



CATÓLICA  
LISBON  
BUSINESS & ECONOMICS

# The Impact of Extreme Levels of Electronic Word-of-Mouth Volume on Suspicion and Choice Intention

Inna Dickmänken

Dissertation written under the supervision of Filipa de Almeida

Dissertation submitted in partial fulfilment of requirements for the  
MSc in Management with Specialization in Strategic Marketing,  
at the Universidade Católica Portuguesa, March 2021

## **Abstract**

**Title:** The Impact of Extreme Levels of Electronic Word-of-Mouth Volume on Suspicion and Choice Intention

**Author:** Inna Dickmanken

Online product reviews have become a relevant information source and sales driver in the online purchase decision process. Being aware of the positive effects, many online sellers engage in measures to increase their amount of highly-rated reviews. Consequently, the instrument of reviews itself and fake reviews in particular became highly relevant for researchers, who put effort into understanding the effects of electronic word of mouth (eWOM) – especially its metrics of valence, the review’s star rating, and volume, the number of reviews – on variables such as purchase intention, perceived and suspicion.

The author elaborates on existing research and uncovers that until now no attention has been devoted to the possible effects of extreme values of review volume, although prevailing, e.g., on Amazon. This thesis attempts to fill the identified research gap by proposing consumers use a heuristic approach to interpret the meaning of these metrics. By means of an experiment, the author analyses whether extreme values of review volume in the beginning of the purchase decision process, namely consideration set formation, work as a suspicious cue that implies suspicion and deters consumers from adding a product to their consideration set. The results reveal that an extreme volume does not lead to consumer suspicion, only valence does. Further, it was shown that this resulting feeling of suspicion weakens choice intention while an extreme volume has no impact on choice intention. This thesis is a first contribution to fill the research gap of extreme values of volume.

**Keywords:** eWOM, eWOM volume, eWOM valence, online reviews, product reviews, fake reviews, heuristics, purchase decision making process

## Sumário

**Título:** O Impacto dos Níveis Extremos do Volume da Palavra Electrónica na Suspeita e Intenção de Escolha

**Autor:** Inna Dickmänken

As opiniões de produtos online tornaram-se uma fonte de informação relevante e um motor de vendas no processo de decisão de compra online. Conscientes dos efeitos positivos, muitos vendedores online tomam medidas para aumentar a quantidade de opiniões altamente positivas. Consequentemente, as opiniões e, em particular, as opiniões falsas tornou-se um tópico altamente relevante para os investigadores, que se têm debruçado sob este assunto na tentativa de compreender os efeitos da *word-of-mouth* electrónica (*eWOM*) - especialmente as métricas de valência, a quantidade de estrelas, e o volume, a quantidade de opiniões - sobre variáveis tais como intenção de compra e suspeita.

Até ao momento não se tem dado atenção aos possíveis efeitos dos valores extremos do volume de opiniões, embora prevaleçam, por exemplo, na Amazon. Esta tese tenta preencher esta lacuna de investigação, propondo que os consumidores usam uma abordagem heurística para interpretar o significado destas métricas. Através de uma experiência, a autora analisa se os valores extremos do volume de opiniões no início do processo de decisão de compra, nomeadamente a formação de um conjunto de considerações, funcionam como uma pista heurística que origina suspeita e dissuade os consumidores de considerar esse produto. Os resultados revelam que o volume em si não resulta em suspeita, mas a valência sim. Para além disso, mostrou-se que este sentimento de suspeita resultante enfraquece a intenção de escolha, enquanto volume extremo não tem impacto na mesma. Esta tese é uma primeira contribuição para preencher a lacuna da investigação de valores extremos de volume.

**Palavras-chave:** *eWOM*, *eWOM*, volume de *eWOM*, valência de *eWOM*, opiniões online, opiniões de produtos, opiniões falsas, heurística, processo de tomada de decisão de compra

## **Acknowledgements**

With this dissertation I am completing my Master of Science degree, which has been an amazing, yet challenging journey of continuous learning and self-development. However, this dissertation as the last step of my academic career for now would not have been possible without the people who supported me along the way.

First, I would like to express my sincere gratitude to my always supporting supervisor, Filipa de Almeida, who guided me through this project by sharing her expertise, ideas and experiences. I thank you for always taking the time, no matter how busy you were, to answer all my questions, to clarify my doubts, to provide insightful feedback and last but not least to always motivate me to proceed. Your support brought my thesis to a greater level and I am deeply grateful for having you as my thesis' supervisor.

Apart from that, I would like to acknowledge my colleagues from Remazing and especially Filip Egert for taking the time to answer all my questions about the online marketplace Amazon in general and product reviews in specific. Obtaining insights from real Amazon experts was a valuable enrichment for this thesis.

Additionally, I would like to thank my friends – new ones from the amazing city of Lisbon that I can call my second home and old ones from my hometown Hamburg – who spent time with me when I needed a break from my thesis and thus provided enjoyable distractions to rest my mind. Finally, I could not have completed this academic degree and especially this dissertation without my loving parents. I thank you for your infinite support and always believing in me.

# Table of Contents

- List of Abbreviations..... III**
- List of Symbols ..... V**
- List of Figures ..... VI**
- 1 Introduction ..... 1**
- 2 Theoretical Background ..... 3**
  - 2.1 Consumer Purchase Decision..... 3
    - 2.1.1 Process..... 3
    - 2.1.2 Impact of Suspicion..... 4
  - 2.2 Concept of Electronic Word-of-Mouth ..... 5
  - 2.3 Electronic Word-of-Mouth in Purchase Decisions ..... 6
    - 2.3.1 Relevance ..... 6
    - 2.3.2 Effects of Electronic Word-of-Mouth Volume ..... 8
    - 2.3.3 Prevalence of Fake Reviews and Suspicion about Electronic Word-of Mouth 10
  - 2.4 Derivation of Hypotheses..... 12
- 3 System of Electronic Word-of-Mouth on Amazon..... 13**
- 4 Empirical Study..... 15**
  - 4.1 Methodology ..... 15
    - 4.1.1 Research Design and Model..... 15
    - 4.1.2 Product Choice ..... 17
    - 4.1.3 Variable Measurements..... 18
    - 4.1.4 Procedure..... 19
    - 4.1.5 Sampling Method ..... 22
  - 4.2 Data Analysis and Results..... 23
    - 4.2.1 Data Cleaning and Final Sample ..... 23
    - 4.2.2 Scale Reliability and Manipulation Checks ..... 23
    - 4.2.3 Hypotheses Testing ..... 24
- 5 Final Discussion ..... 27**
  - 5.1 Summary of Research Findings ..... 27

