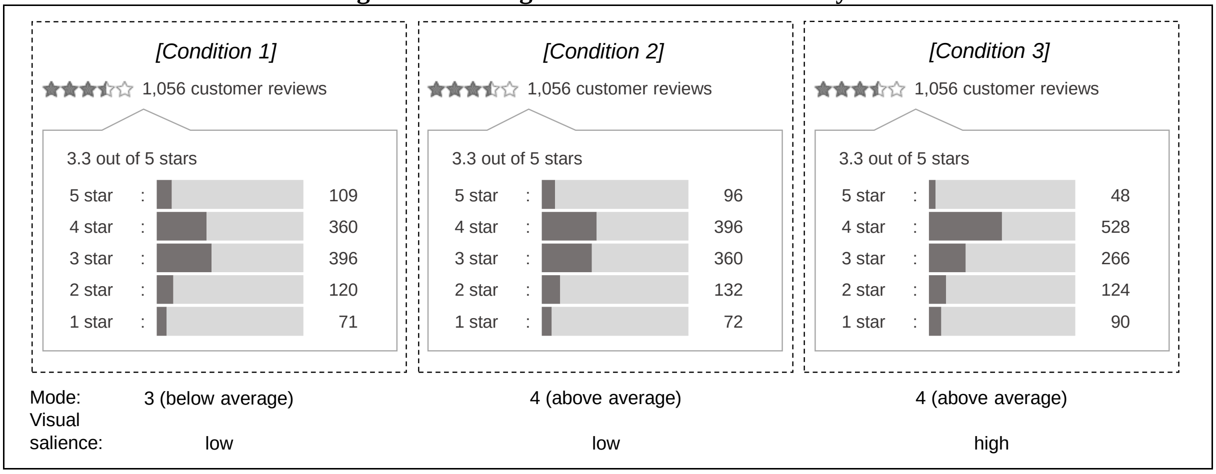
7 Study 4: Manipulating the Visual Salience of the Mode

Having shown that consumers tend to use the mode as a heuristic basis when making product inferences from rating distributions displayed as a bar chart, the objective of Study 4 was to investigate whether the impact of the mode on product evaluations is determined by its visual salience. In the theoretical background of this paper, we have argued that the effect of the mode's location occurs because the bar representing the number of votes assigned to the mode is more visually salient than the bars of all other rating scores. Nonetheless, the extent to which the mode is perceptually salient and, thus, the degree to which people's attention is centered toward the mode may also be determined by how much the bar of the mode stands out from the other bars (e.g., Janiszewski 1998; Milosavljevic et al. 2012). If this is so, an increasing number of votes allotted to the mode may likewise enhance its visual salience and, thus, the extent to which it attracts people's attention. On the basis of this, we propose that:

H3: An increasing visual salience of the mode strengthens the relationship between the location of the mode and product evaluations.

To test this prediction, we constructed three rating distributions that either differed in terms of the location of the mode—i.e., below (condition 1) vs. above the average rating (condition 2 and 3)—or in terms of the visual salience of the mode—i.e., low (condition 1 and 2) vs. high (condition 3). We manipulated the mode's visual salience by varying the extent to which the mode stood out from the other ratings. In the low-salience conditions, the bar of the mode was relatively short (i.e., 37.5% of all ratings were allotted to the mode) such that it only marginally exceeded the length of bars of the remaining rating scores. In contrast, in the high-salience condition, the bar of the mode was considerably longer (i.e., 50.0% of all ratings were allotted to the mode) such that it clearly exceeded the bars of the other rating scores (see Figure 15). All three distributions shared the same number of ratings (N = 1,056), average rating (M = 3.3), and standard deviation (SD = 1.02). In contrast to the rating distributions used in Studies 1–3, all distributions used in Study 4 were left-skewed, though to a different degree (i.e., $\gamma_1 = -.42$, $\gamma_2 = -.49$, $\gamma_3 = -.86$). In this study, we employed a single factor between-subjects design with three experimental conditions such that participants faced only one of the three rating distributions.

Figure 15. Rating Distributions used in Study 4a



Note: The information that appears below the bar charts was not shown to participants.

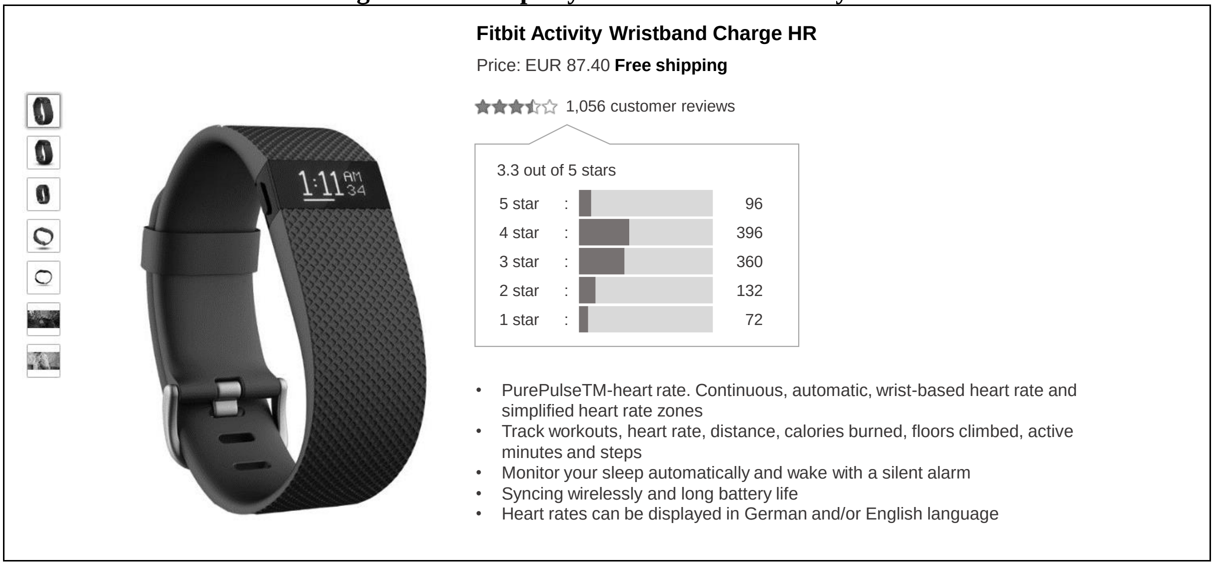
7.1 Study 4a

7.1.1 Participants, Design, and Procedure

One-hundred and forty students (M_{age} = 22.3 years, 38.6% female) participated in this online study for partial course credit. We asked participants to imagine that they were thinking about buying a fitness tracker and, thus, searched the Amazon website to get an overview of current offerings. Subjects then saw a constructed screenshot of a fitness tracker on Amazon including a product picture, product information (e.g., price and performance characteristics) as well as an overview of customer ratings through a bar chart (see Figure 16 for an exemplary stimulus). Participants were randomly assigned to one of the three conditions shown in Figure 15. After processing the provided information, participants indicated their quality perceptions of the illustrated fitness tracker ("The fitness tracker appears to perform satisfactory", "The quality of the fitness tracker seems to be better than average", "The fitness tracker appears to be better than most other fitness trackers", "I think the quality of the fitness tracker is bad/good"; α =

.76). We also measured the perceived realism of the described situation ("It was easy to imagine myself in this situation", "The situation described was realistic"; r = .62, $p < .01$).

Figure 16. Exemplary Stimulus used in Study 4a



Notes: Translated to English. The current average rating of the illustrated product on Amazon.de is very similar to the average rating used in the experiment (fitness tracker: 3.2). The product is listed among the 100 best selling items within its associated product category (i.e., 'activity trackers').

7.1.2 Results

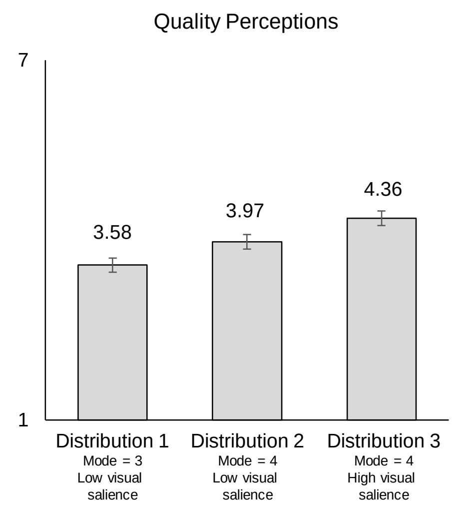
7.1.2.1 Realism Check

Answers to the realism check items (M = 5.65, SD = 1.18) indicated that respondents found the described scenarios to be realistic. Further analysis revealed that reported realism ratings were independent of the experimental conditions ($F(2, 137) = 1.80$, $p = .17$).

7.1.2.2 Hypothesis Tests

Consistent with the mode heuristic, we would predict that participants' product evaluations should be enhanced when they made their inferences based on a distribution wherein the mode is located above the average rating compared to when it is located below the mean. In addition, we expect that the difference in quality perceptions due to a varying location of the mode should increase when the mode becomes more visually salient. An overall one-way ANOVA yielded a significant difference between the three conditions ($F(2, 137) = 10.33$, $p < .01$). Consistent with our prediction, follow-up contrast analyses comparing quality inferences across the three conditions revealed the following: First, considering the two low-salience conditions (condition 1 and 2), participants who were confronted with the distribution wherein the mode was located above the average rating (condition 2) judged the quality of the presented fitness tracker as higher (M = 3.97, SD = .83) than those who were confronted with the distribution wherein the mode was located below the mean (condition 1: M = 3.58, SD = .79; Δ = .39, F(1, 137) = 5.19, p < .05; see Figure 17).

Figure 17. The Effect of Location and Visual Salience of the Mode on Quality Perceptions (Study 4a)



Note: Error bars denote standard errors.

Second, and in line with H3, this difference increased as a function of the mode's visual salience (condition 3: M = 4.36, SD = .83; condition 1 vs. 3: Δ = .78, F(1, 137) = 20.65, p < .01) such that quality perceptions were also significantly different between the low and high visual salience conditions in which the mode was located above the mean (condition 2 vs. 3: Δ = .39, F(1, 137) = 5.36, p < .05).

7.2 Study 4b

The purpose of Study 4b was to replicate the findings obtained from Study 4a in another context. However, aside from using a different context, we also changed the scale of possible rating scores from a five-point scale to a ten-point scale; this change also allowed us to move the distributions used in Study 4a up toward higher rating scores while simultaneously maintaining their basic properties (i.e., location and visual salience of the mode as well as standard deviation) and, thus, to test the robustness of the use of the mode heuristic as well as of the documented visual salience effect for distributions featuring higher average ratings (i.e., 7.3 out of 10).

7.2.1 Participants, Design, and Procedure

One-hundred and twenty-nine students (M_{age} = 21.9 years, 41.9% female) participated in this online study for partial course credit. At the beginning of the experiment, participants were asked to imagine that they had recently seen a movie trailer they found pretty appealing. Before deciding on whether or not to watch the movie in the theatre, they were visiting the IMDb website—an online movie database including, for instance, information about casts, plot summaries, and consumer reviews—to inspect the evaluations of people who had already seen the movie. Participants then saw one of the three rating distributions shown in Figure 18. In all conditions, the movie had received ratings from 176 reviewers with an average rating of 7.3 out of 10. The mode was either located below (condition 1) or above the average rating (condition 2 and 3) and was either rarely (condition 1 and 2) or highly visually salient (condition 3). Participants were randomly assigned to one of the three conditions. After processing the provided information, participants indicated their perceptions of quality of the movie ("The movie seems to be good", "The quality of the movie seems to be better than average", "The movie appears to be better than most other movies", "I think the quality of the movie is bad/good"; $\alpha = .76$) as well as their perceived realism of the described scenario ("It was easy to imagine myself in this situation", "The situation described was realistic"; $r = .79$, $p < .01$).

Figure 18. Stimuli used in Study 4b

Condition 1

Mode: 7 (below the average rating)

Visual salience: Low

Condition 2

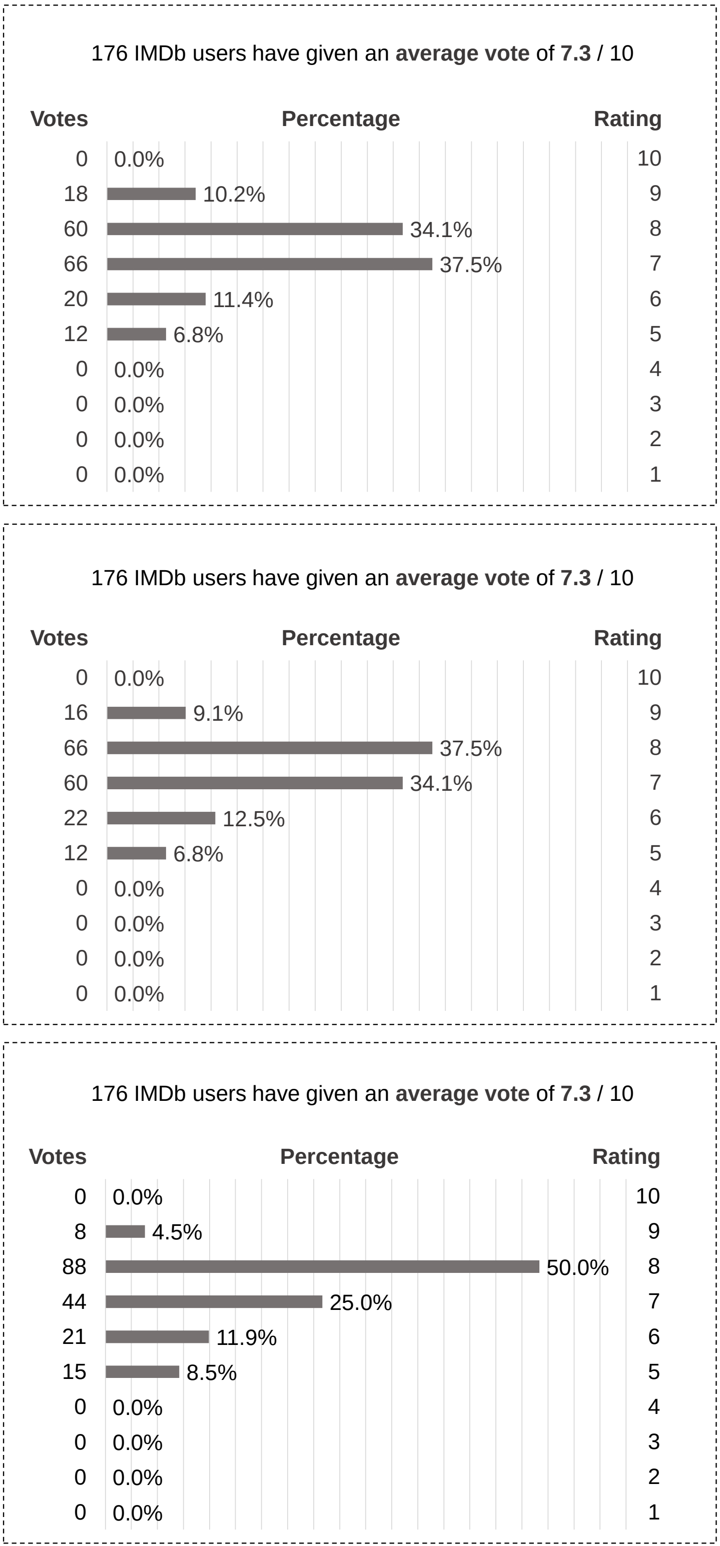
Mode: 8 (above the average rating)

Visual salience: Low

Condition 3

Mode: 8 (above the average rating)

Visual salience: High



Note: Translated to English. The information on the condition that appears on the left was not shown to participants.

7.2.2 Results

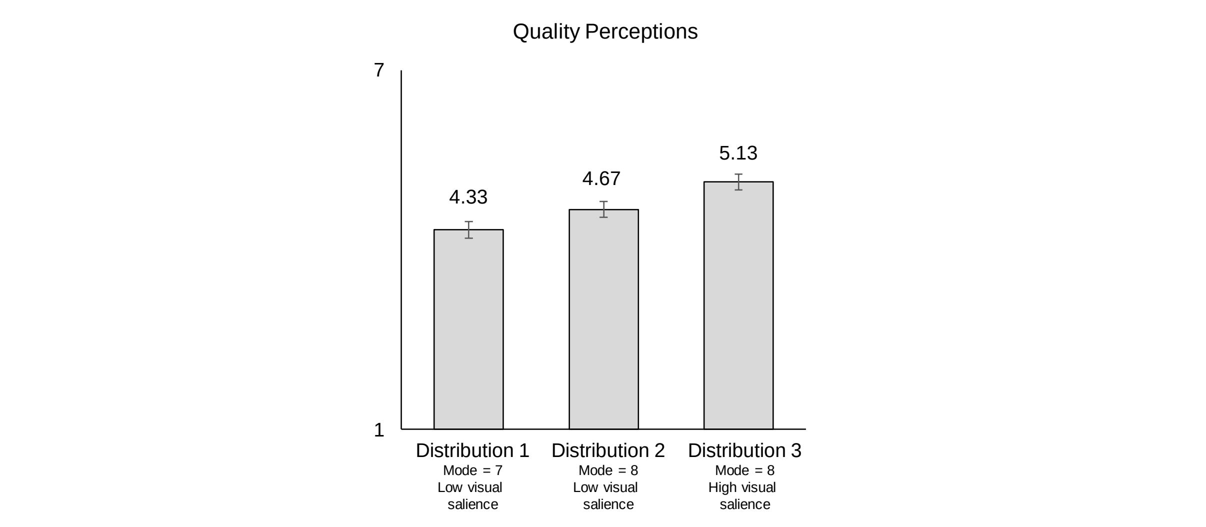
7.2.2.1 Realism Check

The calculated mean of the realism check items (M = 6.02, SD = 1.15) indicated that respondents found the described scenarios to be highly realistic. Further analysis revealed that reported realism ratings were independent of the experimental conditions ($F(1, 126) = .19$, $p = .83$).

7.2.2.2 Hypothesis Tests

An overall one-way ANOVA yielded a significant difference between the three conditions ($F(2, 126) = 8.68$, $p < .01$). Consistent with the results of Study 4a, follow-up contrast analyses comparing quality inferences across the three conditions revealed the following: First, considering the two rating distributions with a low visual salience of the mode, participants who were confronted with the distribution wherein the mode was located above the average rating (condition 2) evaluated the quality of the movie significantly higher (M = 4.67, SD = .98) than those who were confronted with the distribution wherein the mode was located below the mean (condition 1: M = 4.33, SD = .86; $\Delta = .34$, $F(1, 126) = 3.13$, $p < .10$; see Figure 19). In line with H3, this difference increased as a function of the mode's visual salience (condition 3: M = 5.13, SD = .80; condition 1 vs. 3: $\Delta = .80$, $F(1, 126) = 17.17$, $p < .01$) such that quality perceptions were also significantly different between the low and high visual salience conditions in which the mode was located above the mean (condition 2 vs. 3: $\Delta = .46$, $F(1, 126) = 6.00$, $p < .05$).

Figure 19. The Effect of Location and Visual Salience of the Mode on Quality Perceptions (Study 4b)



Note: Error bars denote standard errors.

7.3 Discussion

In sum, Study 4 provides further evidence for the existence of the mode heuristic and extends our findings in two important ways: First, in line with H3, the results of Study 4a revealed that the extent to which a shift in the location of the mode affected consumers' interpretations of rating distributions was dependent on its visual salience. Study 4b, confirmed this finding in a different context (i.e., inferences about the quality of a movie) and when rating distributions were shown on a ten-point scale (instead of a five-point scale) with a higher average rating (i.e., 7.3 out of 10 instead of 3.3 out of 5). Second, since all rating distributions considered in Study 4 were skewed to the left, we can preclude that the findings obtained from Studies 1–3 were merely driven by the diverging direction of skew of the examined rating distributions. However, it remains unclear whether the results were (at least partially) driven by a varying magnitude of skew across the distributions considered in the current study. We will address this concern in Study 5.

8 Study 5: Ruling out Alternative Explanations

Consistent with our mode heuristic account, Studies 1–4 demonstrated that product evaluations inferred from rating distributions systematically vary by the location of the mode. The focus of

Study 5 was on ruling out two alternative explanations for this effect. Specifically, since the location of the mode is typically strongly related to a distribution's median and skewness (e.g., Malhotra 2010; Moore, McCabe, and Craig 2012), it is hardly possible to manipulate the location of the mode without also altering other distribution characteristics. For instance, although all rating distributions used in Studies 1–3 had an equal average rating and standard deviation, they did not only differ regarding the location of the mode but also in terms of their median and direction of skewness. Similarly, although the distributions considered in Study 4 were all skewed to the left, they diverged in terms of the magnitude of skew. Hence, strictly speaking, we cannot explicitly preclude that the documented effects have been caused by changes in the skewness or median of the used distributions rather than by the location and visual salience of the mode as hypothesized. Hence, the purpose of Study 5 was to rule out these alternative accounts of our findings.

Based on von Hippel's (2005) observation of occasions where the interrelationships between mode, median, and skewness of a distribution are disrupted, we constructed three rating distributions with several important properties that allowed us to analyze the impacts of each of the three distribution characteristics under scrutiny (i.e., median, skewness, and location of the mode) in isolation (see Figure 20). First, all three distributions had the same median. Hence, if the detected effects were driven by changes of a distribution's median, participants' quality inferences should be equal across the three distributions. Second, the first and the second distribution merely differed in terms of their skewness; i.e., we increased the extent to which the distributions were negatively skewed. However, the skewness was kept constant between the second and the third distribution. Thus, if the skewness was responsible for the occurrence of the documented effects, quality inferences should differ only between the first and second distribution but not between the second and third. Finally, we manipulated the location of the mode from 3 (i.e., below the mean; distribution 1 and 2) to 4 (i.e., above the mean; distribution 3). Hence, if only the location of the mode was the driver of the reported effects as we predicted, then quality inferences should differ between distribution 2 and 3. We kept the average rating (M = 3.3), standard deviation (SD = 1.02), and rating volume (N = 358) constant across the three conditions.

Figure 20. Rating Distributions used in Study 5a

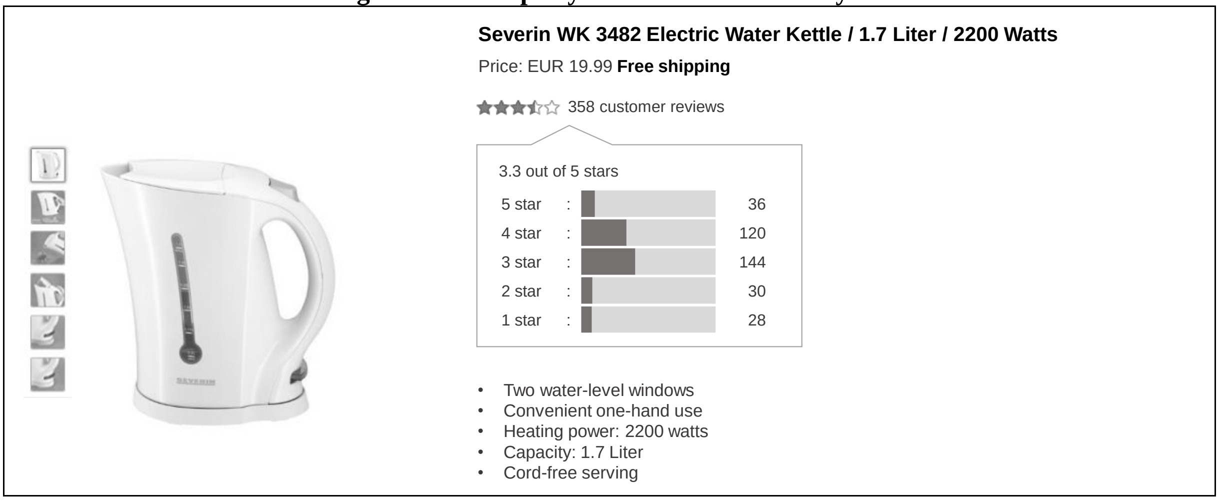
Notes: Numbers written in italics indicate changes relative to the first condition; the information on the distribution characteristics that appears below the bar charts was not shown to participants.

8.1 Study 5a

8.1.1 Participants, Design, and Procedure

One hundred and thirty-eight students (M_{age} = 21.4 years, 35.5% female) participated in this online study for extra course credit. We asked participants to imagine that they were thinking about buying an electric water kettle and, therefore, searched the Amazon website to get an overview of current offerings. Subjects were then exposed to a constructed illustration of an electric kettle on Amazon including a product picture, product information (e.g., price and performance characteristics) as well as an overview of customer ratings displayed as a bar chart (see Figure 21 for an exemplary stimulus). Participants were randomly assigned to one of the three conditions shown in Figure 20. After participants had gone through the scenario, we asked them to indicate their quality perceptions of the illustrated water kettle ("The water kettle appears to perform satisfactory", "The quality of the water kettle seems to be better than average", "The water kettle appears to be better than most other water kettles", "I think the quality of the water kettle is bad/good"; α = .79). We also captured participants' perceived realism of the scenario ("It was easy to imagine myself in this situation", "The situation described was realistic"; r = .72, $p < .01$).

Figure 21. Exemplary Stimulus used in Study 5a



Notes: Translated to English. The current average rating of the illustrated product on Amazon.de is very similar to the average rating used in the experiment (water kettle: 3.9). The product is listed among the 100 best selling items within its associated product category (i.e., 'electric kettles').

8.1.2 Results

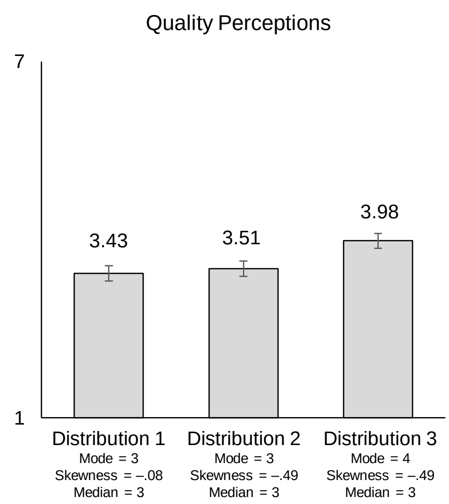
8.1.2.1 Realism Check

The calculated mean of the realism check items (M = 5.55, SD = 1.29) indicated that respondents found the described situation to be realistic. Further analysis revealed that the reported realism ratings were not significantly different across the three experimental conditions ($F(2, 135) = 1.14$, $p = .32$).

8.1.2.2 Main Analyses

An overall one-way ANOVA on quality perceptions revealed a significant difference between the three conditions ($F(2, 135) = 5.51$, $p < .01$). Since we kept the median constant across the three conditions, we can preclude that it was the activator of this effect. Planned contrasts revealed that there was no difference in quality perceptions between the first (M = 3.43, SD = .87) and the second condition (M = 3.51, SD = .80, $F(1, 135) = .16$, $p = .69$; see Figure 22). As the skewness of the presented rating distributions was the only difference between the first and second condition, this null effect rules out that an increasing magnitude of skew per se is instrumental in influencing product inferences. However, consistent with our mode heuristic account, quality inferences from the third rating distribution wherein the mode was located above the average rating were significantly higher (M = 3.98, SD = .91) than those inferred from the first (F(1, 135) = 9.25, $p < .01$) and the second distribution (F(1, 135) = 7.04, $p < .01$) wherein the mode was situated below the mean. In particular, the difference in quality perceptions between the second and the third condition, wherein only the location of the mode differed (i.e., both skewness and median were equal across the two conditions) precludes that other distribution characteristics are essential in producing the effect of the location of the mode on consumers' product inferences.

Figure 22. The Effect of Location of the Mode, Skewness, and Median on Quality Perceptions (Study 5a)



Note: Error bars denote standard errors.

8.2 Study 5b

The purpose of Study 5b was to replicate the findings obtained from Study 5a when considering vertical instead of horizontal bar charts.

8.2.1 Participants, Design, and Procedure

One hundred and forty-eight students (M_{age} = 21.8 years, 37.8% female) participated in this online study for partial course credit. Participants were asked to imagine that they were creating their course schedule for the upcoming summer term and found a course description which sounded appealing. Before making a choice whether or not to enroll in this course, they were checking previous course evaluations on the department's website. Subjects were then exposed to one of the three rating distributions shown in Figure 23. Unlike Study 5a, instead of presenting the frequency of each rating score via horizontal bars, in this study we used vertical bars to illustrate the rating distribution. Participants were randomly assigned to one of the three conditions. After processing the provided information, participants indicated their perceptions of quality of the course ("The course seems to be good", "The quality of the course seems to be better than average", "The course appears to be better than most other courses", "I think the quality of the course is bad/good"; α = .79) as well as the perceived realism of the applied scenarios ("It was easy to imagine myself in this situation", "The situation described was realistic"; r = .71, $p < .01$).

Figure 23. Stimuli used in Study 5b

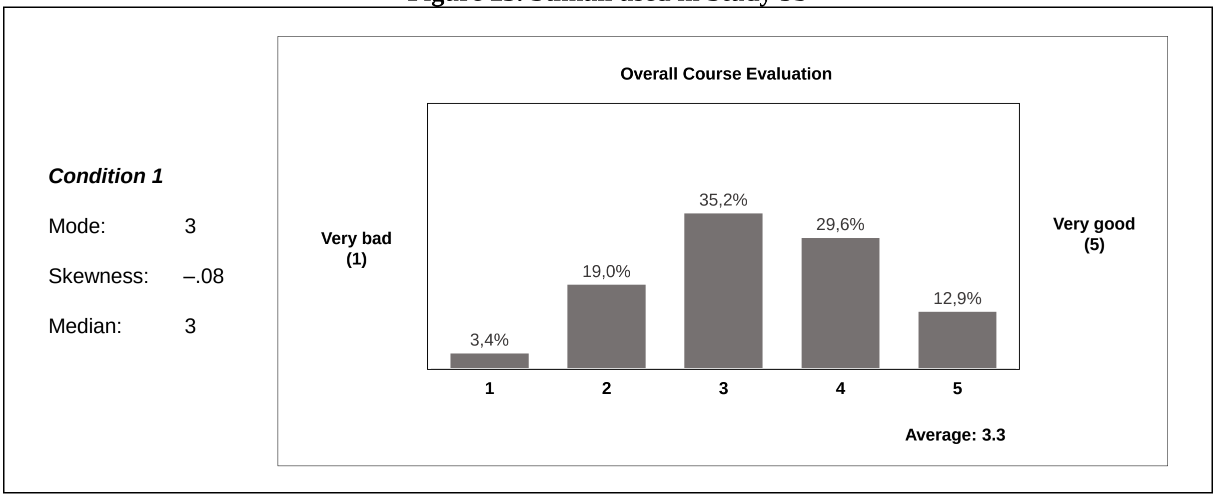
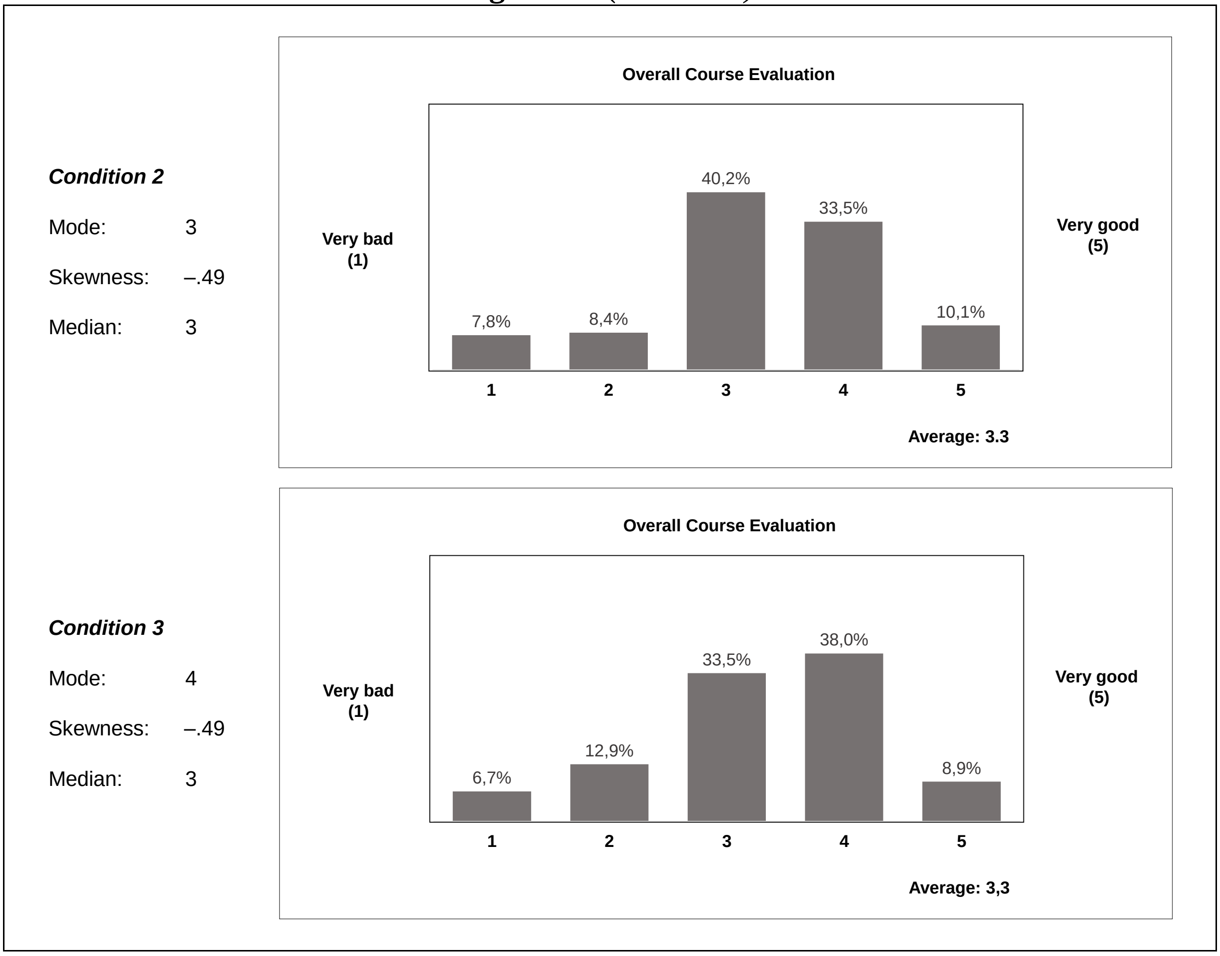


Figure 23. *(continued)*



Note: Translated to English. The information on the condition that appears on the left was not shown to the participants.

8.2.2 Results

8.2.2.1 Realism Check

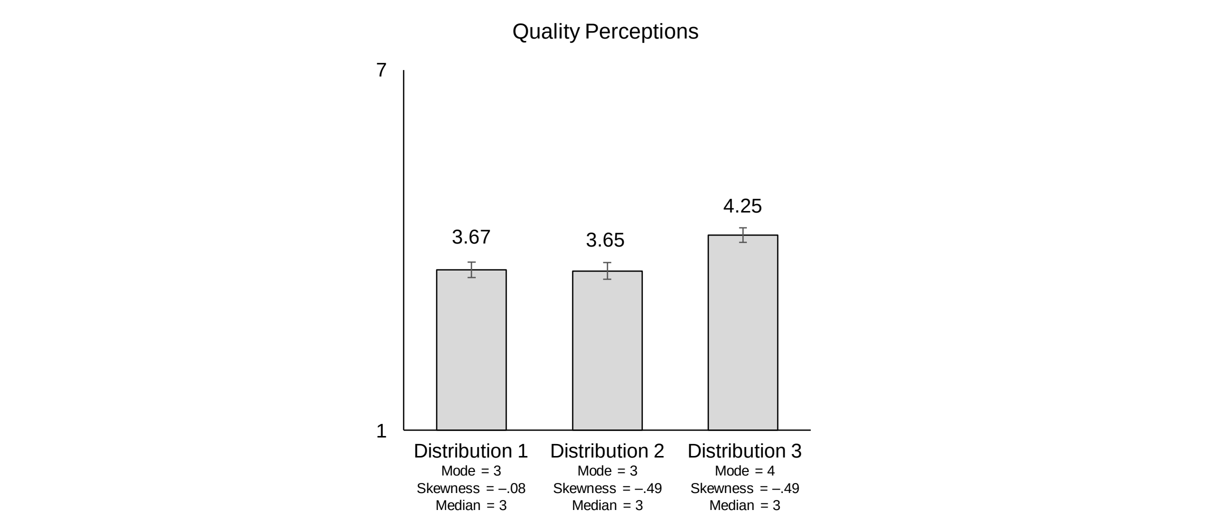
The calculated mean of the realism check items (M = 5.75, SD = 1.23) indicated that respondents found the described scenarios to be highly realistic. Further analysis revealed that reported realism ratings were independent of the experimental conditions (F(2, 145) = 2.022, n.s.).