|          |   |             |
|----------|---|-------------|
| 5.2      | Academic and Managerial Implications..... | 28          |
| 5.3      | Limitations and Future Research.....      | 29          |
| <b>6</b> | <b>Conclusion.....</b>                    | <b>31</b>   |
|          | <b>References .....</b>                   | <b>VII</b>  |
|          | <b>Appendix .....</b>                     | <b>XVII</b> |

## List of Abbreviations

|                |                                  |
|----------------|----------------------------------|
| AC             | Attention Check                  |
| ANOVA          | Analysis of Variance             |
| b              | Regression Coefficient           |
| CA             | Cronbach's Alpha                 |
| CC             | Control VOL - Control VAL (= G1) |
| CH             | Control VOL - High VAL (= G3)    |
| CMMM           | Causal Moderated Mediation Model |
| DV             | Dependent Variable               |
| eWOM           | Electronic Word-of-Mouth         |
| G1             | Group 1 (2-4 respectively)       |
| H <sub>1</sub> | Hypothesis 1 (2-3 respectively)  |
| HC             | High VOL - Control VAL (= G2)    |
| HH             | High VOL - High VAL (=G4)        |
| IV             | Independent Variable             |
| M              | Mean                             |
| MAX            | Maximum Value                    |
| MC             | Manipulation Check               |
| MEDV           | Mediator Variable                |
| MIN            | Minimum Value                    |
| MO             | Mediator-Outcome                 |
| MODV           | Moderator Variable               |
| n              | Sample Size                      |
| OM             | Outcome-Mediator                 |

|                |                              |
|----------------|------------------------------|
| p              | p-value                      |
| PDP            | Product Detail Page          |
| R <sup>2</sup> | Coefficient of Determination |
| SD             | Standard Deviation           |
| SE             | Standard Error               |
| SERP           | Search Results Page          |
| USA            | United States of America     |
| VAL            | Valence                      |
| VOL            | Volume                       |
| WOM            | Word-of-Mouth                |

## List of Symbols

US \$            US Dollar

**$\alpha$**             Alpha

£                Pound Sterling

**List of Figures**

**Figure 1:** Research Model ..... 17

**Figure 2:** Experimental Groups ..... 20

**Figure 3:** Model 7 (Hayes, 2018) ..... 25

**Figure 4:** Moderated Mediation Model ..... 26

# 1 Introduction

Online shopping has become part of consumer's daily lives. In Germany about every third person purchases online on a weekly basis (Idealo, 2020) motivated e.g., by convenience, 24/7 availability, greater product variety and price advantages („Global Consumer Survey 2020“, 2020). However, the missing opportunity to evaluate haptically the characteristics and product quality increases consumer uncertainty and decision complexity of online compared to in-store shopping (Ba & Pavlou, 2002). To make a profound purchase decision and to decrease uncertainty and risk, consumers gather as much information as possible. One of the most relevant pre-purchase information sources, decision aid and, eventually, sales driver are product reviews (Fretwell et al., 2013; Hu et al., 2008; Khammash, 2008). A Nielsen study revealed that about 65% of internet users always or often consider online reviews and evaluate them as a trustful source comparable to recommendations by friends/family (Kats, 2018).

Thus, for Amazon, as one of the largest E-Commerce platforms, reviews became one of the most crucial features. Bestseller products can show up to many thousand reviews and high ratings. Knowing about the review's relevance and being pressured by competitor products with many, highly-rated product reviews, sellers search for approaches – legal and illegal ones – to enhance review creation (F. Egert, personal communication, November 2, 2020; Appendix 1).

This importance is mirrored in the academic sphere, in which the relevance of online reviews has been recognized, resulting in a large body of literature on Electronic Word-of-Mouth (eWOM). Until now research has focused intensively on the positive aspects of eWOM, especially on the two metrics: eWOM valence (numerical star-rating) and eWOM volume (review quantity). Various studies show e.g., how eWOM valence can serve as a convincing quality signal (De Langhe et al., 2015; Tata et al., 2020), how eWOM volume positively impacts consumer awareness (Chen et al., 2014; Duan et al., 2008), increases information credibility (Zhu & Zhang, 2010), and how both metrics are relevant elements and heuristic cues for the decision and thus impacting sales (Magnani, 2020; Maslowska et al., 2016).

However, the downsides of product reviews – particularly of high eWOM volumes – have been discussed to a limited extent. Although volume, as above stated, seems to have a positive impact, investing in review quantity can lead to diminishing effects of profit (Zhao et al., 2013). Additionally, increased purchase complexity is observed as high numbers of reviews can lead to an information overload, if reviews are processed in detail instead of relying on the heuristic cue of the quantity itself (Maslowska et al., 2016).

Yet, despite being heavily present on Amazon (e.g., in the product category of headphones, the average amount of reviews on amazon.com is about 15,000 with a standard deviation of more than 33,600 (Appendix 2), the effect of extremely high values of volume has neither been taken into further consideration by research nor has it been critically scrutinized by managers.

Furthermore, studies that compared effects of high and low volumes of eWOM failed to incorporate extreme values in their design. For example, they focused only on maximum 420 reviews (Kordrostami & Rahmani, 2020) or simply summarized high volumes as “more than five reviews” (Maslowska et al., 2017b). This research lacuna is particularly important as consumers are suspicious about fake reviews (Zhang et al. 2016; Zhuang et al., 2018), which account for a large increase in review numbers (Petrescu et al., 2018). This results in a research gap, which encompasses the unanswered question of whether extreme levels of eWOM volume represent a suspicious cue decreasing choice intention in the consideration set formation on Amazon. The paper at hand attempts to find answers to this research question using an empirical approach.