8.2.2.2 Main Analyses

An overall one-way ANOVA on quality perceptions revealed a significant difference between the three conditions (F(2, 145) = 6.99, *p* < .01). As in Study 5a, follow-up contrast analyses revealed that course quality was rated as significantly higher in the third condition wherein the mode was located above the average rating (M = 4.25, SD = .89) than in the first (M = 3.67, SD

= .90; F(1, 145) = 10.32, $p < .01$) and in the second condition (M = 3.65, SD = .93; F(1, 145) = 10.46, $p < .01$; see Figure 24) wherein the mode was located below the mean. There was no difference in quality perceptions between the first and the second condition (F(1, 145) = .01, $p = .92$).

Figure 24. The Effect of Location of the Mode, Skewness, and Median on Quality Perceptions (Study 5b)



Note: Error bars denote standard errors.

8.3 Discussion

The findings of Study 5 support our theorizing about the mode heuristic and preclude that other distribution characteristics that are typically strongly connected to the location of the mode within a distribution can be held responsible for the identified effects. Precisely, the results of this study suggest that product inferences from rating distributions can be affected by the location of the mode independent of any changes in the skewness of a distribution or its median. Additionally, Study 5b replicates the results of the experiment in Study 5a when considering vertical (instead of horizontal) bar charts. Hence, this study revealed that the use of the mode heuristic cannot only be observed when considering horizontal bar charts but also vertically oriented bar charts and, thus, provides a more holistic picture of bar charts used in business practices.

9 Study 6: Evidence from the Marketplace

The purpose of Study 6 was to substantiate the robustness of our findings by examining whether the demonstrated effects of the mode can also be observed in actual purchase behavior. For this purpose, we collected customer review data from Amazon.de. Amazon is one of the most popular online retailers in Germany with a sales volume of more than $14 billion per year (Amazon 2017). In addition, Amazon's website provides all information necessary for our investigation; i.e., the distribution of rating scores for each product as well as their Amazon Bestseller Ranks as a sales performance indicator.

9.1 Data Description

9.1.1 Data Collection

In Mai 2017, we collected customer review data for the 100 best selling items within 20 consumer electronics product categories including, for example, printers, toasters, fitness trackers, and electric water kettles (see Table 8 for a complete list of all product categories). We restricted the data set to products (1) that were actually available for purchase at the time of data collection and (2) that had already been reviewed by five or more customers. Finally, we removed all items (3) that had been assigned to a wrong best seller list (e.g., we found coffee machines that appeared in the best seller list of electric kettles) and (4) excluded 35 products whose rating distribution did not exhibit a unique, unambiguous mode (i.e., there were multiple rating scores that have likewise received the largest number of votes). In sum, the final data set consisted of 1,536 items that fulfilled all restriction criteria.

Table 8. Product Categories in Study 6

Category	N	Average Rating	Average Price	Average Number of Ratings
Blu-ray players	59	3.87 (.40)	176.86 (163.67)	146.05 (229.52)
Body weight scales	90	4.05 (.44)	35.00 (32.51)	175.81 (337.54)
Coffee machines	92	4.07 (.41)	65.35 (42.88)	151.95 (354.32)
Clock radios	60	3.98 (.49)	43.21 (28.25)	139.30 (252.22)
Electric shavers	55	4.13 (.35)	112.55 (83.23)	147.16 (239.36)
Electric water kettles	85	4.12 (.31)	33.04 (15.58)	174.38 (217.00)
Fitness trackers	44	3.61 (.81)	46.48 (30.30)	412.23 (533.50)
Hair dryers	92	4.23 (.43)	32.45 (18.08)	201.90 (370.30)
Headphones	80	4.14 (.32)	76.26 (82.05)	315.90 (355.72)
Laptop computers	77	3.89 (.42)	507.00 (267.12)	66.77 (59.08)
Lawn mowers	76	4.03 (.48)	234.26 (118.00)	77.92 (143.42)
Microwaves	50	4.07 (.32)	151.07 (94.27)	89.82 (137.48)
Monitors	93	4.25 (.26)	239.03 (166.18)	166.88 (292.36)
Printers	96	3.85 (.33)	129.01 (82.81)	173.92 (195.48)
Projectors	50	4.31 (.26)	520.23 (231.27)	72.00 (93.05)
Reflex cameras	94	4.48 (.26)	769.30 (535.56)	121.68 (130.20)
Smartphones	95	4.09 (.38)	287.42 (165.32)	463.53 (524.98)
Television Sets	90	4.00 (.31)	449.16 (300.52)	100.83 (95.47)
Toasters	77	4.11 (.37)	40.23 (19.55)	124.30 (207.28)
Vacuum cleaners	81	4.26 (.31)	156.42 (91.66)	149.15 (294.30)

Note: SD in parentheses.

9.1.2 Measures

For each product, we extracted the number and mean value of ratings, as well as the frequency of each rating score (e.g., the number of 5 star, 4 star, 3 star ratings, and so forth). Based on this data, we identified the mode of each product's rating distribution. As a measure of the location of the mode we subtracted the average rating from the mode such that positive (negative) values on this measure indicated that the mode was located above (below) the distribution's mean. As a proxy for the mode's visual salience, we used the percentage of votes that have been assigned to the distribution's mode; reflecting the absolute length of the mode's bar on the Amazon website. In addition, we also calculated the skewness of each product's rating distribution as well as its median to provide further evidence that alternative accounts for the effects of a distribution's mode as well as its visual salience do not apply. Building on the finding that the natural log of Amazon's best seller ranks is a negative linear function of a product's logarithmic sales (e.g., Brynjolfsson, Hu, and Smith 2003), we used the log of the sales rank within the product category as an inverse proxy for sales performance (see also Chevalier and Mayzlin 2006; Sun 2012). To control for external influences, we also collected the selling price of each item and included it as a control variable into our analyses (see also Sun 2012).

9.2 Results

We first regressed log of sales rank on the location of the mode, including average rating, the number of ratings, and price as covariates. In order to neutralize all category specific differences in the absolute levels of the predictor variables, we z-standardized all independent variables by product category before analysis such that they had a mean of zero and a standard deviation of one (see de Langhe et al. 2016a for a similar approach). The parameter estimates appear in Table 9 (Model A). In line with our experimental studies, a significant effect of the location of the mode on sales performance emerged ($\beta = -.064$, $t(1531) = 2.425$, $p < .05$) such that the degree to which a distribution's mode exceeded the average rating corresponded with better sales ranks. Next, we included the visual salience of the mode as well as its interaction term with location of the mode into the regression model (Model B). Consistent with H3, the interaction between the location and visual salience of the mode turned out to be significant ($\beta = -.110$, $t(1529) = 3.488$, $p < .01$); indicating that an increasing visual salience strengthened the relationship between location of the mode and sales performance. In a final step, we added the distributions' median (Model C) and skewness (Model D) to the regression model. In line with the findings obtained in Study 5, neither median ($\beta = -.056$, $t(1528) = 1.308$, $p = .19$) nor skewness ($\beta = .018$, $t(1515) = .382$, $p = .70$) were significantly related to sales ranks; ruling out their explanatory power in the identified effects.

9.3 Discussion

Using a huge sample of customer review data from Amazon, Study 6 provides strong evidence for the robustness of the effects of the location and visual salience of the mode in a real-world setting. Consistent with our mode heuristic account, the results revealed that—even when controlling for the effects of rating volume, average rating, and product price—sales performances were positively affected by the extent to which the location of the mode exceeded the average rating; thus, replicating the core findings of our experimental studies and providing additional support for our key hypothesis (H1). In addition, the results demonstrated that the strength of the effect caused by variations of the location of the mode was determined by its visual salience; consistent with H3. Finally, we did not find any evidence that the skewness or the median of a rating distribution could account for these effects; supporting the findings of Study 5.

Table 9. Parameter Estimates (Study 6)

	Model A		Model B		Model C		Model D[1]	
Independent variables	β	t-value	β	t-value	β	t-value	β	t-value
Hypotheses								
Location of the mode	–.064	2.425**	–.141	4.006***	–.104	2.313**	–.148	4.068***
Visual salience of the mode			–.031	.684	.003	.062	–.017	.355
Location × visual salience of the mode			–.110	3.488***	–.085	2.329**	–.116	3.593***
Alternative explanations								
Median					–.056	1.308		
Skewness							.018	.382
Control variables								
Average rating	–.094	3.541***	–.117	2.624***	–.088	1.772*	–.108	2.069**
Number of ratings	–.335	14.114***	–.327	13.754***	–.325	13.632***	–.329	13.749***
Price	.141	5.937***	.139	5.894***	.140	5.921***	.138	5.823***
R^2	.146		.153		.154		.154	

[1]The estimation of Model D is based on 1,523 observations. Thirteen products have been excluded from analysis because the calculation of the skewness entailed an invalid operation (i.e., division by zero); Additional Note: $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$.

10 General Discussion

The present research provides a systematic examination of how consumers' interpretations of rating distributions illustrated as a bar chart are determined by the location of the mode; i.e., the rating score that has received the highest number of votes which is, thus, represented by the most salient bar within the graph. Using data from a series of five experimental studies as well as secondary data, we find support for a tendency to make inferences about the quality of a product based on the location of the mode; a phenomenon we refer to as the mode heuristic. More precisely, across our studies, we (1) provide strong empirical evidence for consumers' use of the mode as a heuristic basis for pre-purchase product evaluations in a variety of different contexts, (2) shed light on this phenomenon at the process level, and (3) demonstrate how quality inferences based on the mode heuristic depend on the visual salience of the mode. In the following sections, we discuss the contribution of our research to theory, implications for business practices, and opportunities for further research.

10.1 Theoretical Contributions

With this research, we add to the large body of literature studying the impacts of product ratings in consumers' online shopping behavior (see Babić Rosario et al. 2016 for an overview). To date, extant research on consumers' response to different distribution characteristics of online ratings has mostly focused on the effects of the number of ratings, average ratings, as well as rating dispersion, while our investigation is the first that places the mode of rating distributions under scrutiny. By establishing this previously neglected distribution characteristic as an important parameter in people's interpretations of rating distributions we broaden our understanding regarding the impacts of customer ratings on pre-purchase product evaluations in online shopping environments.

On a more general note, we contribute to the literature on graphical perceptions. Our results are consistent with the idea that people's interpretations of graphical formats are not only determined by the provided content per se but also by the physical properties of the elements of a chart (e.g., Jarvenpaa 1990; Simkin and Hastie 1987; Spence 1990; Stone et al. 1997; Stone et al. 2003; Weber and Kirsner 1997). More precisely, we found that interpretations of graphical illustrations of frequency distributions are crucially affected by the most perceptually salient element; i.e., the mode of a distribution. However, in contrast to extant studies in this field, we also provide explicit process evidence using an eye-tracking method (Study 2). Our results reveal that, in fact, the attention paid to an object is dependent on its visual salience and that the allocation of visual attention across different elements of a graphical display determines people's conclusions drawn from it. Thereby, we respond to several calls for research to provide direct evidence for the process underlying salience effects in people's interpretations of graphical formats (e.g., Jarvenpaa 1990; Raghubir and Das 2010; Stone et al. 2003).

Finally, our study complements the wide array of research on heuristics in judgment and decision making documenting the use of a variety of simple cues and rules of thumb as a simplification of evaluation processes (see Gilovich et al. 2002). In extension of this field of research, we demonstrate that the mode of a rating distribution can serve as such a heuristic basis when processing and interpreting graphical displays illustrating the distribution of product rating scores. Therefore, the documented mode heuristic aligns with other previously reported heuristics entailing that the weighting of available informational cues in judgment formation deviates from a normative mindset; e.g., the anchoring (Tversky and Kahneman 1974) and availability heuristic (Tversky and Kahneman 1973), as well as the peak-end rule (Kahneman et al. 1993).

10.2 Managerial Implications

Our findings have important implications for business practices. By highlighting the relevance of the mode in consumers' product inferences from online rating distributions we provide marketers a new key figure which—aside from the number of ratings, average ratings, and rating dispersion—should be involved when monitoring, analyzing, and evaluating review data. In other words, with our results in mind, marketers can better anticipate the consequences of different rating distributions of their offerings on customers' pre-purchase product quality evaluations and, consequently, on product sales performances. In times when consumers place more trust in the opinions of unknown people posted online than in any form of communication initiated by a company (Nielsen 2015), it is essential to develop a deeper understanding of how consumers respond to the abundance of product evaluations provided by previous customers that circulate in the marketplace.

10.3 Limitations and Future Research Directions

Although our work provides valuable new insights into customers' interpretations of rating distributions it has some limitations that offer promising opportunities for further research. First and foremost, since our studies highlight the important role of the mode of rating distributions, our findings might encourage researchers to gain deeper insights into this distribution characteristic. In this vein, a worthwhile issue for future research might be to investigate whether the mode is associated with a semantic meaning. For instance, it is conceivable that the mode is interpreted as the majority's opinion. Since the majority's position has been demonstrated to be (sometimes disproportionately) influential in attitude formation (e.g., Mackie 1987; Martin and Hewstone 2003; Martin, Hewstone, and Martin 2007), this may further explain the use of the mode as a heuristic basis for quality judgments beyond our reasoning about its eye-catching physical salience. In this context, it would be interesting to examine whether people use the mode heuristic consciously or on a non-conscious level (Gigerenzer and Gaissmaier 2011). People may deliberately concentrate on the mode when interpreting rating distributions (e.g., because of its majority status) or, alternatively, the use of the mode heuristic could be a rather automatic process. Furthermore, future research could explore conditions under which the mode heuristic is more or less likely to be used and, thereby, contribute to an explanation as to when and why different locations of the mode influence product perceptions and subsequent purchase behavior. Building on factors that have been found to prompt heuristic (rather than systematic) information processing, potentially relevant aspects may include situational factors such as time constraints (e.g., Nowlis 1995; Suri and Monroe 2003) and outcome relevance (e.g., Martin et al. 2007) as well as personal characteristics such as motivation, task involvement, and need for cognition (e.g., Chaiken 1980; Maoz and Tybout 2002).

Second, in our studies, we have established the mode heuristic in consumers' response to rating distributions in terms of inferred evaluations of the quality of a reviewed product. However, investigating whether the mode heuristic is also applied in other judgmental tasks in the context of online ratings might be another promising route for future research. For instance, extant literature has suggested that consumers' rating behavior is influenced by already existing ratings (e.g., Moe and Trusov 2011; Sridhar and Srinivasan 2012; see also Pincus and Waters 1977). These studies, however, typically focus on the relationship between rating valence in terms of average ratings and subsequent ratings. Given the reported effects of the mode on people's pre-purchase product evaluations it would be worthwhile to examine whether consumers also anchor their (post-purchase) assessments on the mode of an existing rating distribution when giving a product rating in a similar manner.

Third, the aggregated summaries of customer ratings by means of bar charts in our studies were strongly geared to the illustrations used by marketers in the real world. However, extant studies on graphical perceptions have examined and compared people's interpretations of a variety of different types of graphs (i.e., pie charts, line graphs, and scatter diagrams) in a wide array of different contexts (e.g., financial and health risks). Consequently, future research could investigate, whether the overpowering effects of the most perceptually salient element within a chart are robust among different graph formats, informational contents, and evaluation contexts beyond product evaluations.

D Empirical Research Paper 2: Should We Reach for the Stars? Examining the Convergence between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance

Abstract

By documenting that online product ratings poorly correlate with quality scores provided by Consumer Reports—presumably a measure of 'objective' product quality—de Langhe et al. (2016a) found that consumers rely more heavily on such ratings when making quality inferences than they should. Aside from replicating this finding, this research examines how the convergence between objective and rated quality alters over the product life cycle and investigates which quality indicator is a better predictor of sales performance.

Additional note:
A shorter version of this paper, co-authored by Sören Köcher, has been accepted for publication in the *Journal of Marketing Behavior* (Köcher, Sarah, and Sören Köcher, "Should We Reach for the Stars? Examining the Convergence between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance," *Journal of Marketing Behavior*).

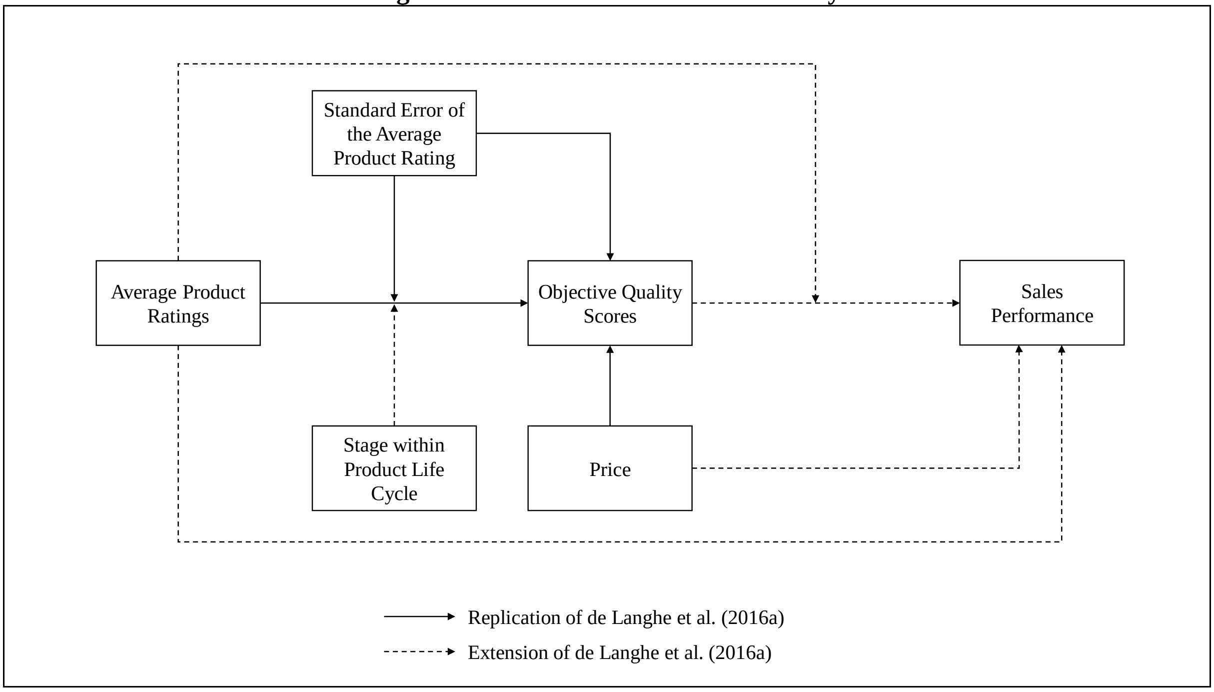
1 Introduction

Today, when making purchase decisions consumers increasingly use online product reviews to draw conclusions about the quality of the available purchase options (e.g., Hu, Liu, and Zhang 2008; Li and Hitt 2008; Simonson and Rosen 2014). Unsurprisingly, an abundance of research has documented that these reviews are extremely influential in driving sales and related performance metrics (e.g., Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Moe and Trusov 2011; see Babić Rosario et al. 2016 for a recent meta-analysis). However, a few studies raised doubts about whether online ratings can actually reflect the ‘true’ quality of a product (e.g., de Langhe, Fernbach, and Lichtenstein 2016a; Hu, Pavlou, and Zhang 2006; Koh, Hu, and Clemons 2010). For instance, de Langhe et al. (2016a) recently discovered a substantial gap between the extent to which consumers trust in online product ratings when making inferences about the quality of a product and the actual validity of such ratings as an indicator of a product’s ‘objective’ performance. More precisely, across a series of consumer studies the authors found that people place enormous weight on the average product rating when making predictions about the quality of a product, while the convergence between average ratings and the quality scores provided by *Consumer Reports* (CR)—presumably a measure of objective quality (see also Gerstner 1985; Hardie, Johnson, and Fader 1993; Lichtenstein and Burton 1989)—and, thus, their validity as a quality indicator, is evidentially weak. These findings have caused a lively discussion among several marketing researchers (de Langhe et al. 2016b; Kozinets 2016; Simonson 2016; Winer and Fader 2016) primarily questioning the actual relevance of the reported results, the reliability of CR scores as a measure of objective quality, as well as the simplicity of analysis neglecting consumer heterogeneity and dynamic changes in product ratings over time. However, as de Langhe et al. (2016b) note “these assertions are proffered without a shred of evidence” (p. 852).