The dissertation’s structure comprises four parts. First, the theories of the consumer purchase decision process and eWOM will be presented. In a subsequent chapter both concepts will be combined and current research findings about eWOM as a major information source for purchase decisions will be portrayed. Based on the already existing findings the research gap will be identified, hypotheses derived, and a conceptual framework developed. Before conducting an experiment, a short introduction about the review system of Amazon will be given, as it is the platform the study will be based on. The focus of the third part is the empirical study. A 2 (volume: control, extremely high) x 2 (valence: control, extremely high) between-subjects design will be used to gather data which will provide answers for the research question. The chapter includes the methodology, data analysis and finally states the results to the tested hypotheses. The final fourth part will focus on summarizing and interpreting the results, emphasizing the academical and managerial implications and finally presenting the limitations and providing recommendations for further research.

## 2 Theoretical Background

### 2.1 Consumer Purchase Decision

#### 2.1.1 Process

A consumer's purchase decision process traditionally consists of five phases: problem recognition, information search, evaluation of alternatives, purchase decision and post-purchase behaviour (Kotler et al., 2008). During information search and alternative evaluation, the marketing literature traditionally refers to a linear process – a funnel – in which consumers narrow down the number of products from the total to the awareness, the consideration and finally the choice set (Kotler et al., 2009). The process starts with many products, which are reduced step-wise based on available information until one is left.

However, as realized by McKinsey (Court et al., 2009) this process is outdated in today's digital environment, which confronts consumers with even more products and channels providing easy comparable information. Consequently, instead of a linear reduction, McKinsey defines the new process as a decision journey with a dynamic loop of adding and subtracting products. This circular model also includes an initial consideration set, comprising all brands that are familiar to the consumer. However, while the consumers are proceeding to the final purchase stage, they are faced with further brands and information. In contrast to the traditional approach, brands can be added to increase the consideration set or can even supersede previously considered strong brands. Hence, strong brands cannot be simply assumed to remain in the consideration set, but might be ignored once new, more appealing products become attractive to the consumer. This makes the stage of consideration set formation as one of the most crucial ones for brands.

Today more than ever, to make decisions about including or excluding brands and the final purchase, individuals actively collect and process information such as reviews or word-of-mouth recommendations (Court et al., 2009; Gross, 2014). Based on factors like involvement individuals either chose taking a systematic or a heuristic route to gather the necessary information to develop judgements – known as the Heuristic-Systematic-Model (Chaiken, 1980).

The systematic route is chosen for purchase decisions with high product involvement, that is, more expensive and high-risk products involving the individual more emotionally and monetary in the decision (Gu et al., 2012; Traylor, 198). Decisions within the systematic route require high cognitive effort and consumers' motivation for an extensive information search. So, they can form a profound judgement, ensuring to satisfy needs and reduce risks.

In contrary, as human beings are economy-minded individuals and cognitive resources are limited, the heuristic route, in which decisions are made rather fast, is preferred if the motivation to process information is low, which tends to be the case for decisions about low involvement products which have lower price and risk levels (Traylor, 1981). Also, if detailed information is not accessible or can only be processed with high cognitive effort, they will be substituted by other attainable attributes. These attributes are easily accessible information cues or simplified metrics, called heuristics, that are used by consumers to form an opinion about any product feature (Kahnemann & Frederick, 2004). Based on this heuristic attribute substitution process consumers are able to derive interpretations and eventually make decisions in a fast way.

### **2.1.2 Impact of Suspicion**

Today's online environment provides consumers with an immense amount of information. Hence, no matter if the systematic or heuristic route is chosen, this information overload makes it even more difficult, yet important, for consumers to identify which facts can be trusted for decision making (Ramadani et al., 2014). Trust towards the information given by online sellers is built upon the consumer's expectation that they act in an honest way and will not exploit the consumer's vulnerability and his disadvantageous position in what concerns information asymmetry (Hong & Cha, 2013). Hence, if positive product information (e.g., a product review) is evaluated as trustworthy it will be a predictor for further steps, such as including it in the consideration set or the final purchase (Augusto et al., 2019; Cheung et al., 2009; Gross, 2014).

In contrast, opposite effects can be identified if there is lack of trust, which arises suspicion (Sinaceur, 2010). The source of suspicion is a deviation between a consumer's expectations or general norms and the given information. Suspicious cues, such as extreme values (e.g., overly positive product reviews), that are exceeding the norm (Filiari, 2016; Yang, 2018), can lead to the identification of this deviation (Zhuang et al., 2018).

For information that is perceived as trustworthy, further information search is reduced as the given one is accepted. In contrast, if consumers face suspicious cues, information search will either be extended to profoundly evaluate the information before decision making occurs (Sinaceur, 2010) or they will distance themselves from interacting with or purchasing products from that brand (Shareef, 2013). The more extensive this information search, the more cognitive effort is required. Whether a consumer is likely to perform a detailed examination or distance himself from the product, according to the Heuristic-Systematic-Model, will likely depend on the cognitive route the consumer follows. Under the heuristic route avoiding cognitive effort, a

suspicious cue can rather immediately result in disengaging from a decision of considering a product. Under the systematic route, further search for information becomes the most likely outcome. Nevertheless, if this lack of trust about certain product information cannot be minimized by further search, it leads to a major purchase barrier (Augusto et al., 2019).

## **2.2 Concept of Electronic Word-of-Mouth**

As previously stated, information is an essential element during a consumer's decision journey. Especially in the online environment characterised by information overload and high risk, eWOM can be seen as a major information source (Almana & Mirza, 2013).