The purpose of the present study is to contribute to this debate by empirically testing three of the critics’ annotations: First, Simonson (2016) doubted that CR scores actually capture objective product quality; inter alia referring to an occasion where CR's methodology has come under severe criticism. We tested whether the claimed inaccuracy and distortedness of CR ratings per se are the ultimate source of the low convergence between rated and objective quality by replicating de Langhe et al.'s (2016a) findings using a database very similar to the one they have used. However, instead of CR scores, we collected the quality scores provided by *Stiftung Warentest*—the German equivalent of CR—and inspected their convergence with average customer ratings provided on Amazon's German website. Second, Winer and Fader (2016) asserted that the correlation between rated and objective quality is determined by the dynamics of reviews and criticize the lack of dynamic aspects in the original study. Among other considerations, they suggested that the correlation between objective and rated quality may change over a product's life cycle. Inspired by this assumption we examined potential differences in the convergence between objective and rated quality scores across older and newer products and, thereby, extend the original work. Finally, motivated by a further suggestion of the discussants (Simonson 2016; Winer and Fader 2016) we investigated the extent to which different pieces of quality information influence purchase behavior by examining the impact of rated and objective quality on sales performance. Figure 25 illustrates the relationships put under scrutiny in the present study.

The rest of this paper is organized as follows: We first decribe the data set used throughout the analyses documented in this paper. We then replicate the findings reported in de Langhe et al. (2016a) about the weak convergence between average product ratings and measures of objective quality. In a next step, we document two additional analyses centered toward changes in the relationship between rated and objective quality over time as well as their concurrent impacts on sales performances. We conclude with a discussion of the contributions of our findings and provide recommendations on when and why consumers should be rather reluctant to trust online ratings as an indicator of product quality.

Figure 25. Overview of the Present Study



2 Data Description

The German consumer organization *Stiftung Warentest* publishes a monthly magazine with tests of a variety of consumer products. We downloaded all tests of consumer electronic products that had been published between 2014 and 2017 from the organization's website and extracted the reported quality scores[8] for each of the tested items. This resulted in quality ratings for 2,473 products across 352 categories. As in de Langhe et al. (2016a), we defined product categories at the lowest level of abstraction (e.g., we considered over-ear and on-ear Bluetooth headphones as separate categories). In addition, if a category had been tested multiple times during the observation period, we treated each test as an individual subcategory (e.g., smartphones tested in February 2014 and smartphones tested in November 2017 were considered as separate subcategories) such that items within a subcategory were relatively homogeneous and, thus, quality ratings were comparable. We then searched the Amazon.de website for each product for which we had a *Stiftung Warentest* score and recorded all customer

[8] In contrast to the quality scores provided by *Consumer Reports*, *Stiftung Warentest* scores range from 1 to 6; with lower values indicating better quality. To facilitate the comparison of our results with those reported by de Langhe et al. (2016a), we reversed these scores in our data set.

product ratings, selling prices, bestseller ranks, and launch dates (i.e., the date first available). Out of the 2,473 products that were evaluated by *Stiftung Warentest*, we were able to find 1,833 products across 339 categories on the Amazon.de website. Subsequently, we restricted the database to items that have been rated five or more times, and categories comprising at least three products. Our final data set consisted of 1,322 products across 224 categories with, in total, 239,906 ratings. See Appendix for a complete list of all product categories.

3 Replications

3.1 Simple Correlations

Analogously to de Langhe et al. (2016a), we first calculated the Pearson correlation between average ratings and *Stiftung Warentest* scores for each of the 224 product categories. Similar to the original findings, the average correlation was only 0.18 (vs. $\bar{r}_{\text{de Langhe et al. 2016a}} = 0.18$), and a proportion of 36.3% (vs. 34% in de Langhe et al. 2016a) of the correlations were negative.

3.2 Regression Analyses

We then examined the convergence between average ratings and *Stiftung Warentest* scores using regression analyses. In a first step, we regressed *Stiftung Warentest* scores on the average rating, the standard error (SE) of the average rating—as a measure of the accuracy of the average rating—and the interaction between the average rating and the SE. As in de Langhe et al. (2016a), we standardized all variables by subcategory before analysis such that they had a mean value of zero and a standard deviation (SD) of one. A comparison of parameter estimates and confidence intervals (CIs) with de Langhe et al.'s (2016a) results appears in Table 10.

Table 10. Parameter Estimates (and Confidence Intervals) for the Original and Present Study

	De Langhe et al. (2016a)			Present study (replication)		
	Model A	Model A’	Model B	Model A	Model A’	Model B
Dependent variable		*Consumer Reports* quality scores			*Stiftung Warentest* quality scores	
Independent variables						
Average rating	0.16 (0.10 to 0.22)	not reported	0.09 (0.03 to 0.15)	0.13 (0.07 to 0.19)	0.22 (0.14 to 0.30)	0.08 (0.02 to 0.13)
Price			0.34 (0.28 to 0.39)			0.31 (0.26 to 0.36)
Standard error	–0.13 (–0.20 to –0.07)		–0.15 (–0.21 to –0.09)	–0.15 (–0.21 to –0.09)		–0.18 (–0.23 to –0.12)
Average rating × standard error	–0.06 (–0.12 to –0.01)		–0.07 (–0.12 to –0.02)	–0.01 (–0.06 to 0.04)		–0.01 (–0.05 to 0.04)
Number of ratings		0.12 (0.07 to 0.18)			0.15 (0.10 to 0.21)	
Standard deviation		0.06 (–0.01 to 0.13)			0.08 (0.00[a] to 0.16)	
Data source of independent variables		Amazon.com			Amazon.de	
Number of observations		N = 1,272 products across 120 categories			N = 1,322 consumer electronic products across 224 categories	

Note: [a]CI_{95}: 0.002 to 0.158.

In line with the original findings, a weak relationship between average ratings and objective quality scores emerged (*b* = 0.13, CI_{95}: 0.07 to 0.19; see Model A). However, while we also found a negative main effect of SE (*b* = –0.15, CI_{95}: –0.21 to –0.09) on *Stiftung Warentest* scores, we could not support the reported interaction between average rating and SE (*b* = –0.01, CI_{95}: –0.06 to 0.04); i.e., the correspondence between average ratings and *Stiftung Warentest* scores in our sample was independent of the SE.

We then regressed *Stiftung Warentest* scores on the two components of the SE—i.e., the number of ratings and the SD—in addition to the average rating (see Model A') to shed light on where the detected negative effect of SE came from. Similar to de Langhe et al. (2016a), we found that the number of ratings (*b* = 0.15, CI_{95}: 0.10 to 0.21) was positively related to *Stiftung Warentest* scores; consistent with de Langhe et al.'s (2016a) conjecture that products with higher objective quality scores might be more popular or sold for a longer period of time, which would lead to a larger number of ratings. In addition, the relationship between the SD of ratings and *Stiftung Warentest* quality scores (*b* = 0.08, CI_{95}: 0.00 to 0.16) turned out to be significant; indicating that people seem to stronger agree in their evaluations of products with a high objective quality than in their assessments of low quality items.

Subsequently, we benchmarked the effect of average ratings on *Stiftung Warentest* scores against that of price (see Model B). This analysis revealed significant main effects of average rating (*b* = 0.08, CI_{95}: 0.02 to 0.13), SE (*b* = –0.18, CI_{95}: –0.23 to –0.12), and price (*b* = 0.31, CI_{95}: 0.26 to 0.36) on *Stiftung Warentest* scores. Again, the interaction between average rating and SE was not significant (*b* = –0.01, CI_{95}: –0.05 to 0.04). In a final step, we computed squared semipartial correlations in order to evaluate the relative amount of unique variance in *Stiftung Warentest* scores explained by price and average rating. Interestingly, price uniquely explained 9.61% of the variance in *Stiftung Warentest* quality scores, 21 times more than average rating

($sr^2_{average\ rating}$ = 0.46%); suggesting that the price of a product is a much better indicator of its quality than its average rating.

3.3 Discussion

In sum, our findings reveal that the convergence between average product ratings and objective quality scores is considerably weak and, thus, match up with the results reported by de Langhe et al. (2016a). Consequently, average ratings should be considered a rather imprecise predictor of a product’s objective performance. In addition, given that we used quality scores provided by *Stiftung Warentest* we can preclude that the CR scores used by de Langhe et al. (2016a) were responsible for the detected low convergence between rated and objective quality (Simonson 2016). However, it should be noted that our analyses did not support the interaction between average rating and its SE reported by de Langhe et al. (2016a); implying that average product ratings in our study setting did not become better predictors of objective product performance as the number of ratings increased and/or rating dispersion decreased.

4 Extensions

4.1 Does the Convergence between Rated and Objective Quality Change over the Product Life Cycle?

4.1.1 Theoretical Considerations

In their commentary on the article by de Langhe et al. (2016a), Winer and Fader (2016) speculated that the correlation between online ratings and objective performance may be different early in a product’s life cycle versus later. This conjecture is in line with extant literature suggesting that early buyers of a new product tend to have greater experience with similar products and might be more knowledgeable than later adopters (e.g., Dee Dickerson and Gentry 1983; Hirschman 1980; Huh and Kim 2008; Park and Kim 2008). Thus, they may also be able to judge a product's quality more accurately. As a consequence, the correlation between early adopters' ratings and objective quality may be higher than that of later adopters. Since early adopters are those who write the first reviews (Li and Hitt 2008), the correspondence of average ratings of products situated in early stages of their life cycle with objective quality scores should be stronger than that of older products which have been longer available for being reviewed and, thus, their average ratings might be contaminated by product evaluations from later buyers with limited abilities to appropriately evaluate a product's quality. Accordingly, the correspondence should decline as a function of progress through the product life cycle. Thus, in line with Winer and Fader's (2016) assertion, we propose:

H1: *The convergence between average ratings and objective performance will decrease over the product life cycle.*

4.1.2 Data Description

To test this postulate, we screened our database for product categories that have been evaluated by *Stiftung Warentest* more than once during the last four years. This resulted in a data set of 546 products across 29 categories that have been tested, on average, 2.8 times. As an approximation for a product's stage in its life cycle, we calculated for how long each product had already been available on Amazon.de using its launch date (see Babić Rosario et al. 2016 for a similar approach). The average age of the products in this data set was 32.6 months and the average range of product age within the 29 categories was 35.2 months.

4.1.3 Results

We regressed *Stiftung Warentest* scores on the average rating, the age of the product, and the interaction between average rating and product age. Before running the analysis, we z-standardized *Stiftung Warentest* scores and average ratings by each test of each product category (e.g., smartphones tested in February 2014 vs. smartphones tested in November 2017) and standardized product age by category (e.g., smartphones). As shown in Table 11 (Model C), this analysis revealed a positive main effect of average rating on *Stiftung Warentest* scores (b = 0.17, CI_{95}: 0.09 to 0.25) qualified by a significant average rating × product age interaction (b = –0.11, CI_{95}: –0.20 to –0.03). In support of H1, the relationship between average ratings and *Stiftung Warentest* quality scores was stronger for newer products in our data set (–1 SD from the average age: b = 0.28, CI_{95}: 0.16 to 0.40) than for older products (+1 SD from the average age: b = 0.10, CI_{95}: –0.06 to 0.18). The main effect of product age was not statistically significant (b = –0.01, CI_{95}: –0.09 to 0.07). Controlling for product price did not affect the direction or significance of these results (see Model D).

Table 11. The Moderating Effect of Product Age on the Relationship between Rated and Objective Product Quality

	Model C	Model D
Dependent variable	*Stiftung Warentest* quality scores	
Independent variables		
Average rating	0.17 (0.09 to 0.25)	0.13 (0.05 to 0.21)
Product age	–0.01 (–0.09 to 0.07)	0.01 (–0.07 to 0.08)
Average rating × Product age	–0.11 (–0.20 to –0.03)	–0.09 (–0.17 to –0.00[a])
Price		0.31 (0.23 to 0.39)

Note: [a]CI_{95}: –0.169 to –0.003.

4.1.4 Discussion

In sum, these findings provide support for our postulate that the convergence between average ratings and objective performance would decrease as a function of progress through the product life cycle (H1). Importantly, it should be noted that the relationship between rated and objective quality did not only decline over time, but also diminished to insignificance when considering products situated in later stages within their life cycles.

4.2 What is the Better Predictor of Sales Performance, Product Ratings or Objective Quality Scores?

4.2.1 Theoretical Considerations

Winer and Fader (2016) and Simonson (2016) likewise expressed their interest in a better understanding regarding the extent to which different pieces of available quality information (e.g., rated and objective quality) affect decision making and, consequently, sales performance. According to cue utilization theory (e.g., Olson 1978; Olson and Jacoby1972; Richardson, Dick, and Jain 1994; Zeithaml 1988) consumers draw quality inferences from two different types of informational cues; namely, intrinsic cues to quality—i.e., physical properties of a product which cannot be changed without altering the product itself—and extrinsic cues—i.e., attributes that are outside the product such as its price or brand name. Extant literature argues that, due to their high predictive value, intrinsic product attributes are often used as quality indicators (e.g., Olson and Jacoby 1972; Rigaux-Bricmont 1982; Szybillo and Jacoby 1974). Accordingly, prior research has shown that measures of objective quality such as CR scores are a good predictor of sales (e.g., Hardie et al. 1993; Narasimhan, Ghosh, and Mendez 1993; Trandel 1991; see also Simonsohn 2011). Building on the intrinsic-extrinsic dichotomy, online reviews can be classified as external cues to product quality (Khare, Labrecque, and Asare 2011); often argued to be less diagnostic indications of product quality than intrinsic cues (e.g., Olson and Jacoby 1972; Roggeveen, Grewal, and Gotlieb 2006). Although online reviews are highly accessible and easy to interpret, previous research has shown that consumers may not blindly follow them (Duan, Gu, and Whinston 2008; Liu 2006). For instance, Duan et al. (2008) found that, after controlling for several movie characteristics as well as rating volume, the average rating of a movie did not play an essential role in affecting box office revenues. In a similar vein, Liu (2006) argued that although rating valence might be highly influential in consumers' attitude formation, it is questionable whether this effect actually carries over to sales performance measures because the relationship between attitude and behavior is often weak (e.g., Ajzen and Fishbein 1977; Liska 1984, 1974). Therefore, we propose:

H2: *Objective quality scores are a better predictor of sales performance than average ratings*[9].

However, the concurrent effects of objective and rated quality on sales may not be independent of one another. More precisely, processing intrinsic product information often requires a lot of time and mental effort, and some product characteristics might be even too difficult to evaluate (Zeithaml 1988). In such cases, when persuasive extrinsic cues are available, consumers may rely on them more heavily because they are easier to access and to evaluate. For instance, strong brands may act as informational shortcut to infer product quality and, thereby, reduce evaluation costs (e.g., Häubl and Elrod 1994; Jo, Nakamoto, and Nelson 2003; Richardson et al. 1994). Assuming that high average ratings are similar persuasive as strong brands, consumers' might be less likely to engage in an extensive processing of product features to estimate the quality of a product if it has received mostly favorable online ratings. Hence, the effect of objective quality scores on sales performance may diminish as a function of increasing average ratings. On the other hand, previous research suggests that negative reviews might be more helpful and influential than positive reviews (e.g., Casaló et al. 2015; Chevalier and Mayzlin 2006; Ludwig et al. 2013; Yin, Bond, and Zhang 2014). Thus,

[9] Please note that consumers do not necessarily need to be informed about the actual objective quality scores provided by *Stiftung Warentest* to be able to evaluate a product's objective performance (e.g., Lichtenstein and Burton 1989). Instead, consumers processing product descriptions may be able to assess objective performance on their own, at least to some degree of accuracy. For instance, when comparing different vacuum cleaners consumers may correctly infer that a product with an input power of 800 watts should perform better than a product with only 600 watts.

consumers' may refrain from processing further product information required to assess the quality of a product if it is accompanied by an unfavorable average rating. Given these contradicting predictions, we investigated the moderating effect of average ratings on the relationship between objective quality scores and sales performance in an exploratory manner.

4.2.2 Data Description

We used Amazon's bestseller ranks as an inverse indicator of sales performance (see Chevalier and Mayzlin 2006, Sun 2012, and Floyd et al. 2014 for a similar approach). Hence, aside from the initially described restriction criteria, we had to further limit our database to product categories comprising three or more products that have been ranked within a common category on Amazon.de (e.g., 'Camera & Photo'). This resulted in a database of 1,220 products across 213 categories.

4.2.3 Results

We stepwise investigated the impacts of rated and objective quality on sales performance. Therefore, we first regressed the bestseller rank on the average rating and *Stiftung Warentest* score (Model E). Then, we added the interaction between rated and objective quality to the regression model (Model F). Finally, we incorporated selling prices as a covariate (Model G). In sum, across the three estimated models, we found that both objective and rated quality were positively related to sales performance (i.e., lower sales ranks; see Table 12).