The concept of Word-of-Mouth (WOM) has developed over time. Traditionally, WOM is seen as the act of communicating and sharing opinions – negative and positive – e.g., about products or services, consumers have bought and already experienced. This personal information source has been proven to be highly relevant as it is more credible to consumers than any advertising attempt by a brand (Aslam et al., 2011). EWOM is an enhancement of traditional WOM by including the electric component. Similar to WOM, it also encompasses positive or negative comments. However, the powerful difference is that it is shared via the internet on an online platform and is therefore made available to an immense number of individuals (one-to-many). In contrast, the traditional concept is based on one-to-one communication in a rather private manner (Cheung & Thadani, 2012; Kozinets et al., 2010).

There are different ways in which an eWOM statement can be shared in the online environment, such as posts, “likes”, “tweets”, or reviews. The type depends on the channel which is used to provide the information. There are three prevailing channels where eWOM messages can be found. Posts, likes and tweets appear in social media channels (e.g., Facebook, blogs.). Review sites (e.g., Yelp) and e-commerce platforms (Amazon) entail reviews. For purchase decisions consumers are mainly influenced by reviews on e-commerce platforms (Rosario et al., 2016).

Online reviews commonly consist of two components. Individuals who are writing a review about a product usually are asked first to give a quantitative rating and second to write a qualitative text (Mudambi & Schuff, 2010). The effects of the qualitative element have been researched regarding multiple variables such as text subjectivity level, spelling errors, readability scores (Ghose & Ipeiritos, 2011) or sentiments (Hu et al., 2014; Li et al., 2019). However, to answer the research question the focus of this dissertation will encompass the quantitative components of reviews (eWOM metrics). There are two prevailing metrics: eWOM valence and

eWOM volume. Another, albeit less considered, metric is eWOM variance. However, as this value is not immediately visible to consumers (Chapter 3) it will not be further discussed.

EWOM valence quantitatively reflects the overall positive or negative evaluation of a product, measured by the star rating on a Likert scale ranging from 1 to 5. It can either be considered isolated as a rating of a single review or as a summarized, averaged value for many reviews. An advantage of this quantitative evaluation is that it is less prone to misinterpretation in comparison to qualitative content (Maslowska, 2017a). However, it is also observed that ratings are positively biased (e.g., 4- and 5-star reviews make up 70%) which complicates the interpretation of a rating as the neutral reference point is indistinct (Yang, 2018). The second metric eWOM volume refers to the overall number of reviews given for a product.

## **2.3 Electronic Word-of-Mouth in Purchase Decisions**

The previous two chapters elaborated on the process of decision making and the concept of eWOM. Both make up essential theoretical background knowledge for this paper. In this chapter both are combined to understand the relevance of eWOM for decision making in general and to discuss the identified effects of eWOM volume and suspicion about eWOM in purchase decisions. It will serve as a basis to derive hypotheses.

### **2.3.1 Relevance**

The existing literature suggests there is great interest in investigating the effects of eWOM on variables like purchase intention (e.g., Augusto et al., 2019; Park et al., 2007), sales, a sales-proxy like the Amazon sales rank<sup>1</sup> (e.g., Chevalier & Mayzlin 2006; Cui et al., 2012; Zhu & Zhang, 2010) or profit from incentivized reviews (e.g., Zhao et al., 2013). Although all studies use various models, there is agreement that reviews have a major impact on purchase decisions.

There are many reasons why reviews are so relevant in purchase decisions. A main explanation is that reviews serve as additional information, whose source is opinions of other customers. Thus, they decrease consumers' perceived risk and uncertainty in the e-commerce environment, in which they are at the undesirable end of information asymmetry (Mavlanova et al., 2012). Due to this extra amount of information, consumers are also able to make more profound purchase decisions for products, increasing the likelihood to perfectly suit their needs. Further,

---

<sup>1</sup> Amazon Sales Rank is an indication for product sales. It uses an inverse relationship, i.e., bestselling products have a lower sales rank than products which are barely sold. It is widely accepted by many researchers to use it a proxy for sales (Floyd et al., 2014).

especially when the numerical metrics are taken into account, search and evaluation efforts are decreased and the comparison between products is facilitated. Finally, individuals use reviews to find social assurance supporting their choices (Bailey, 2005).

Although the power of product reviews in decisions is evident, the relevance within the actual process of decision making has to be discussed. Regarding the Heuristic-Systematic-Model, product reviews can be taken into account for both routes. For systematic information processing the review's qualitative content is read thoroughly. It requires cognitive effort not only due to the fact of intensive reading, but also because the consumer needs to actively execute a click or scroll down as the review content is not accessible immediately (Maslowska et al., 2017b; Park et al., 2007). In contrast, the heuristic route relies on the quantitative components of the review. The average rating and the review volume are usually presented prominently and thus serve as a heuristic cue (Maslowska et al., 2017a).

Which route is chosen and subsequently which information of the review are processed does not merely depend on product involvement, as stated in Chapter 2.1.1, but can also vary depending on the stage of the decision process. In general, during the formation of the consideration set, mostly the quantitative metrics are taken into account. They serve as helpful heuristic cues, especially if the list of alternative products is large. Hence, the consumer skims through the list of alternatives, forms a pre-judgement and intuitively decides whether to include or exclude a certain product in the consideration set (Li et al., 2020; Magnani, 2020). In contrast, the qualitative part of a review has higher relevance for the final choice. It provides the persuasive information an individual needs to finalize a choice (Hu et al., 2014). Combining this with the knowledge about involvement, it is important to realize that the just explained process of using both quantitative and qualitative review elements is more typical for high involvement products. Thus, high involved consumers do not only follow the systematic route, but rather use a mixed approach when it comes to review information processing (Filiari & McLeay, 2014). For low involvement products, however, the consumer predominantly relies merely on the quantitative information for consideration set formation and then also makes the final choice upon these quantitative cues.

As the relevance and the effects of eWOM during purchase decisions are not alike for all products, it is important to discuss how the product category may fulfil a moderating role. Various categorizations for products are possible. A study by Dai et al. (2019) analysed the use of reviews for material and experiential purchases. Performing a material purchase means to buy a tangible product which can be possessed, whereas experiential purchases are about events that

can be lived through (Van Boven & Gilovich, 2003). The study's results show that reviews are more important for material products as consumers tend to rely more on reviews in a purchase decision for these goods than for experiential goods. This is because the evaluation of an experiential good is based on a more subjective perception, whereas a material purchase can be evaluated based on an objective quality assessment.

Another approach categorizes goods into search and experience products<sup>2</sup>. Search goods can be evaluated by consumers before the purchase based on own past experiences, visible product features or descriptions. In contrast, experience goods can only be evaluated after consumption leading to a higher level of uncertainty during the purchase decision (Nelson, 1974). Although Nelson classifies goods as search or experience, it is important to mention that a single product – although categorized as a search product – can have some experience characteristics and vice-versa (Jourdan, 2001). EWOM can decrease the uncertainty and allows for a certain degree of pre-evaluation. Thus, eWOM appears to have a more relevant role in decision making for experience goods than for search goods (Park & Lee, 2009; Zhao et al., 2013). However, to give a complete overview of the current literature, it has to be mentioned that there are also opposite findings available: Hao et al. (2010) find evidence that reviews have a stronger effect for search than for experience goods. The underlying premise is that an evaluation of experience goods rather states the individual experience of the review writer instead of the real intrinsic product quality. This experience might be different for each individual which makes it harder to transfer. As search goods can easily be described based on uniform and definite standards, an evaluation about them most likely reflects the actual product quality.

### **2.3.2 Effects of Electronic Word-of-Mouth Volume**

The preceding chapter considered the relevance of eWOM in purchase decisions in general. As the research question concentrates on eWOM volume, past research findings about this metric should be emphasized now.

Most research on eWOM volume discussed its positive impact on the final purchase decision and, accordingly, sales. For example, for experience goods like books (Amblee & Bui, 2011; Chen et al., 2004; Chevalier & Mayzlin, 2006; Sun, 2012), and for search goods like digital cameras (Zhang et al., 2013), a significant impact of review volume sales or sales ranks was found. Additionally, the fact that review volume increases consumer awareness and is a

---

<sup>2</sup> Often also the category of credence products is considered. These are products that even after consumption can hardly be evaluated due to a lack of knowledge. Consequently, they are less relevant for a study about eWOM, which is based on consumers' evaluations.

heuristic cue implying popularity and enthusiasm indirectly leads to more sales (Liu, 2006; Maslowska et al., 2017b; Park et al., 2007). This effect is especially helpful for less popular products. With rising review numbers those products are perceived as more popular and users are rather convinced to purchase them (Zhu & Zhang, 2010). Although there is evidence for the impact of eWOM volume on sales it is important to understand that it is not a one-way street. In fact, volume rises with more people purchasing and sharing opinions (Elwalda & Lü, 2016). Further, although most research found significant results for the relation of volume and sales there are also studies failing to show this relationship (Chintagunta et al., 2010; Clemons et al., 2006; Gopinath et al., 2014), making it even more important to investigate this metric further.

Critical effects of eWOM volume that are relevant for the paper at hand are still limited as they do not focus on the heuristic cue of the displayed count of reviews but rather on the review content. Maslowska et al. (2017a) for example studied the impact of volume on purchase intention and found contrasting results to the positive effects stated above. They showed a decreasing effect on purchase intention with rising numbers of volume. However, the result was attributed to information overload. It arises when the consumer focuses on the review texts. The more information (more reviews) is available to evaluate and incorporate in a decision the higher the chance of information overload (Park & Lee, 2008). The authors believe, the effect would have been positive, if eWOM volume had been considered as a heuristic cue.

Wang et al. (2018b) also showed that consumers tend to evaluate increasing volumes as less helpful on Amazon as further reviews only lead to marginal additional information – still not considering review volume as a heuristic. The relevance of the very first reviews was also investigated by Zhao et al. (2013). As firms know about the positive impact of eWOM on purchases they tend to incentivize users for writing a product review. Using experience goods, books, they examined to what extent this act of incentivization is profitable with an increasing volume. Their results revealed as well diminishing returns of rising review numbers. While the very first posted reviews still have a relevant impact on profits and market share, investing in further ones is decreasingly important and can even cause losses.

Although there is evidence for the relevance of eWOM volume in a purchase decision, researchers understood that this concept cannot be considered isolated. Thus, to get a more detailed understanding of the dynamics of eWOM volume, studies that incorporated the metric of valence should not be ignored.

The positive effects of high volume are, naturally, not valid for all conditions of valence. Two products both with high volumes can elicit a different impression on a consumer, depending on

valence. High valence will attract, whereas low ratings paired with high volume will deter the consumer. Thus, the credibility and the consumer's confidence in the positive or negative score is enhanced by volume (Khare et al., 2011; Kostyra et al., 2016).

Using a conjoint analysis, Kostyra et al. (2016) reveal that volume does not play a moderating role for the effect of valence on sales for low and medium levels, but for high valence levels it does. Further, this effect is diminishing as trustworthiness in review ratings – which is achieved through volume – does not increase linearly. These results are supported by the results from Maslowska et al. (2017b), who additionally looked at product risk. They found that review scores impact purchase probability more for high than for low levels of eWOM volume. This effect is steeper if the risk related to the purchase is low. For a low-priced product (implying lower risk) a small increase in review volume already has a measurable effect on purchase probability. In contrast, for high-priced products, it is harder to convince consumers solely based on the rating. Consequently, a larger volume is needed to see the effect.

Kordrostami and Rahmani (2020) complemented these findings from the opposite perspective. They give proof that volume has a stronger impact on purchase intention for moderate and high valence ratings. However, a rating score within a lower range (2.1-2.5 stars) is already an explicit signal for low satisfaction and, in that case, volume has no impact anymore.

Although based on the current research findings one might believe that eWOM volume can serve as a choice driver, it is necessary to critically scrutinise whether the relationship between volume and sales is valid for all volume levels. Until now research (except for Khare et al., 2011) has been neglecting extreme eWOM levels and focusing on review volume levels around hundreds (Appendix 3). Hence, the question of whether extreme levels of eWOM volume show different effects – e.g., instead of diminishing marginal effects (Kostyra et al., 2016), increasing scepticism with volume levels of many thousand reviews – is still left unanswered.