Table 12. The Effects of Rated and Objective Product Quality on Sales Performance

	Model E	Model F	Model G
Dependent variable		*Amazon Bestseller Ranks*	
Independent variables			
Average rating	–0.15 (–0.20 to –0.09)	–0.14 (–0.20 to –0.09)	–0.15 (–0.21 to –0.10)
Stiftung Warentest quality scores	–0.22 (–0.28 to –0.17)	–0.22 (–0.27 to –0.16)	–0.28 (–0.33 to –0.22)
Average rating × *Stiftung Warentest* quality scores		0.08 (0.02 to 0.14)	0.07 (0.01 to 0.13)
Price			0.19 (0.14 to 0.25)

Note: All variables were z-standardized by subcategory before analysis.

Next, we inspected the relative amount of unique variance in sales performance explained by each predictor using squared semipartial correlations. Interestingly, objective quality scores uniquely explained the highest proportion of variance in bestseller ranks ($sr^2_{SW\ scores}$ = 6.60%; sr^2_{price} = 3.34%; $sr^2_{average\ rating}$ = 2.22%: Model G), such that the effect of objective quality scores on sales performance (b = –0.28, CI_{95}: –0.33 to –0.22) was significantly stronger than that of average ratings (b = –0.15, CI_{95}: –0.21 to –0.10; Δ = 0.13, t = 3.11, p < .01; Chin 2000).
In addition, we also found a significant average rating × *Stiftung Warentest* score interaction; indicating that the influence of each of the two pieces of quality information decreases with the favorability of the other.

4.2.4 Discussion

These findings support our postulate that objective quality scores are a better predictor of a product's sales performance than average ratings (H2). In fact, the information these scores represent is three times more influential in driving sales performance than average ratings. Please note, this finding does not necessarily imply that consumers actually consulted the quality judgments provided by *Stiftung Warentest*. As already mentioned, consumers might be at least partially able to infer a product's objective performance on their own using the provided product information. On the other hand, the finding that the influence of average ratings on purchase decisions appears to be rather small when compared to objective quality scores indicates that consumers do not blindly follow them and seem to form their own opinion about a product's quality. Finally, the detected interaction between rated and objective quality implies that the effects of the two pieces of quality information do not affect sales performance independent of one another. Pessimistically stated, this interaction suggests that purchase behavior becomes less dependent upon the objective performance of a product as its rated quality increases.

5 General Discussion

5.1 Contributions and Implications for Consumers

This research replicates and extends de Langhe et al.'s (2016a) seminal work on the limited convergence between online product ratings and measures of objective product performance which has been controversially discussed among several eminent marketing researchers. With this paper, we contribute to this discussion in three important ways.

First, by replicating the original findings using a different data source for objective quality information (i.e., product assessments published by *Stiftung Warentest*), we can rule out that the detected low convergence has to be merely ascribed to methodological defects in the product evaluations provided by CR (Simonson 2016). Given that consumers often consult average ratings in order to get an impression of a product's objective quality (de Langhe et al. 2016a), the confirmed low convergence suggests that the degree to which average ratings provide objective quality information is much lower than what consumers commonly believe. In fact, we found that even a product's price is a better indicator of its quality than its average rating. Hence, although online consumer ratings undoubtedly provide valuable information about other customers' subjective experiences with the reviewed product, consumers should interpret them carefully and should refrain from jumping to conclusions about a product's objective quality from average ratings.

Second, inspired by Winer and Fader (2016), we examined potential changes in the correlation between rated and objective quality over time. In line with extant literature suggesting that early adopters of a new product tend to be more knowledgeable than later buyers (e.g., Dee Dickerson and Gentry 1983; Hirschman 1980; Huh and Kim 2008; Park and Kim 2008), our findings reveal that the convergence between average ratings and objective quality scores deteriorates with product age. Hence, in particular, when trying to get an impression of the quality of older products, customers should use average ratings with utmost care.

Third, our investigation of the degree to which both rated and objective quality influence sales performance reveals that the information conveyed by objective quality scores is three times more influential in driving sales than average ratings; indicating that consumers rather attempt to get an own impression of a product's quality than blindly follow average ratings. This finding is particularly interesting in the light of Simonson's (2016) comment on de Langhe et al.'s (2016a) work arguing that consumers may not even care about objective assessments of product quality. Although we cannot make a statement about whether or not consumers care about the quality scores provided by *Stiftung Warentest* per se, our findings revealing a comparable large impact of these scores on sales performance indicate that consumers seem to at least care about what they convey. In addition, the documented interaction between rated and objective quality on sales performance indicates that the impacts of these two pieces of quality information are not independent of one another. More precisely, this interaction implies that the relationship between objective quality and sales ranks diminishes as average ratings increase. Thus, high customer ratings seem to be able to disguise a product's objective quality at least to some degree. As a consequence, consumers have to be careful not to be misled by enticingly high average ratings.

5.2 Future Research Directions

Although our work answers several questions raised after de Langhe et al.'s (2016a) study has been published, some limitations may offer opportunities for further research. First, just as the original study, our findings are based on an investigation of rated and objective quality scores for search products whose quality can be assessed at least with a certain degree of accuracy when inspecting performance-related product characteristics. However, extant literature argues that extrinsic cues to quality such as online consumer ratings might become more relevant in quality estimations when considering experience products (e.g., Zeithaml 1988) and services (e.g., Hartline and Jones 1996). Thus, future research should investigate whether objective quality scores are still a better predictor of sales performance than average ratings when considering products or services that can be typically evaluated only during or after consumption. In this context, however, developing an appropriate concept and operationalization of 'objective' quality might be challenging.

Second, although our results are in line with Winer and Fader's (2016) posit that the correlation between online ratings and objective performance decreases over a product's life cycle, a more conservative and explicit test of this postulate would investigate the relationship between objective quality scores and periodical average ratings using the dates on which online ratings have been posted while keeping the products under consideration constant.

Third, average ratings have often been argued to be subject to a variety of biases; including statistical, sampling, and evaluation issues (e.g., de Langhe et al. 2016a, 2016c, Hu et al. 2006). For instance, since usually not all customers who bought a product provide a review, sample sizes are often not sufficiently large from a statistical standpoint. Furthermore, the subset of customers who leave a review is typically not representative of the entire population of all product users (e.g., de Langhe et al. 2016c; Askalidis, Kim, and Malthouse 2017). In addition, more strictly speaking, online rating scores are commonly provided on an ordinal scale such that the assumption of equal distances between different rating scores might be violated and, thus, the calculation of average ratings is an invalid operation form a statistical perspective (e.g. Hair et al. 2010). Hence, future research may focus on other measures describing the distribution of rating scores (e.g., mode or median) and investigate the extent to which they could provide a valid indication of quality.

E Conclusion

Over the last 15 years, an abundance of academic research has demonstrated that online consumer reviews factor heavily into consumers' purchase decisions, making insights into how judgments are made on the basis of such evaluations provided by previous, typically unknown customers a worthy pursuit. Consequently, generating a better understanding regarding the pieces of information inherent in consumer reviews that shape decision making processes is highly relevant from both a practical and a theoretical perspective. Adding to the rich body of research on the impacts of different characteristics of the distribution of online rating scores, this doctoral thesis establishes consumers' use of the mode of rating distributions when making predictions about product quality and sheds light on the informational value of average ratings as an indication of a product's objective performance. The subsequent sections summarize the major findings of this thesis and discuss their contributions to existing literature. Thereafter, managerial implications for marketers and recommendations for consumers will be derived. This thesis concludes with an outline of limitations and future research avenues.

1 Summary of Findings

Extending previous research on the effects of different characteristics of rating distributions (i.e., average rating, rating volume, and dispersion), the first manuscript presented in this thesis provided a systematic examination of how consumers' interpretations of rating distributions illustrated as a bar chart are determined by the location of the mode; i.e., the rating score that has received the highest number of votes which is, thus, represented by the most salient bar within the graph. The results of a series of experimental studies in different product and service domains as well as secondary data from Amazon covering a variety of different product categories provided substantial empirical support for the tendency to make inferences about the quality of a product based on the location of the mode; a phenomenon labeled as the mode heuristic. More precisely, the studies documented that consumers use the mode as a heuristic basis for pre-purchase product evaluations in a variety of different contexts in such a way that products were judged more (less) favorably if the mode was located above (below) the average rating. Shedding light on this phenomenon at the process level, further analyses revealed that people's attention was directed toward more favorable (unfavorable) product ratings when the mode was located above (below) the average rating and that this shift in the allocation of visual attention prompted more favorable (unfavorable) product evaluations. Finally, quality inferences based on the mode heuristic were found to depend on the visual salience of the mode such that the effect of the mode's location on quality perceptions increased with its visual salience. Table 13 provides a summarizing overview of the key findings of Paper 1.

Table 13. Key Findings of Empirical Research Paper 1

Research Questions	Findings
How are consumers' inferences about the quality of a product affected by the location of a rating distribution's mode?	The mode of a rating distribution serves as a heuristic cue in consumers' product evaluations such that products are judged more (less) favorably if the mode is located above (below) the average rating.
What is the process underlying the relationship between the location of the mode and quality inferences?	The effect of the location of the mode on consumers' product evaluations is mediated by the allocation of visual attention to individual rating scores.
Which factors determine the relationship between the location of the mode and quality inferences?	An increasing visual salience of the mode strengthens the relationship between the location of the mode and product evaluations.

The second research paper in this thesis replicated and extended de Langhe et al.'s (2016a) seminal work on the limited convergence between online consumer ratings and more objective measures of product performance which has been controversially discussed in prior literature. The findings of this manuscript contribute to the discussion in three important ways. First, by replicating the original findings using a different data source for objective quality information (i.e., product assessments published by Stiftung Warentest), it can be ruled out that the originally reported low convergence has to be merely ascribed to methodological defects in the product evaluations provided by Consumer Reports. Second, potential changes in the correlation between rated and objective quality over time were examined. In line with the assumption that early adopters of a new product are able to judge a product's quality more accurately than later buyers, the findings revealed that the convergence between average product ratings and objective product performance decreases as a function of product age. Third, further analyses demonstrated that the information conveyed by objective quality scores was three times more influential in driving sales performance than average ratings. However, average ratings were found to moderate the relationship between objective quality and sales performance such that objective quality became less important when average ratings increased. Thus, not only that online ratings are an inaccurate indicator of product quality, they also seem to disguise a product's objective performance at least to some degree. Table 14 provides a summarizing overview of the key findings of Paper 2.

Table 14. Key Findings of Empirical Research Paper 2

Research Questions	Findings
Is the average product rating an adequate indicator of a product's 'objective' performance?	The convergence between average ratings and objective measures of product performance is remarkably weak; implying that average ratings are a rather imprecise predictor of a product's quality.
Does the convergence between rated and objective quality change over the product life cycle?	The convergence between average ratings and objective performance decreases over the product life cycle such that the relationship between rated and objective quality is weaker for older products that have been on the market for a longer period of time than for newer products.
What is the better predictor of sales performance, product ratings or objective quality scores?	Objective quality scores are a better predictor of sales performance than average ratings; they uniquely explain three times more variance in sales ranks than average ratings.

2 Theoretical Contributions

From a theoretical perspective, this doctoral thesis yields several important contributions to (1) prior literature on consumers' interpretations of different characteristics of the distribution of online ratings as well as to (2) research on the validity of online reviews as a quality indicator.

To date, extant research studying consumers' response to different rating distribution characteristics has primarily concentrated on the impacts of rating valence, rating volume, and rating dispersion (see Babić Rosario et al. 2016 for an overview), while the studies documented in Paper 1 are the first that place the mode of rating distributions under scrutiny. Hence, by establishing this previously neglected distribution characteristic as an important parameter in people's interpretations of rating distributions this work broadens our understanding regarding the effects of consumer ratings on pre-purchase product evaluations in online shopping environments.

Furthermore, Paper 1 complements the wide array of research on heuristics in judgment and decision making documenting the use of a variety of simple cues and rules of thumb as a simplification of evaluation processes (see Gilovich et al. 2002). In extension of this field of research, the present work demonstrates that the mode of a rating distribution can serve as such a heuristic basis when processing and interpreting graphical displays illustrating the distribution of online rating scores. Therefore, the documented mode heuristic aligns with other previously reported heuristics entailing that the weighting of available informational cues in judgment formation deviates from a normative mindset; e.g., the anchoring (Tversky and Kahneman 1974) and availability heuristic (Tversky and Kahneman 1973), as well as the peak-end rule (Kahneman et al. 1993).

In addition, though on a more minor note, this thesis also contributes to the literature on graphical perceptions. The empirical results reported in the first paper are consistent with the idea that people's interpretations of graphical formats are not only determined by the provided content per se, but also by the physical properties of the elements of a chart (e.g., Jarvenpaa 1990; Simkin and Hastie 1987; Spence 1990; Stone et al. 1997; Stone et al. 2003; Weber and Kirsner 1997). Specifically, people's interpretations of graphical illustrations of frequency distributions has been found to be crucially affected by the most perceptually salient element; i.e., the mode of a distribution. However, in contrast to extant studies in this field, this thesis provides explicit process evidence using an eye-tracking methodology (see Paper 1, Study 2). The findings revealed that, in fact, the attention paid to an object is dependent on its visual salience and that the allocation of visual attention across different elements of a graphical display determines the conclusions drawn from it. Thereby, this thesis responds to several calls for research to provide direct evidence for the process underlying salience effects in people's interpretations of graphical formats (e.g., Jarvenpaa 1990; Raghubir and Das 2010; Stone et al. 2003).

Furthermore, by replicating and extending previous findings on the weak convergence between average product ratings and more objective measures of product quality (de Langhe et al. 2016a) the second paper presented in this thesis adds to the recent debate about whether online reviews can adequately serve as an indicator of a products' true quality. In particular the finding that the accuracy of average ratings as a quality indicator diminishes as a function of a product's age generates a better understanding as to when and why the explanatory power of average ratings is questionable. In addition, in response to several calls for research (Simonson 2016; Winer and Fader 2016) the investigation of how both rated and objective quality concurrently affect a product's sales performance—discovering that the influence of each of the two pieces of quality information decreases with the favorability of the other—suggests that average ratings can distract consumers from the true quality of a product.

3 Implications for Business Practices

Online reviews have become a popular and powerful tool in driving customers' quality perceptions, purchase intentions, and sales. Empowered by the Internet, customers can easily share their opinions about the products and services they have experienced. On the one hand, from a company's perspective, review systems can be strategically used to increase sales and, thus, a firm's profitability and success. On the other hand, from a customer perspective, online reviews may enhance shopping experiences and help in making the 'right' choice.

The findings presented in this doctoral thesis generate a better understanding of how online reviews can influence consumers' decision making processes and draw attention to their weak predictive value as an indicator of a product's 'objective' performance. Hence, it is crucial to create awareness for both companies employing review systems on their websites as well as for consumers using such reviews in the belief that they allow them to make more informed choices. Therefore, the following subsections discuss managerial implications for marketers and recommendations for customers.

3.1 Managerial Implications for Marketers

The managerial implications of this thesis are multifaceted. Recognized as an instrument of strategic importance for practitioners (Jin et al. 2014; Packard and Berger 2017; Pan and Zhang 2011; Simonson and Rosen 2014) online consumer reviews provide important information for a variety of management activities including product development and quality assurance as well as customer acquisition and retention (Dellarocas 2003). In addition, consumer-generated product evaluations can be used as a basis to forecast sales performances (e.g., Dellarocas et al. 2004; Fan et al. 2017). In times when consumers place more trust in online reviews posted by unknown people than in any form of communication initiated by a company (Nielsen 2015), it is essential to develop a deeper understanding of how consumers respond to the abundance of product evaluations provided by previous customers that circulate in the modern marketplace.

By highlighting the relevance of the mode in consumers' product inferences from online rating distributions this thesis provides marketers a further useful measure that should be considered in analyzing and interpreting online review data. In other words, when using the mode in addition to the already established criteria (e.g., the number of ratings, average ratings, and rating dispersion), managers can better anticipate the consequences of different rating distributions of their offerings on consumers' pre-purchase product quality evaluations and, consequently, improve their estimations about product sales performances.

In addition, considering that consumers aim to make predictions about the quality of the available purchase options before buying, practitioners should also draw attention to the extent to which online shoppers incorporate the information inherent in online reviews into decision making. In this vein, the findings of Paper 2 reveal that objective quality was three times more influential in affecting sales ranks than average ratings indicating that favorable online reviews are not the major driver of a product's sale performance. This finding does not necessary imply that consumers do not care about online reviews per se but that review characteristics other than the average rating—such as, for instance, the mode of a rating distribution—are also relevant when making inferences about product quality. In addition, this result also suggests that although customers increasingly use online reviews as decision aids they do not follow them blindly. Hence, both manufactures and retailers should still keep the important role of a product's true quality in mind.

Finally, given the finding that the accuracy of average ratings as a predictor of product quality decreases over time, online marketers may contemplate reporting weighted instead of simple averages that might be more precise indicators of a products objective performance. This thesis suggests that one of the weighting factors could be the timing when ratings have been posted such that reviews written by earlier product adopters might be weighted more heavily.

3.2 Recommendations for Consumers

Although online reviews undoubtedly provide important and useful information about the experiences and opinions of other customers, in particular the results of the second paper presented in this thesis should draw consumers' attention to the fact that online reviews reflect less objective quality information than what one might believe. In order to illustrate the implications of this finding please image an online shopper who is trying to decide which smartphone to buy on an online retailers' website. He or she may look at the average ratings of the offered alternatives and might choose the product with the highest average rating. But, how likely does this customer choose the item that has likewise received the highest Stiftung Warentest score? In order to answer this question, it has been examined how often the product that has been awarded as a category test winner by Stiftung Warentest is also the product that has received the highest average rating on Amazon's website (among the products that have been tested by the consumer organization) using the data set from Paper 2 covering 224 tested product categories. This examination reveals that only in 30.8 percent (i.e., 69 out of 224 tested categories) the product that has received the best evaluation by Stiftung Warentest coincides with the one with the highest average rating in the corresponding category. Hence, consumers should avoid jumping to conclusions about a product's quality from average ratings.

In addition, the findings documented in Paper 2 also indicate that the convergence between average ratings and objective quality scores deteriorates with product age. Hence, in particular, when trying to get an impression of the quality of older products by means of online ratings, consumers might be well advised to be rather cautious not to be misled by average ratings. Finally, the documented interaction between rated and objective quality on sales performance indicates that the relationship between objective quality and sales ranks diminishes as average ratings increase. In other words, high consumer ratings seem to be able to disguise a product's objective quality at least to some degree. Thus, consumers should be careful not to be misled by enticingly high average ratings.