### **2.3.3 Prevalence of Fake Reviews and Suspicion about Electronic Word-of-Mouth**

A main driver for rising levels of eWOM is not only sales, as stated in the previously, but also the emergence of fake reviews (Petrescu et al., 2018). Fake reviews are characterized by being false, deceptive, and misleading because their content is inconsistent with the real product evaluation (Mathews Hunt, 2015). These inconsistencies arise as the ones writing the review have not necessarily experienced the product, but just pretend to be a real consumer (Banerjee & Chua, 2017; Hu et al., 2012).

Due to the power of eWOM in purchase decisions, managers are considering eWOM as a powerful marketing opportunity (Yang, 2018) and hence are increasingly inclined to engage in fake review creation. The main motivations go beyond the financial nature due to the review's impact on product perception, sales volumes, and profits, and include being pressured by fierce competition in the transparent online environment (Wu et al., 2020).

A common tactic to make users engage in writing fake reviews is to incentivize them either with a financial reward or free products, which leads to biases and limited objectivity (Thakur et al., 2018; Wang et al., 2018a). It is also possible that consumers are intrinsically motivated to write fake reviews, e.g., due to a desire for revenge (Thakur et al., 2018). However, for the paper at hand extrinsically motivated and merchant-initiated approaches are considered.

Companies are increasingly exploiting these approaches which leads to an explosion of fake reviews and internet users who become more suspicious about online product reviews (Zhang et al., 2019). Nowadays only 14% of Amazon-shoppers in the USA fully trust the reviews they are confronted with (Tinuiti, 2020) and the percentage of people that are sceptical about reviews are rising (Murphy, 2018). This opens a new area of research and scholars have found great interest in examining the impact of fake reviews on consumers' purchase decisions. Research refers to distrust, suspicion, low credibility, and uncertainty about the review as the dominating psychological reactions due to fake reviews (e.g., DeAndrea et al., 2018; Jin Ma & Lee, 2014; Zhao et al., 2013; Zhuang et al., 2018). On a large scale, if consumers are faced with those deceptive reviews, the actual positive effects of eWOM are compromised (Zhao & Zang, 2020). Consequently, consumers tend to disengage from considering a product for a purchase decision which arouses the stated negative reactions. Accordingly, purchase intention is weakened and sales decrease (Zhuang et al., 2018).

Nevertheless, it is important to mention, that these negative effects only occur if the consumer is able to recognize a review as fake due to available cues that imply the manipulation. Hu et al. (2012) investigated whether consumers are able to detect fake reviews rather through quantitative ratings or sentiments of the review statement. Their findings identify the review rating as the more dominant cue indicating manipulation. An example for such a cue are overly positive ratings (full 5-star rating). Consumers who are faced with such a rating are more likely to distrust the evaluation and become suspicious (Maslowska, 2017a; Yang, 2018). Hence, products that exhibit a suspicious rating are most likely not considered for a purchase decision.

Zhuang et al. (2018) also investigated the effects of suspicious cues and show that the effect of fake reviews on sales follows an inverted u-shape: As the very first product reviews are believed

to be from authentic writers, adding fake reviews at a stage of low volume levels does not add much damage and can still lead to the known positive effects. However, these effects wear out and, finally, turn into negative ones with increasing numbers of fake reviews, which provide a larger presence of manipulative cues that arise suspicion.

Besides, the author finds evidence that this effect of suspicion on purchase intention is moderated by the consumer's expertise. Users with little expertise in the e-commerce environment are more susceptible to be deceived, whereas experienced consumers are more likely to become suspicious. Further, if consumers are already sceptical about eWOM they are also more inclined to detect those suspicious cues (Zhang & Barnes, 2019). Similar effects can also be seen for low vs. high involvement purchases. While users that are highly involved in a decision are more likely to examine a review's trustworthiness and notice suspicious cues, low involved consumers more easily fall prey to fake reviews (Filiari, 2016). Additionally, brand strength also impacts the negative effects of fake reviews: Weak brands suffer more from fake reviews than strong ones as prior certainty about brand knowledge and overall reputation makes consumers less suspicious (Zhuang et al, 2018).

Apart from ratings, sentiment, or extreme values, further cues that were found to draw the consumers attention to (un-)trustworthy reviews are review source, pattern in reviews and website type (Filiari, 2016). To the best of the author's knowledge, up to the present moment, the factor of extreme levels eWOM volume as a cue arising suspicion was neither incorporated in any study nor critically scrutinized. This opens a research gap which this paper attempts to close.

## **2.4 Derivation of Hypotheses**

To fill the stated research gap, this dissertation focuses on finding an answer to the question of whether extreme levels of eWOM volume are a suspicious cue decreasing choice intention in consideration set formation on Amazon. The answer will be delivered by means of a study that attempts to validate the following four hypotheses. The concrete research design and model will be introduced in detail in Chapter 4.1.1.

Based on the literature presented above it is known that extreme values which outrange the common norms can work as cues that arise suspicion. Thus, it is not unreasonable to assume that an extreme number of reviews can serve as such a cue. This is further supported by the fact that fake reviews are among others one driver for review volume and consumers are aware of fake reviews prevailing in the e-commerce environment. Consequently, one can derive that

consumers might associate a high review count with the existence of fake reviews. Hence, the following is hypothesized:

**H<sub>1</sub>:** *Extreme high volumes of eWOM will lead to higher perceived suspicion of manipulation compared to lower levels of eWOM volume.*

To understand the effects of eWOM volume it is suggested that this metric should not be analysed isolated but rather in combination with valence. Studies showed that extreme high ratings (all five stars) can seem untrustworthy to consumers and make them suspicious. If extreme high volume is paired with an overly positive review a consumer's evaluation is affected by two suspicious cues. As suspicion rises the more cues are recognized, the second hypothesis is:

**H<sub>2</sub>:** *The effect of an extreme eWOM volume level on suspicion of manipulation is moderated by an extreme high valence rating.*