4 Limitations and Future Research Directions

Although this thesis provides valuable new insights into customers' interpretations of rating distributions as well as the use of online reviews as a quality indicator, it is not without limitations that offer promising opportunities for further research.

First and foremost, the studies reported in Paper 1 highlighting the important role of the mode of rating distributions might encourage researchers to gain deeper insights into this distribution characteristic. Although the studies reported in this thesis provide substantial empirical evidence of consumers' use of the mode as a heuristic cue when making predictions about the quality of a product, they, however, leave some aspects unanswered that could be addressed in future investigations. For instance, a worthwhile issue for future research might be to investigate whether the mode is associated with a semantic meaning and not only affects consumers' product evaluations due to its eye-catching physical salience but also because of the information it conveys. For instance, it is possible that the mode is interpreted as the majority's opinion or the most likely outcome. In a similar vein, future research could investigate whether the use of the mode heuristic occurs on a conscious or on a non-conscious level and whether the overpowering effects of the most perceptually salient element within a chart are robust among different graphical formats (e.g., pie charts, line graphs, and scatter diagrams), informational contents (e.g., financial and health risks), and judgmental contexts beyond product evaluations. In addition, future studies could explore conditions under which the mode heuristic is more or less likely to be used and, thereby, contribute to an explanation as to when and why different locations of the mode influence product perceptions and subsequent purchase behavior. In a similar vein, Gottschalk and Mafael (2017) recently identified five eWOM processing types (i.e., “The Efficients”, “The Meticulous”, “The Quality-Evaluators”, “The Cautious Critics”, and “The Swift Pessimist”) which differ in regard to how consumers use information cues when looking for decision aids in the online shopping environment. Hence, future research could examine whether the use of the mode heuristic varies across different eWOM processing types.

Second, against the background of the finding that online reviews heavily impact purchase decision making in online shopping environments (see Babić Rosario et al. 2016 for an overview; see also Paper 1) it is even more important to shed light on the question how accurate such reviews can reflect the ‘true’ quality of a product. Picking up one facet of online rating distributions, namely, average ratings, the second paper presented in this thesis highlights that this measure is a rather imprecise predictor of a product’s objective performance and, thereby, adds to the body of literature questioning the validity and predictive value of online reviews (e.g., Langhe et al. 2016a; Hu et al. 2006; Koh et al. 2010). A limitation of this study, however, is that the database used merely comprised search products whose quality can be assessed at least with a certain degree of accuracy when inspecting performance-related product features. Extant literature argues that extrinsic cues to quality such as online ratings might become more relevant in quality estimations when considering experience products and services (e.g., Hartline and Jones 1996; Zeithaml 1988). Thus, future research should investigate whether objective quality scores are still a better predictor of sales performance than average ratings when considering products or services that can be typically evaluated only during or after consumption. In addition, average ratings per se have often been argued to be subject to a variety of biases; including statistical, sampling, and evaluation problems (e.g., de Langhe et al. 2016a, 2016c; Hu et al. 2006). Hence, it would be interesting to examine the extent to which these different issues impair the accuracy and, thus, validity of online reviews as an indication of product quality. Furthermore, future research may focus on other measures describing the distribution of online ratings (e.g., mode or median) and investigate the degree to which they could provide a more valid and adequate measure of quality.

The huge amount of review data available online is not only invaluable for consumers and companies, but also for researchers. With the advent of eWOM about products and services, consumers' experiences, opinions, and evaluations have become easily accessible and measureable. Hence, the enormous amount of information that circulates around just waits to be processed and analyzed to shed light on a variety of still unanswered managerial and consumer-relevant research questions.

References

Ajzen, Icek, and Martin Fishbein (1977), "Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research," *Psychological Bulletin,* 84 (5), 888–918.

Ajzen, Icek and Martin Fishbein (1980), *Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs*, NJ: Prentice Hall.

Anderson, Eugene W. (1998), "Customer Satisfaction and Word of Mouth," *Journal of Service Research,* 1 (August), 5–17.

Amazon (2017), "Annual Report 2016," (accessed October 30, 2017), [available at http://annualreports.com/HostedData/AnnualReports/PDF/NASDAQ_AMZN_2016.pdf].

Aribarg, Anocha, Rik Pieters, and Michel Wedel (2010), "Raising the BAR: Bias Adjustment of Recognition Tests in Advertising," *Journal of Marketing Research,* 47 (June), 387–400.

Askalidis, Georgios, Su Jung Kim, and Edward C. Malthouse (2017), "Understanding and Overcoming Biases in Online Review Systems," *Decision Support Systems*, 97 (May), 23–30.

Atalay, A. Selin, H. Onur Bodur, and Dina Rasolofoarison (2012), "Shining in the Center: Central Gaze Cascade Effect on Product Choice," *Journal of Consumer Research,* 39 (December), 848–866.

Babić Rosario, Ana, B., Francesca Sotgiu, Kristine De Valck, and Tammo H. A. Bijmolt (2016), "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research*, 53 (June), 297–318.

Bakos, Yannis, and Chrysanthos Dellarocas (2011), "Cooperation without Enforcement? A Comparative Analysis of Litigation and Online Reputation as Quality Assurance Mechanisms," *Management Science,* 57 (November), 1944–1962.

Banerjee, Shankhadeep, Samadrita Bhattacharyya, and Indranil Bose (2017), "Whose Online Reviews to Trust? Understanding Reviewer Trustworthiness and its Impact on Business," *Decision Support Systems*, 96 (April), 17–26.

Bao, Tong, and Tung-lung Steven Chang (2014), "Why Amazon Uses Both the New York Times Best Seller List and Customer Reviews: An Empirical Study of Multiplier Effects on Product Sales from Multiple Earned Media," *Decision Support Systems,* 67 (November), 1–8.

Berger, Jonah, Alan T. Sorensen, and Scott J. Rasmussen (2010), "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales," *Marketing Science*, 29 (September/October), 815–827.

Bertini, Marco, Elie Ofek, and Dan Ariely (2009), "The Impact of Add-On Features on Consumer Product Evaluations," *Journal of Consumer Research*, 36 (June), 17–28.

Brynjolfsson, Erik, Yu Hu, and Michael D. Smith (2003), "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49 (November), 1580–1596.

Casaló, Luis V., Carlos Flavián, Miguel Guinalíu, and Yuksel Ekinci (2015), "Avoiding the Dark Side of Positive Online Consumer Reviews: Enhancing Reviews' Usefulness for High Risk-Averse Travelers," *Journal of Business Research*, 68 (September), 1829–1835.

Chaiken, Shelly (1980), "Heuristic versus Systematic Information Processing and the Use of Source versus Message Cues in Persuasion," *Journal of Personality and Social Psychology*, 39 (November), 752–766.

Chatterjee, Patrali (2001), "Online Reviews—Do Consumers Use Them?," in *Advances in Consumer Research*, ed. M. C. Gilly and J. Myers-Levy, Provo, UT: Association for Consumer Research, 129–134.

Chandon, Pierre, J. Wesley Hutchninson, Eric T. Bradlow, and Scott H. Young (2009), "Does In-Store Marketing Work? Effects of the Number and Position of Shelf Facings on Brand Attention and Evaluation at the Point of Purchase," *Journal of Marketing*, 73 (November), 1–17.

Chen, Pei-Yu, Samita Dhanasobhon, and Michael D. Smith (2008), "All Reviews are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.Com," (May), [available at: https://ssrn.com/abstract=918083].

Chen, Yubo, Yong Liu, and Jurui Zhang (2011), "When Do Third-Party Product Reviews Affect Firm Value and What Can Firms Do? The Case of Media Critics and Professional Movie Reviews," *Journal of Marketing*, 75 (September), 116–134.

Chen, Yubo, Qi Wang, and Jinhong Xie (2011), "Online Social Interactions: A Natural Experiment on Word of Mouth versus Observational Learning, "*Journal of Marketing Research*, 48 (April), 238–254.

Cheng, Yi-Hsiu, and Hui-Yi Ho (2015), "Social Influence's Impact on Reader Perceptions of Online Reviews," *Journal of Business Research*, 68 (April), 883–887.

Cheung, Christy MK, and Dimple R. Thadani (2012), “The Impact of Electronic Word-of-Mouth Communication: A Literature Analysis and Integrative Model,” *Decision Support Systems*, 54 (December), 461–470.

Chevalier, Judith A., and Dina Mayzlin (2006), “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research*, 43 (August), 345–359.

Chin, Wynne W. (2000), “Frequently Asked Questions—Partial Least Squares & PLS-Graph,” (accessed December 10, 2017), [available at http://disc-nt.cba.uh.edu/chin/plsfaq/plsfaq.htm].

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), “The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets,” *Marketing Science*, 29 (September/October), 944–957.

Clemons, Eric K., Goudon Gordon Gao, and Lorin M. Hitt (2006), “When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry,” *Journal of Management Information Systems*, 23 (December), 149–171.

Cleveland, William S., and Robert McGill (1984) “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,” *Journal of the American Statistical Association*, 79 (September), 531–554.

Corbetta, Maurizio, Erbil Akbudak, Thomas E. Conturo, Abraham Z. Snyder, John M. Ollinger, Heather A. Drury, Martin R. Linenweber, Steven E. Peterson, Marcus E. Raichle, David C. Van Essen, and Gordon L. Shulman (1998), “A Common Network of Functional Areas for Attention and Eye Movements,” *Neuron*, 21 (October), 761–773.

Dabholkar, Pratibha A. (1996), “Consumer Evaluations of New Technology-Based Self-Service Options: An Investigation of Alternative Models of Service Quality,” *International Journal of Research in Marketing*, 13 (February), 29–51.

Dawar, Niraj, and Philip Parker (1994), “Marketing Universals: Consumers' Use of Brand Name, Price, Physical Appearance, and Retailer Reputation as Signals of Product Quality,” *Journal of Marketing*, 58 (April), 81–95.

De Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016a), “Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings,” *Journal of Consumer Research*, 42 (April), 817–833.

De Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016b), “Star Wars: Response to Simonson, Winer/Fader, and Kozinets,” *Journal of Consumer Research*, 42 (April), 850–857.

De Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016c), “High Online User Ratings Don’t Actually Mean You’re Getting a Quality Product,” *Harvard Business Review*, (July), 2–6.

Dellarocas, Chrysanthos (2003), “The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms,” *Management Science*, 49 (October), 1407–1424.

Dellarocas, Chrysanthos (2006), “Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms,” *Management Science*, 52 (October), 1577–1593.

Dellarocas, Chrysanthos, Neveen Awad, and Michael Zhang (2004), “Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning,” *Proceedings of the Twenty-Fifth International Conference on Information Systems*, 379–386.

Dellarocas, Chrysanthos, and Ritu Narayan (2006), “A Statistical Measure of a Population’s Propensity to Engage in Post-Purchase Online Word-of-Mouth,” *Statistical Science*, 21 (May), 277–285.

Dellarocas, Chrysanthos, Xiaoquan M. Zhang, and Neveen Awad (2007), “Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures,” *Journal of Interactive Marketing*, 21 (Fall), 23–45.

Deng, Xiaoyan, Barbara E. Kahn, H. Rao Unnava, and Hyojin Lee (2016), “A ‘Wide’ Variety: Effects of Horizontal Versus Vertical Display on Assortment Processing, Perceived Variety, and Choice,” *Journal of Marketing Research*, 53 (October), 682–698.

Dee Dickerson, Mary, and James W. Gentry (1983), “Characteristics of Adopters and Non-Adopters of Home Computers,” *Journal of Consumer Research*, 10 (September), 225–235.

Dodds, William B., Kent B. Monroe, and Dhruv Grewal (1991), “Effects of Price, Brand, and Store Information on Buyers’ Product Evaluations,” *Journal of Marketing Research*, 28 (August), 307–319.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), “The Dynamics of Online Word-of-Mouth and Product Sales—An Empirical Investigation of the Movie Industry,” *Journal of Retailing*, 84 (June), 233–242.

Fan, Zhi-Ping, Yu-Jie Che, and Zhen-Yu Chen (2017), “Product Sales Forecasting Using Online Reviews and Historical Sales Data: A Method Combining the Bass Model and Sentiment Analysis,” *Journal of Business Research*, 74 (May), 90-100.

Filieri, Raffaele (2015), “What Makes Online Reviews Helpful? A Diagnosticity-Adoption Framework to Explain Informational and Normative Influences in e-WOM,” *Journal of Business Research*, 68 (June), 1261–1270.

Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Y. Cho, and Traci Freling (2014), "How Online Product Reviews Affect Retail Sales: A Meta-Analysis," *Journal of Retailing*, 90 (June), 217–232.

Fredrickson, Barbara L., and Daniel Kahneman (1993), "Duration Neglect in Retrospective Evaluations of Affective Episodes," *Journal of Personality and Social Psychology*, 65 (July), 45–55.

Gerstner, Eitan (1985), "Do Higher Prices Signal Higher Quality?," *Journal of Marketing Research*, 22 (May), 209–215.

Gigerenzer, Gerd, and Wolfgang Gaissmaier (2011), "Heuristic Decision Making," *Annual Review of Psychology*, 62 (January), 451–482.

Gilovich, Thomas, Dale Griffin, and Daniel Kahneman (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, UK: University Press.

Godes, David, and Dana Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (November), 545–560.

Godes, David, and José C. Silva (2012), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, 31 (May/June), 448–473.

Gottschalk, Sabrina A., and Alexander Mafael (2017), "Cutting Through the Online Review Jungle—Investigating Selective eWOM Processing," *Journal of Interactive Marketing,* 37 (February), 89–104.

Grewal, Dhruv, R. Krishnan, Julie Baker, and Norm Borin (1998), "The Effect of Store Name, Brand Name and Price Discounts on Consumers' Evaluations and Purchase Intentions," *Journal of Retailing, 74* (Fall), 331–352.

Hair, Joseph F., William C. Black, Barry J. Babin, and Rolph E. Anderson (2010), *Multivariate Data Analysis – A Global Perspective,* Upper Saddle River, NJ: Pearson.

Hardie, Bruce GS, Eric J. Johnson, and Peter S. Fader (1993), "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 12 (Fall), 378–394.

Hartline, Michael D., and Keith C. Jones (1996), "Employee Performance Cues in a Hotel Service Environment: Influence on Perceived Service Quality, Value, and Word-of-Mouth Intentions," *Journal of Business Research,* 35 (March), 207–215.

Hayes, Andrew F. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis*. New York: Guilford.

Häubl, Gerald, and Terry Elrod (1999), "The Impact of Congruity between Brand Name and Country of Production on Consumers' Product Quality Judgments," *International Journal of Research in Marketing,* 16 (September), 199–215.

He, Stephen X., and Samuel D. Bond (2015), “Why is the Crowd Divided? Attribution for Dispersion in Online Word of Mouth,” *Journal of Consumer Research*, 41 (April), 1509–1527.

Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler (2004), “Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?,” *Journal of Interactive Marketing*, 18 (Winter), 38–52.

Hennig-Thurau, Thorsten, and Gianfranco Walsh (2003), “Electronic Word-of-Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet." *International Journal of Electronic Commerce*, 8 (Winter), 51–74.

Hennig-Thurau, Thorsten, Caroline Wiertz, and Fabian Feldhaus (2015), “Does Twitter Matter? The Impact of Microblogging Word of Mouth on Consumers’ Adoption of New Movies,” *Journal of the Academy of Marketing Science*, 43 (May), 375–394.

Henricks, M. (1998), “Spread the Word,” *Entrepreneur*, 26 (February), 120–125.

Herr, Paul M., Frank R. Kardes, and John Kim (1991), “Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective,” *Journal of Consumer Research*, 17 (March), 454–462.

Hess, Ronald L., Shankar Ganesan, and Noreen M. Klein (2003), “Service Failure and Recovery: The Impact of Relationship Factors on Customer Satisfaction,” *Journal of the Academy of Marketing Science*, 31 (March), 127–145

Hirschman, Elizabeth C. (1980), “Innovativeness, Novelty Seeking, and Consumer Creativity,” *Journal of Consumer Research*, 7 (December), 283–295.

Ho-Dac, Nga N., Stephen J. Carson, and William L. Moore (2013), “The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter?” *Journal of Marketing*, 77 (November), 37–53.

Huang, Ni, Gordon Burtch, Yili Hong, and Evan Polman (2016), “Effects of Multiple Psychological Distances on Construal and Consumer Evaluation: A Field Study of Online Reviews,” *Journal of Consumer Psychology*, 26 (October), 474–482.

Hu, Nan, Ling Liu, Indranil Bose, and Jialie Shen (2010), “Does Sampling Influence Customers in Online Retailing of Digital Music?” *Information Systems and e-Business Management*, 8 (November), 357–377.

Hu, Nan, Ling Liu, and Jie Jennifer Zhang (2008), “Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects,” *Information Technology and Management*, 9 (September), 201–214.

Hu, Nan, Paul A. Pavlou, and Jennifer Zhang (2006), “Can Online Reviews Reveal Product’s True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication.” *Proceeding of the 7th ACM Conference on Electronic Commerce*, (June), 324–330.

Huh, Young Eun, and Sang-Hoon Kim (2008), “Do Early Adopters Upgrade Early? Role of Post-Adoption Behavior in the Purchase of Next-Generation Products,” *Journal of Business Research*, 61 (January), 40–46.

Hutchinson, J. Wesley, Joseph W. Alba, and Eric M. Eisenstein (2010), “Heuristics and Biases in Data-Based Decision Making: Effects of Experience, Training, and Graphical Data Displays,” *Journal of Marketing Research*, 47 (August), 627–642.

Ibrekk, Harald, and M. Granger Morgan (1987), “Graphical Communication of Uncertain Quantities to Nontechnical People,” *Risk Analysis*, 7 (December), 519–529.

Jacoby, Jacob, Jerry C. Olson, and Rafael A. Haddock (1971), “Price, Brand Name, and Product Composition Characteristics as Determinants of Perceived Quality,” *Journal of Applied Psychology*, 55 (December), 570–579.

Janiszewski, Chris (1998), “The Influence of Display Characteristics on Visual Exploratory Search Behavior,” *Journal of Consumer Research*, 25 (December), 290–301.

Jarvenpaa, Sirkka L. (1990), “Graphic Displays in Decision Making—The Visual Salience Effect,” *Journal of Behavioral Decision Making*, 3 (October/December), 247–262.