Decisions are based on available information. If the given information does not seem trustworthy and arouses suspicion, a consumer is likely to distance from these cues for decision making to avoid risks. Consequently, products or brands that display such suspicious cues are unlikely to be included in the consideration set. Therefore, the third hypothesis focuses on the consequences of extreme high eWOM volume and increased suspicion:

**H<sub>3</sub>:** *Higher perceived suspicion of manipulation will lead to lower choice intention.*

Although the impact of extreme eWOM volume on choice intention were hardly studied until now, one can assume that it might be considered as a heuristic cue that negatively influences a consumer due to many additional aroused feelings besides suspicion, such as the lack of trust and credibility, scepticism etc. Thus, the fourth hypothesis states that extreme volume will result on lower choice intention as a result of any possible explanatory variable(s) (mediator):

**H<sub>4</sub>:** *Extreme high levels of eWOM will lead to lower choice intention.*

### **3 System of Electronic Word-of-Mouth on Amazon**

As there is a strong effect of eWOM on sales for e-commerce platforms, the experiment will use examples from such a platform (Rosario et al., 2016). Due to its popularity and high market share in the e-commerce sector the online platform Amazon is the one that will be used (eMarketer, 2019). Further, eWOM volumes reach high levels here which affords less manipulation and thus increases external validity. Before the experiment will be explained and results presented, an introduction about the review system of Amazon will be provided.

Already since 1995, Amazon allows consumers to evaluate the purchased products in form of reviews on their website (Ante, 2009). Since then, reviews have developed into one of the platform's major features that consumers do not want to miss anymore for purchase decisions ("Demand", 2020). On Amazon an overview of the reviews is displayed prominently underneath the product title on the search results page (SERP) and the product detail page (PDP; Appendix 4). On the SERP the star-rating and the number of reviews is shown. It is important to mention, that already with a rating of 4.8 stars all five stars are displayed. A distribution of the number reviews for each valence level – as a proxy of the eWOM variance – only becomes visible if the user actively hovers over the reviews. This summarized overview of valence, volume and variance is also displayed at the top of the PDP. However, to see the details of the reviews and the qualitative content a user has to either click on the review overview or scroll down.

Brands that are selling their products on Amazon are aware that reviews have a positive influence on the Amazon algorithm and so are, as such, a major driver for purchase decisions. An interview with Filip Egert (

Appendix 1), who works in the management of a major Amazon agency in Germany, was conducted to get insights about the importance of reviews for sellers. Egert emphasizes that especially the fierce competition on the platform does not let firms without any reviews survive. Hence, sellers use different approaches to increase their numbers of highly rated reviews. According to Egert these approaches can be categorized into three groups.

First, there are two official programs offered by Amazon – Amazon Vine and Amazon Early Reviewer Program – that enable merchants to get honest, but high qualitative reviews. In both cases the review author is incentivised for writing the review. Further, the review is labelled with “Vine Customer Review” or “Early Reviewer Rewards”. However, both programs can only be used for the creation of the very first reviews. The second approach is to strategically organize one's product presentation. Reviews are cumulated if products with different variations (e.g., size, colour, style) are grouped together. Thus, for example if a single style has a low number of reviews it yet appears as if it has a lot because it benefits of the other items' reviews. The third way to increase reviews are numerous and rather critical approaches that do not provide authentic evaluations but rather fake reviews. The most common technique is to offer a financial incentive or free products in return for positive 5-star-reviews at a large scale through Facebook groups (Lee & Murphy, 2020; Schiffer, 2020).

This fake review generation has become popular. It is estimated that about 42% of 720 million reviews are deceptive and unreliable on Amazon (Lee, 2020). Amazon is taking action against

fake reviews, e.g., by using AI and machine learning technologies to filter and detect such reviews (Gibbs, 2015; Schoolov, 2020) and even by suing fake review authors (Gani, 2015).

High eWOM volumes are prevailing on Amazon in different categories. Still, one must mention that high review volumes cannot only be attributed to the actively taken measures presented above, but also due to organic increase. The most obvious reason is sales. If a product is bought very frequently, there is a higher chance of users writing reviews. Additionally, Egert states that products which are offered already for a long time on Amazon might also have collected many reviews. Furthermore, Amazon did change the review system in 2019. While first it was mandatory to also provide a written text, nowadays it is possible to only rate the product. As it is a faster process more consumers agree to provide a rating (Kaziukenas, 2019). Thus, the question is in how far extreme high volumes of eWOM might harm sellers on Amazon, no matter if the number eventually is rooted in fake reviews or actually in real authentic reviews that e.g., have been collected over years.

## **4 Empirical Study**

### **4.1 Methodology**

#### **4.1.1 Research Design and Model**

Apart from the qualitative approach of conducting an interview with an Amazon expert (Appendix 1 and Chapter 3), for finding validation for the stated hypotheses the major focus lies in a quantitative study in the form of an online survey, based on Qualtrics. Among others, benefits of an online survey are control of the sampling and the survey logic, low administration cost, and ease of data analysis (Evans & Mathur, 2005).

The study's intention is to uncover a causal relationship between extreme levels of eWOM volume and choice intention. According to Malhotra et al. (2017) experimentation is an appropriate research approach that enables researchers to infer causal relationships as the one mentioned above. In such an experiment the researchers manipulate the independent variable(s) (IV), called treatments, and measures the effect on a dependent variable (DV). Simultaneously, they control for any extraneous variables, that might influence the DV's outcome and thereby ensure internal validity (Bryman & Bell, 2011).