Jiménez, Fernando R., and Norma A. Mendoza (2013), “Too Popular to Ignore: The Influence of Online Reviews on Purchase Intentions of Search and Experience Products,” *Journal of Interactive Marketing*, 27 (August), 226–235.

Jin, Liyin, Bingyan Hu, and Yanqun He (2014), “The Recent versus the Out-Dated: An Experimental Examination of the Time-Variant Effects of Online Consumer Reviews,” *Journal of Retailing*, 90 (December), 552–566.

Jo, Myung-Soo, Kent Nakamoto, and James E. Nelson (2003), “The Shielding Effects of Brand Image Against Lower Quality Countries-of-Origin in Global Manufacturing,” *Journal of Business Research*, 56 (August), 637–646.

Kahneman, Daniel, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier (1993), “When More Pain is Preferred to Less: Adding a Better End,” *Psychological Science*, 4 (November), 401–405.

Kahneman, Daniel, and Amos Tversky (1979), “Prospect Theory: An Analysis of Decision Under Risk,” *Econometrica*, 47 (March), 263–291.

Khare, Adwait, Lauren I. Labrecque, and Anthony K. Asare (2011), "The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings," *Journal of Retailing*, 87 (March), 111–126.

Kim, Junyong, and Pranjal Gupta (2012), "Emotional Expressions in Online User Reviews: How they Influence Consumers' Product Evaluations," *Journal of Business Research*, 65 (July), 985–992.

King, Robert Allen, Pradeep Racherla, and Victoria D. Bush (2014), "What We Know and Don't Know about Online Word-of-Mouth: A Review and Synthesis of the Literature," *Journal of Interactive Marketing*, 28 (August), 167–183.

Koh, Noi Sian, Nan Hu, and Eric K. Clemons (2010), "Do Online Reviews Reflect a Product's True Perceived Quality? An Investigation of Online Movie Reviews across Cultures," *Electronic Commerce Research and Applications*, 9 (September/October), 374–385.

Kostyra, Daniel S., Jochen Reiner, Martin Natter, and Daniel Klapper (2016), "Decomposing the Effects of Online Customer Reviews on Brand, Price, and Product Attributes," *International Journal of Research in Marketing*, 33 (March), 11–26.

Kozinets, Robert V. (2016), "Amazonian Forests and Trees: Multiplicity and Objectivity in Studies of Online Consumer-Generated Ratings and Reviews, A commentary on de Langhe, Fernbach, and Lichtenstein," *Journal of Consumer Research*, 42 (April), 834–839.

Kronrod, Ann, and Shai Danziger (2013), ""Wii Will Rock You!" The Use and Effect of Figurative Language in Consumer Reviews of Hedonic and Utilitarian Consumption," *Journal of Consumer Research*, 40 (December), 726–739.

Kumar, V., JeeWon Brianna Choi, and Mallik Greene (2017), "Synergistic Effects of Social Media and Traditional Marketing on Brand Sales: Capturing the Time-Varying Effects," *Journal of the Academy of Marketing Science*, 45 (March), 268–288.

Kustov, Alexander A., and David Lee Robinson (1996), "Shared Neural Control of Attentional Shifts and Eye Movements," *Nature*, 384 (November): 74–77.

Lamberton, Cait, and Andrew T. Stephen (2016), "A Thematic Exploration of Digital, Social Media, and Mobile Marketing: Research Evolution from 2000 to 2015 and an Agenda for Future Inquiry," *Journal of Marketing*, 80 (November), 146–172.

Langan, Ryan, Ali Besharat, and Sajeev Varki (2017), "The Effect of Review Valence and Variance on Product Evaluations: An Examination of Intrinsic and Extrinsic Cues," *International Journal of Research in Marketing*, 34 (June), 414–429.

Lewandowsky, Stephan, and Ian Spence (1989), “The Perception of Statistical Graphs,” *Sociological Methods & Research*, 18 (November), 200–242.

Li, Xinxin, and Lorin M. Hitt (2008), “Self-Selection and Information Role of Online Product Reviews,” *Information Systems Research*, 19 (December), 456–474.

Lichtenstein, Donald R., and Scot Burton (1989), “The Relationship between Perceived and Objective Price-Quality,” *Journal of Marketing Research*, 26 (November), 429–443.

Lipkus, Isaac M. (2007), “Numeric, Verbal, and Visual Formats of Conveying Health Risks: Suggested Best Practices and Future Recommendations,” *Medical Decision Making*, 27 (September/October), 696–713.

Liska, Allen E. (1974), “Emergent Issues in the Attitude-Behavior Consistency Controversy,” *American Sociological Review*, 39 (April), 261–272.

Liska, Allen E. (1984), “A Critical Examination of the Causal Structure of the Fishbein/Ajzen Attitude-Behavior Model,” *Social Psychology Quarterly*, 47 (March), 61–74.

Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70 (July), 74–89.

Lu, Qi, Qiang Ye, and Rob Law (2014), “Moderating Effects of Product Heterogeneity between Online Word-of-Mouth and Hotel Sales,” *Journal of Electronic Commerce Research*, 15 (February), 1–12.

Luca, Michael (2011), “Reviews, Reputation, and Revenue: The Case of Yelp.com,” Working Paper No. 12-016, *Harvard Business School, NOM Unit*.

Luca, Michael, and Georgios Zervas (2016), “Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud,” *Management Science*, 62 (December), 3412–3427.

Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels, and Gerard Pfann (2013), “More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates,” *Journal of Marketing*, 77 (January), 87–103.

Lurie, Nicholas H., and Charlotte H. Mason (2007), “Visual Representation: Implications for Decision Making,” *Journal of Marketing*, 71 (January), 160–177.

MacKenzie, Scott B. (1986), “The Role of Attention in Mediating the Effect of Advertising on Attribute Importance,” *Journal of Consumer Research*, 13 (September), 174–195.

Mackie, Diane M. (1987), “Systematic and Nonsystematic Processing of Majority and Minority Persuasive Communications,” *Journal of Personality and Social Psychology*, 53 (July), 41–52.

Malbon, Justin (2013), "Taking Fake Online Consumer Reviews Seriously," *Journal of Consumer Policy*, 36 (June), 139–157.

Malhotra, Naresh K. (2010), *Marketing Research – An Applied Orientation*, Upper Saddle River, NJ: Pearson.

Mandel, Naomi, and Eric J. Johnson (2002), "When Web Pages Influence Choice: Effects of Visual Primes on Experts and Novices," *Journal of Consumer Research*, 29 (September), 235–245.

Mannan, Sabira K., Christopher Kennard, and Masud Husain (2009), "The Role of Visual Salience in Directing Eye Movements in Visual Object Agnosia," *Current Biology*, 19 (March), R247-R248.

Maoz, Eyal, and Alice M. Tybout (2002), "The Moderating Role of Involvement and Differentiation in the Evaluation of Brand Extensions," *Journal of Consumer Psychology*, 12 (January), 119–131.

Martin, Robin, and Miles Hewstone (2003), "Majority versus Minority Influence: When, not Whether, Source Status Instigates Heuristic or Systematic Processing," *European Journal of Social Psychology*, 33 (November), 313–330.

Martin, Robin, Miles Hewstone, and Pearl Y. Martin (2007), "Systematic and Heuristic Processing of Majority and Minority-Endorsed Messages: The effects of Varying Outcome Relevance and Levels of Orientation on Attitude and Message Processing," *Personality and Social Psychology Bulletin*, 33 (July), 43–56.

Matakos, Antonis, and Panayiotis Tsaparas (2016), "Temporal Mechanisms of Polarization in Online Reviews," *Advances in Social Networks Analysis and Mining (ASONAM)*, IEEE/ACM International Conference, (August) 529–532.

Mathwick, Charla, and Jill Mosteller (2017), "Online Reviewer Engagement: A Typology based on Reviewer Motivations," *Journal of Service Research*, 20 (May), 204–218.

Mayzlin, Dina (2006), "Promotional Chat on the Internet," *Marketing Science*, 25 (Spring), 155–163.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014), "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review* 104 (August), 2421–2455.

Meißner, Martin, Andres Musalem, and Joel Huber (2016), "Eye Tracking Reveals Processes that Enable Conjoint Choices to Become Increasingly Efficient with Practice," *Journal of Marketing Research*, 53 (February), 1–17.

Milosavljevic, Milica, Vidhya Navalpakkam, Christof Koch, and Antonio Rangel (2012), "Relative Visual Saliency Differences Induce Sizable Bias in Consumer Choice," *Journal of Consumer Psychology*, 22 (January), 67–74.

Moe, Wendy W., and David A. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31 (May/June), 372–386.

Moe, Wendy W., and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (June), 444–456.

Monroe, Kent B. (1973), "Buyers' Subjective Perceptions of Price," *Journal of Marketing Research*, 10 (February), 70–80.

Moon, Sangkil, Paul K. Bergey, and Dawn Iacobucci (2010), "Dynamic Effects among Movie Ratings, Movie Revenues, and Viewer Satisfaction," *Journal of Marketing*, 74 (January), 108–121.

Moon, Sangkil, and Wagner A. Kamakura (2017), "A Picture is Worth a Thousand Words: Translating Product Reviews into a Product Positioning Map," *International Journal of Research in Marketing*, 34 (March), 265–285.

Moore, David S., George P. McCabe, and Bruce A. Craig (2012), *Introduction to the Practice of Statistics*, New York, NY: Freeman.

Moore, Sarah G. (2015), "Attitude Predictability and Helpfulness in Online Reviews: The Role of Explained Actions and Reactions," *Journal of Consumer Research*, 42 (June), 30–44.

MSI (2014), "Marketing Science Institute 2014–2016 Research Priorities" (accessed April 17, 2018), [available at http://www.msi.org/uploads/files/MSI_RP14-16.pdf].

MSI (2016), "Marketing Science Institute 2016–2018 Research Priorities" (accessed April 17, 2018), [available at https://www.msi.org/uploads/articles/MSI_RP16-18.pdf].

Muchnik, Lev, Sinan Aral, and Sean J. Taylor (2013), "Social Influence Bias: A Randomized Experiment," *Science*, 341 (August), 647–651.

Mudambi, Susan M., and David Schuff (2010), "Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly*, 34 (March), 185–200.

Narasimhan, Ram, Soumen Ghosh, and David Mendez (1993), "A Dynamic Model of Product Quality and Pricing Decisions on Sales Response," *Decision Sciences*, 24 (September), 893–908.

Naylor, Rebecca Walker, Cait Poynor Lamberton, and David A. Norton (2011), "Seeing Ourselves in Others: Reviewer Ambiguity, Egocentric Anchoring, and Persuasion," *Journal of Marketing Research*, 48 (June), 617–631.

Nielsen (2015), "Global Trust in Advertising: Winning Strategies for an Evolving Media Landscape," (accessed October 30, 2017), [available at http://www.nielsen.com/content/dam/nielsenglobal/apac/docs/reports/2015/nielsen-global-trust-in-advertising-report-september-2015.pdf].

Nowlis, Stephen M. (1995), "The Effect of Time Pressure on the Choice between Brands that Differ in Quality, Price, and Product Features," *Marketing Letters*, 6 (October), 287–295.

NRF (2017), "2017 Top 50 E-Retailers Chart" (accessed April 17, 2018), [available at https://nrf.com/2017-top-50-e-retailers-chart].

Olson, Jerry C. (1978), "Inferential Beliefs in the Cue Utilization Process," in *Advances in Consumer Research*, Vol. 5, ed. H. Keith Hunt, Ann Arbor, MI: Association for Consumer Research, 706-713.

Olson, Jerry C., and Jacob Jacoby (1972), "Cue Utilization in the Quality Perception Process," in *Advances in Consumer Research*, ed. M. Venkatesan, Chicago, IL: Association for Consumer Research, 167–179.

Packard, Grant, and Jonah Berger (2017), "How Language Shapes Word of Mouth's Impact," *Journal of Marketing Research*, 54 (August), 572–588.

Pan, Yue, and Jason Q. Zhang (2011), "Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews," *Journal of Retailing*, 87 (December), 598–612.

Park, Do-Hyung, and Sara Kim (2008), "The Effects of Consumer Knowledge on Message Processing of Electronic Word-of-Mouth via Online Consumer Reviews," *Electronic Commerce Research and Applications*, 7 (Winter), 399–410.

Park, Se-Bum, and Do-Hyung Park (2013), "The Effect of Low- versus High-Variance in Product Reviews on Product Evaluation," *Psychology & Marketing*, 30 (June), 543–554.

Parkhurst, Derrick, Klinton Law, and Ernst Niebur (2002), "Modeling the Role of Salience in the Allocation of Overt Visual Attention," *Vision Research*, 42 (January), 107–123.

Pieters, Rik, Edward Rosbergen, and Michel Wedel (1999), "Visual Attention to Repeated Print Advertising: A Test of Scanpath Theory," *Journal of Marketing Research*, 36 (November), 424–438.

Pincus, Steven, and L. K. Waters (1977), "Informational Social Influence and Product Quality Judgments," *Journal of Applied Psychology*, 62 (October), 615–619.

Pinker, Steven (1990), "A Theory of Graph Comprehension," in *Artificial Intelligence and the Future of Testing*, R. Freedle, ed., Hillsdale, NJ: Erlbaum, 73–126.

Purnawirawan, Nathalia, Martin Eisend, Patrick De Pelsmacker, and Nathalie Dens, (2015), "A Meta-Analytic Investigation of the Role of Valence in Online Reviews," *Journal of Interactive Marketing*, 31 (August), 17–27.

Purnawirawan, Nathalia, Patrick De Pelsmacker, and Nathalie Dens (2012), "Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions," *Journal of Interactive Marketing*, 26 (November), 244–255.

Raghubir, Priya, and Eric A. Greenleaf (2006), "Ratios in Proportion: What Should the Shape of the Package Be?" *Journal of Marketing*, 70 (April), 95–107.

Raghubir, Priya, and Aradhna Krishna (1999), "Vital Dimensions in Volume Perception: Can the Eye Fool the Stomach?," *Journal of Marketing Research*, 36 (August), 313–326.

Raghubir, Priya, and Sanjiv R. Das (2010), "The Long and Short of it: Why are Stocks with Shorter Runs Preferred?," *Journal of Consumer Research*, 36 (April), 964–982.

Rao, Akshay R., and Kent B. Monroe (1989), "The Effect of Price, Brand Name, and Store Name on Buyers' Perceptions of Product Quality: An Integrative Review," *Journal of Marketing Research*, 26 (August), 351–357.

Ratneshwar, Srinivasan, Luk Warlop, David Glen Mick, and Gail Seeger (1997), "Benefit Salience and Consumers' Selective Attention to Product Features," *International Journal of Research in Marketing*, 14 (July), 245–259.

Reimer, Thomas, and Martin Benkenstein (2016), "When Good WOM Hurts and Bad WOM Gains: The Effect of Untrustworthy Online Reviews," *Journal of Business Research*, 69 (December), 5993–6001.

Reyes, Robert M., William C. Thompson, and Gordon H. Bower (1980), "Judgmental Biases Resulting from Differing Availabilities of Arguments," *Journal of Personality and Social Psychology*, 39 (July), 2–12.

Richardson, Paul S., Alan S. Dick, and Arun K. Jain (1994), "Extrinsic and Intrinsic Cue Effects on Perceptions of Store Brand Quality," *Journal of Marketing*, (October), 28–36.

Rigaux-Bricmont, Benny (1982), "Influences of Brand Name and Packaging on Perceived Quality," in *Advances in Consumer Research*, Vol. 9, ed. Andrew Mitchell, Ann Abor, MI: Association for Consumer Research, 472–477 [available at: http://acrwebsite.org/volumes/6050/volumes/v09/NA-09].

Roggeveen, Anne L., Dhruv Grewal, and Jerry Gotlieb (2006), "Does the Frame of a Comparative Ad Moderate the Effectiveness of Extrinsic Information Cues?," *Journal of Consumer Research*, 33 (June), 115–122.

Sanfey, Alan, and Reid Hastie (1998), "Does Evidence Presentation Format Affect Judgment? An Experimental Evaluation of Displays of Data for Judgments," *Psychological Science*, 9 (March), 99–103.

Schlosser, Ann E. (2011), "Can Including Pros and Cons Increase the Helpfulness and Persuasiveness of Online Reviews? The Interactive Effects of Ratings and Arguments," *Journal of Consumer Psychology*, 21 (July), 226–239.

Sen, Shahana, and Dawn Lerman (2007), "Why Are You Telling Me This? An Examination into Negative Consumer Reviews on the Web," *Journal of Interactive Marketing*, 21 (Fall), 76–94.

Shavitt, Sharon, and Russell H. Fazio (1991), "Effects of Attribute Salience on the Consistency between Attitudes and Behavior Predictions," *Personality and Social Psychology Bulletin*, 17 (October), 507–516.

Shoham, Meyrav, Sarit Moldovan, and Yael Steinhart (2017), "Positively Useless: Irrelevant Negative Information Enhances Positive Impressions," *Journal of Consumer Psychology*, 27 (April), 147–159.

Simkin, David, and Reid Hastie (1987), "An Information–Processing Analysis of Graph Perception," *Journal of the American Statistical Association*, 82 (398), 454–465.

Simonsohn, Uri (2011), "Lessons from an "Oops" at Consumer Reports: Consumers Follow Experts and Ignore Invalid Information," *Journal of Marketing Research*, 48 (February), 1–12.

Simonson, Itamar (2016), "Imperfect Progress: An Objective Quality Assessment of the Role of User Reviews in Consumer Decision Making, A Commentary on de Langhe, Fernbach, and Lichtenstein," *Journal of Consumer Research*, 42 (April), 840–845.

Simonson, Itamar, and Emanuel Rosen (2014). *Absolute Value: What Really Influences Customers in the Age of (Nearly) Perfect Information*. New York, NY: HarperCollins.

Singh, Jyoti P., Sena Irani, Nripendra P. Rana, Yogesh K. Dwivedi, Sunil Saumya, and Pradeep K. Roy (2017), "Predicting the "helpfulness" of online consumer reviews," *Journal of Business Research*, 70 (January), 346–355.

Spence, Ian (1990), "Visual Psychophysics of Simple Graphical Elements," *Journal of Experimental Psychology: Human Perception and Performance*, 16 (November), 683–692.

Spiegelhalter, David, Mike Pearson, and Ian Short (2011), "Visualizing Uncertainty about the Future," *Science*, 333 (September), 1393–1400.

Sridhar, Shrihari, and Raji Srinivasan (2012), "Social Influence Effects in Online Product Ratings," *Journal of Marketing*, 76 (September), 70–88.

Stone, Eric R., Winston R. Sieck, Benita E. Bull, J. Frank Yates, Stephanie C. Parks, and Carolyn J. Rusch (2003), "Foreground:Background Salience: Explaining the Effects of Graphical Displays on Risk Avoidance," *Organizational Behavior and Human Decision Processes*, 90 (May), 19–36.

Stone, Eric R., J. Frank Yates, and Andrew M. Parker (1997), "Effects of Numerical and Graphical Displays on Professed Risk–Taking Behavior," *Journal of Experimental Psychology: Applied*, 3 (December), 243–256.

Sun, Monic (2012), "How Does the Variance of Product Ratings Matter?" *Management Science*, 58 (April), 696–707.

Sun, Yan, Shu Li, and Nicolao Bonini (2010), "Attribute Salience in Graphical Representations Affects Evaluation," *Judgment and Decision Making*, 5 (June), 151–158.

Sundaram, Dinesh S., Kaushik Mitra, and Cynthia Webster (1998), "Word-of-Mouth Communications: A Motivational Analysis," in *Advances in Consumer Research*, Vol. 25, ed. Joseph W. Alba, J. Wesley Hutchinson, Provo, UT: Association for Consumer Research, 527–531.

Suri, Rajneesh, and Kent B. Monroe (2003), "The Effects of Time Constraints on Consumers' Judgments of Prices and Products," *Journal of Consumer Research*, 30 (June), 92–104.

Szybillo, George J., and Jacob Jacoby (1974), "Intrinsic Versus Extrinsic Cues as Determinants of Perceived Product Quality," *Journal of Applied Psychology*, 59 (February), 74–78.

Taylor, Valerie A., and William O. Bearden (2002), "The Effects of Price on Brand Extension Evaluations: The Moderating Role of Extension Similarity," *Journal of the Academy of Marketing Science*, 30 (March), 131–140.

Teixeira, Thales, Michel Wedel, and Rik Pieters (2012), "Emotion-Induced Engagement in Internet Video Advertisements," *Journal of Marketing Research*, 49 (April), 144–159.

Townsend, Claudia, and Barbara E. Kahn (2014), "The 'Visual Preference Heuristic': The Influence of Visual versus Verbal Depiction on Assortment Processing, Perceived Variety, and Choice Overload," *Journal of Consumer Research*, 40 (February), 993–1015.

Trandel, Gregory A. (1991), "The Bias Due to Omitting Quality When Estimating Automobile Demand," *The Review of Economics and Statistics*, 73 (August), 522–525.

TripAdvisor (2018), "About TripAdvisor" (accessed April 17, 2018), [available at https://tripadvisor.mediaroom.com/US-about-us].

Tsang, Alex S. L., and Gerard Prendergast (2009), “Is a “Star” Worth a Thousand Words? The Interplay between Product-Review Texts and Rating Valences,” *European Journal of Marketing*, 43 (November/December), 1269–1280.

Tversky, Amos, and Daniel Kahneman (1973), “Availability: A Heuristic for Judging Frequency and Probability,” *Cognitive Psychology*, 5 (September), 207–232.

Tversky, Amos, and Daniel Kahneman (1974), “Judgment under Uncertainy: Heuristics and Biases,” *Science*, 185 (September), 1124–1131.

Van Doorn, Jenny, Katherine N. Lemon, Vikas Mittal, Stephan Nass, Doreén Pick, Peter Pirner, and Peter C. Verhoef (2010), “Customer Engagement Behavior: Theoretical Foundations and Research Directions,” *Journal of Service Research*, 13 (August), 253–266.

Venkatraman, Vinod, Angelika Dimoka, Paul A. Pavlou, Khoi Vo, William Hampton, Bryan Bollinger, Hal E. Hershfield, Masakazu Ishihara, and Russell S. Winer (2015), “Predicting Advertising Success Beyond Traditional Measures: New Insights From Neurophysiological Methods and Market Response Modeling,” *Journal of Marketing Research*, 52 (August), 436–452.

Villarroel Ordenes, Francisco, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (2017), “Unveiling what is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media,” *Journal of Consumer Research*, 43 (April), 875–894.

Visschers, Vivianne H. M., Ree M. Meertens, Wim W. F. Passchier, and Nanne N. K. De Vries (2009), “Probability Information in Risk Communication: A Review of the Research Literature,” *Risk Analysis*, 29 (February), 267–287.

Von Hippel, Paul T. (2005), “Mean, Median, and Skew: Correcting a Textbook Rule,” *Journal of Statistics Education*, 13 (July), [available at http://ww2.amstat.org/publications/jse/v13n2/ vonhippel.html].

Wang, Feng, Xuefeng Liu, and Eric Er Fang (2015), “User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects,” *Journal of Retailing*, 91 (September), 372–389.

Weber, Elke, and Britt Kirsner (1997), “Reasons for Rank-Dependent Utility Evaluation,” *Journal of Risk and Uncertainty*, 14 (January), 41–61.

West, Patricia M., and Susan M. Broniarczyk (1998), “Integrating Multiple Opinions: The Role of Aspiration Level on Consumer Response to Critic Consensus,” *Journal of Consumer Research*, 25 (June), 38–51.

Wilson, Andrew E., Michael D. Giebelhausen, and Michael K. Brady (2017), "Negative Word of Mouth can be a Positive for Consumers Connected to the Brand," *Journal of the Academy of Marketing Science*, 45 (July), 534–547.

Winer, Russell S. and Peter S. Fader (2016), "Objective vs. Online Ratings: Are Low Correlations Unexpected and Does It Matter? A Commentary on de Langhe, Fernbach, and Lichtenstein," *Journal of Consumer Research*, 42 (April), 846–849.

Yang, Liu, Olivier Toubia, and Martijn G. De Jong (2015), "A Bounded Rationality Model of Information Search and Choice in Preference Measurement," *Journal of Marketing Research*, 52 (April), 166–183.

Ye, Qiang, Rob Law, Bin Gu, and Wei Chen (2011), "The Influence of User-Generated Content on Traveler Behavior: An Empirical Investigation on the Effects of e-Word-of-Mouth to Hotel Online Bookings," *Computers in Human behavior*, 27 (March), 634–639.

Yin, Dezhi, Samuel Bond, and Han Zhang (2014), "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews," *MIS Quarterly*, 38 (June), 539–560.

You, Ya, Gautham G. Vadakkepatt, and Amit M. Joshi (2015), "A Meta-Analysis of Electronic Word-of-Mouth Elasticity," *Journal of Marketing*, 79 (2015), 19–39.

Zeithaml, Valarie A. (1988), "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence," *Journal of Marketing*, 52 (July), 2–22.

Zhang, Xiaoquan Michael (2006), "Tapping into the Pulse of the Market: Essays on Marketing Implications of Information Flows," doctoral dissertation, MIT, Cambridge, MA.

Zhang, Xiaoquan, and Chrysanthos Dellarocas (2006), "The Lord of the Ratings: Is a Movie's Fate is Influenced by Reviews?," *Proceedings of the Twenty-Seventh International Conference on Information Systems*, 1959–1978.

Zhang, Jie, Michel Wedel, and Rik Pieters (2009), "Sales Effects of Attention to Feature Advertisements: a Bayesian Mediation Analysis," *Journal of Marketing Research*, 46 (October), 669–681.

Zhao, Yi, Sha Yang, Vishal Narayan, and Ying Zhao (2013), "Modeling Consumer Learning from Online Product Reviews," *Marketing Science*, 32 (January/February), 153–169.

Zhu, Feng, and Xiaoquan Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74 (March), 133–148.

Appendix

Product Categories and Pearson Correlations between Average Ratings from Amazon.de and *Stiftung Warentest* Quality Scores

Category	Issue	Correlation
All-in-one inkjet printers (✓)		
*with fax**	4/15	-0.711
	4/17	0.041
with fax and automatic document feeder	4/16	-0.392
*without fax**	4/16	-0.667
	4/17	0.756
All-in-one laser printers (✓)		
*black and white laser (with fax)**	4/14	-0.226
	9/17	0.966
*black and white laser (without fax)**	10/14	0.081
	9/16	0.881
color laser (with fax)	9/17	0.345
color laser (without fax)	9/16	-0.662
Baby monitors (✓)		
audio models	5/15	0.505
video models	5/15	0.904
Blood pressure monitors (✓)		
wrist models	5/16	0.656
arm models	5/16	0.226
Bluetooth headphones		
over-ear	6/17	-0.388
on-ear	6/17	0.344
Bluetooth receivers	8/17	0.263
Bluetooth speakers*	6/15	0.206
	4/16	0.255
	9/17	-0.287
Blu-ray players	1/16	-0.785
Built-in refrigerators (*small size models)* (✓)	5/17	0.500
Camcorders (✓)	6/17	0.718
*action camcorders**	8/14	-0.262
	7/16	0.619
HD camcorders with hard drive	10/14	0.791
Cameras (✓)		
*small models with large zoom**	9/14	-0.016
	9/15	0.230
*large models with extra large zoom**	9/14	-[a]
	9/15	-0.058
Simple compact models		
super zoom	9/16	0.410
standard zoom	9/16	0.475
Premium compact models	12/14	-0.471
standard zoom	9/16	-0.347
compact models	12/15	0.476
with zoom lens	1/14	0.160
robust cameras	7/14	-0.817
Camera travel lenses (✓)		
for Canon	3/16	-0.829
for Nikon	3/16	-0.213

Product Categories and Pearson Correlations between Average Ratings from Amazon.de and *Stiftung Warentest* Quality Scores *(continued)*

Category	Issue	Correlation
Coffee makers (✓)		
coffee makers cup brew and dispense models	11/15	-0.059
espresso machines	12/14	0.643
espresso makers	12/16	0.396
Coffeemaker combos	12/16	0.081
with automatic milk frother	12/17	-0.099
Computer monitors (✓)		
widescreen (16:9 ratio)	5/15	0.333
ultrawide (21:9 ratio)	5/15	0.786
Computer tablets (✓)		
large computer tablets	12/14	0.719
small computer tablets	12/14	0.660
*computer tablets with keyboard**	8/16	-0.085
	1/17	0.546
	7/17	0.972
*6.8 – 8.4 inch models**	6/14	-0.381
	1/15	0.673
	12/15	0.532
	1/16	0.229
6.9 – 8 inch models	8/16	0.920
7 – 8 inch models	12/16	0.945
8.7 – 9.8 inch models	6/14	0.614
8.7 – 10.9 inch models	12/15	0.933
8.9 – 10 inch models	1/15	0.081
8.9 – 10.9 inch models	7/15	0.409
9.4 – 10 inch models	7/17	0.588
9.6 – 10.1 inch models	12/16	0.870
10 inch models	8/16	1.000
Conventional dishwashers *(60 cm)* (✓)*	5/15	-0.506
	6/16	-0.121
Cooktops *(ceramic cooktops)*	2/15	-0.028
Cordless hedge trimmers		
hedge trimmers	8/17	0.056
pole hedge trimmers	8/17	0.168
Cordless phones (✓)		
simple models	1/14	-0.871
comfort models	1/14	0.712
with base station	9/15	0.612
without base station	9/15	-0.533
with touchscreen	1/14	-0.747
Digital radios		
DAB+	7/15	0.664
DAB+ and Internet radio	7/15	-0.738
Digital scales (✓)	1/14	0.232
Drones *(with GPS)*	12/17	-0.111
Drilling machines (✓)		
Cordless drills/drivers		
light use cordless use drills/drivers	3/15	-0.720
cordless impact drills/drivers	3/15	0.995
impact drills/drivers	3/15	0.134
rotary hammer drills	3/15	-0.758
DVB-T2 HD receivers *(with decoder)*	2/17	0.602
DVB-T2 outdoor antennas	3/17	0.299
E-book readers *(Black and White)* (✓)	2/14	0.083
Electric grills (✓)		
contact grills	6/15	-0.251
electric griddles	6/15	0.158

Product Categories and Pearson Correlations between Average Ratings from Amazon.de and *Stiftung Warentest* Quality Scores *(continued)*

Category	**Issue**	**Correlation**	**Category**	**Issue**	**Correlation**
Electric toothbrushes*	3/16	0.586	High pressure washers	4/14	-0.175
	1/17	0.685	Indoor antennas for DVB-T2	2/17	0.780
	11/17	-0.533	Inkjet printers (✓)	4/15	-0.348
Electric toothbrushes for kids	1/15	0.502	Jig saws		
Electric mixers (✓)			*corded barrel-grip*	3/16	0.150
up to 1000 watts	10/16	0.873	*cordless top handle*	3/16	0.958
more than 1000 watts	10/16	0.686	Kitchen machines (*with heating mode)*	12/15	-0.546
Electric razors (✓)	5/17	0.389			
Exercise bikes *(upright ergometer)*	1/15	-0.463	Laptop computers and ultrabooks		
			ultrabook PCs with Windows	4/17	-0.674
Fitness trackers (✓)	12/17	-0.581	*convertibles with Windows*	4/17	-0.278
with heart rate monitor	1/16	0.947	Laser printers (✓)		
without heart rate monitor	1/16	0.893	*black and white laser*	10/14	-0.129
Fitness watches	12/17	0.258	*printers**	9/17	0.937
Freezers			*color laser printers*	9/15	-0.111
small size freezers	8/15	0.762	Lawn mowers		
large size freezers	8/15	-0.961	*cordless lawn mower*	4/17	-0.616
freestanding Freezers (large size)	8/17	0.721	*corded lawn mower*	4/14	0.567
			Microwaves (✓)		
GPS navigators (✓)			*with grill and oven*	8/16	0.085
*5 inch screen size**	2/14	0.742	*with grill*	8/16	0.061
	2/15	0.540	Mini Hi-Fi systems	12/15	-0.138
*6 – 7 inch screen size**	2/14	-0.716	Mini PCs	10/16	-0.917
	2/15	0.035	Network receivers (*AV receivers*)	8/17	0.865
Hair dryers (*ionic*) (✓)	1/15	0.470			
Headphones (✓)			PC sticks	10/16	0.985
in-ear headphones	8/15	-0.305	Personal clouds		
wired headphones	5/14	0.091	*single drive*	2/16	0.250
Sports headphones			*dual drive*	2/16	0.139
wired sports headphones	8/16	0.362			
with bluetooth	8/16	0.151			

Product Categories and Pearson Correlations between Average Ratings from Amazon.de and *Stiftung Warentest* Quality Scores *(continued)*

Category	Issue	Correlation	Category	Issue	Correlation
Power banks			Smartphones* *(continued)*	11/16	0.325
2200 – 3000 mAh capacity	6/16	-0.448		5/17	0.426
5200 – 6000 mAh capacity	6/16	-0.214		11/17	0.517
Projectors					
long throw	6/16	0.263	Smartphones for seniors (*simple models*)	1/17	0.525
short throw	6/16	0.196			
full HD	6/14	-0.691	Smartwatches*	10/15	0.712
Refrigerators (✓)				12/17	0.997
large size models	5/17	0.780	Smoke alarms		
compact models	8/14	-0.135	*battery operated smoke alarm*	1/16	0.442
Refrigerator freezer combos (✓) *(without chill compartment)*	7/16	0.190	*interconnected battery operated smoke alarm*	1/16	-0.781
Routers			Smoothie mixers (✓)	10/16	0.873
DSL	11/17	0.551	Soundbars and soundplates (✓)	12/14	-0.480
with ADSL modem	8/14	-0.936	*soundbars**	11/15	0.272
with VDSL and ADSL modem	8/14	0.932		11/17	0.702
			soundbar bundles with wireless bass module	11/17	0.104
Satellite TV receivers	4/14	-0.094			
single tuner	6/15	0.245	*soundplates*	11/15	0.383
twin tuner	6/15	0.327	Steam irons (✓)		
Security cameras			*conventional*	12/16	-0.053
outdoor	10/17	0.174	*steam ironing systems*	12/16	0.994
indoor	10/17	-0.430	System cameras (✓)		
Small water filters (✓)	5/15	-0.376	*with viewfinder*	3/14	0.352
Smartphones*	2/14	-0.203	*with electronic viewfinder**	3/15	-0.106
	7/14	0.194		3/16	-0.885
	11/14	0.274		4/17	0.165
	3/15	-0.032	*with optical viewfinder**	3/15	0.713
	8/15	-0.146		3/16	-0.135
	1/16	0.341		4/17	0.159
	5/16	0.554			

Product Categories and Pearson Correlations between Average Ratings from Amazon.de and *Stiftung Warentest* Quality Scores *(continued)*

Category	Issue	Correlation	Category	Issue	Correlation
System cameras (✓) *(continued)*			TV's *(continued)*		
*without viewfinder**	3/15	0.175	*48-50 inches**	2/16	0.619
	3/16	0.610		6/16	-0.988
	4/17	0.611		10/17	0.229
Tankless water heaters (*electric models*)	1/15	0.958	*49 inches*	10/16	0.893
			49-50 inches (LCD models)	12/17	0.175
Telephoto lenses (✓)			*55 inches (OLED models)*	12/17	0.621
for Canon cameras			*55-58 inches**	12/16	-0.680
large maximum aperture	7/17	0.706		12/17	-0.645
small maximum aperture	7/17	0.744	Vacuum cleaners (✓)		
for Nikon cameras			*bagged**	6/15	-0.088
large maximum aperture	7/17	0.967		5/16	0.944
Thermostats (✓)				7/17	-0.495
programmable thermostats	1/17	-0.792	*bagless**	5/16	-0.725
thermostats with Wi-Fi	1/17	-0.086		7/17	0.810
Toasters (✓)	4/16	0.697	*cord-free vacuum*	2/16	0.025
Tumble dryers	10/17	-0.597	*robotic vacuum cleaners*	2/15	0.590
*with heat pump**	9/14	0.301	Washing machines (*front load washer*)*	11/15	-0.357
	9/15	-0.120		11/16	0.072
	9/16	0.755	Wi-Fi receivers		
without heat pump	9/16	0.689	*network audio players*	8/17	-0.731
TV's (✓)			*connectors*	8/17	-0.011
*32 inches**	10/16	0.663	Wi-Fi speakers	12/16	0.457
	2/17	-0.094	Wireless speakers	11/14	0.032
	10/17	-0.719			
*40-43 inches**	10/15	0.896			
	12/15	-0.078			
	2/16	-0.287			
	10/16	0.556			
	10/17	0.993			

Note: The check mark indicates that the same or a similar product category was also in de Langhe et al.'s (2016a) database; [a]no variation in *Stifitung Warentest* quality scores; asterisked categories have been used to explore the moderating effect of product age on the relationship between average ratings and *Stiftung Warentest* scores.