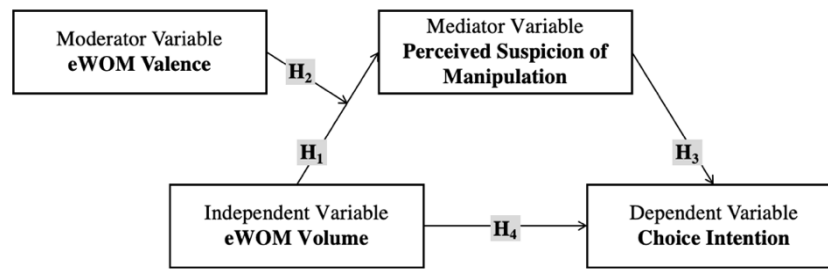
In the study at hand, the test units of the experiment are faced with two different treatment levels of eWOM volume. The outcome of the two levels can be identified by a change in the

DV, the choice intention. In addition to eWOM volume, valence is considered as a third variable, in particular as a moderator variable (MODV) with two different treatment levels. A MODV is used to explain if the strength of a relation between the IV and the DV varies depending on different levels of the MODV (Wu & Zumbo, 2007).

Including the MODV results in a 2 (volume) x 2 (valence) factorial design, which is a suitable experimental design to understand the impact of the two IVs at different levels on the DV, namely choice intention, while also enabling to identify interactions between the IV and MODV. Concretely, the two levels of each treatment variable are combined in all possible ways leading to four cells. During the experiment, participants are randomly assigned to one of these four cells. Thus, extraneous factors are equally distributed among the groups and prior equality of test groups is maximized, which is an important contribution to internal validity (Malhotra et al., 2017). The fact that participants are allocated to one experimental group (one cell) and thus are only exposed to one combination of treatments is the structure of a between-subjects design. It allows to compare the participants' behaviour between the groups while avoiding carry-over effects. This effect can occur, especially in within-subjects designs, if participants' reactions to the treatment are influenced by experiences from a previous exposure to another treatment (Charness et al., 2012).

Finally, to enhance the understanding of the causality between the IV and DV, a mediator variable (MEDV) – perceived suspicion of manipulation – is included in the model as a fourth variable. A MEDV is a variable that is caused by the IV and then impacts the DV, thereby transmitting the effect from IV to DV (MacKinnon, 2011). It is often a responsive variable, e.g., a cognitive state, that is able to change within a test unit after being exposed to a treatment condition (Wu & Zumbo, 2007). Suspicion as being a feeling of uncertainty, a perception of malintent together with increased cognitive processing can be defined as such a cognitive state (Bobko et al., 2014). Together with the DV, IV and MODV, this variable completes the research model (Figure 1) – a causal moderated mediation model (CMMM). This model aims to reveal inferences about the indirect effects of eWOM volume on choice intention through perceived suspicion of manipulation, while its strength depends on the moderator of eWOM valence.

**Figure 1:** Research Model



#### 4.1.2 Product Choice

In order to increase external validity – the degree of generalisation of findings – of the experiment, not only a real marketplace but also a concrete product listing is used (Bryman & Bell, 2011). Furthermore, to control that changes in the DV are actually caused by the treatment variables the same product listing – with unchanged images, titles, price, etc. – is used for each experimental group increasing internal validity.

The decision for an appropriate product category is based on seven criteria. First, the product has to be a material instead of an experiential good, as only these are sold on Amazon and reviews are more important for this category (Chapter 2.3.1). Second, as research about the effect of reviews for search and experience goods are still inconsistent (Chapter 2.3.1), a product should be selected with both search and experience characteristics, avoiding the extremes. Third, high review volumes should be prevailing in the chosen product category to decrease manipulative intervention and increase external validity. Fourth, the participants should be faced with a certain level of risk, implied by the price that is higher than a few US dollars. Considering prospect theory one can assume, that consumers do not endeavour to question the trustworthiness of a product review if prospective losses are minimal. However, if prices are higher and potential losses are more severe, consumers are likely to pay more attention to possible suspicious cues such as extreme review counts (Kahnemann & Tversky, 1979). Fifth, to avoid gender biases a unisex product is chosen that neither gender is especially familiar with. Sixth, the product should not show an extreme design that might influence the participants' answers. Seventh, to avoid the limited negative effects of fake reviews for popular brands and biased responses due to brand preference, the product should be from a rather unknown brand.

A product, which is a material good and matches all other criteria are headphones. They show concrete search characteristics, such as battery lifetime, while still having experience characteristics like sound quality or ergonomic fit. Further, high amounts of reviews are typical for the headphone category. On Amazon.com the average volume for headphones is 14,988 with a

SD of 33,668 (Appendix 2). The average price lies at US\$ 41.53 with a SD of US\$ 53.97 supporting the criteria of a moderate risk level. As headphones are not only designed for one gender (e.g., like electric razors) the fifth criteria is also fulfilled. A concrete product that not only fits the already stated conditions, but also the sixth and seventh are the TOZO T10 headphones<sup>3</sup> due to their simple design and their less popular brand TOZO compared to strong brands like Bose, Apple etc. Further, they show an extreme high volume with over 136,500 reviews and a price of US\$ 49.99 which does not require further manipulation.

### 4.1.3 Variable Measurements

In the following, the concrete levels, constructs and items per variable, that were already introduced briefly in Chapter 4.1.1, will be presented.

The IV is the eWOM volume. It is a categorical variable and shows two conditions of volume: one control condition and one extremely high condition. To make a decision about the concrete volume level for each condition and to minimize manipulation, the review volumes for headphones on Amazon were taken into account. Headphones on average have 14,988 reviews with a SD of 33,668. Due to the high SD, indicating high variation, the median (= 3,852) being more robust to outliers is considered. This value is used for the control condition. Although the initial plan was to use a value around hundreds, that existing research has already examined and resulted in positive effects, the crawler data has shown that the overall review volume within this category reaches a high level. Thus, with the demand for external validity in mind and compared to the high review volumes in this category, it seems reasonable to take 3,852 as the control condition. Extremely high volumes are those that exceed the average. Therefore, the second condition is based on the concrete review volume of the chosen product, again to increase external validity and decrease manipulation. The TOZO headphones show 136,571 reviews, which is not only a value strongly above the average but also serves as a suspicious cue.

The DV measures the users' choice intention of including or excluding the product in their consideration set. Although the actual behaviour of the participants on Amazon will not be measured in the experiment, it can be assumed that the intention for or against choosing the product is a main indicator of the real behaviour (Kim et al., 2008). In previous studies, purchase intention is commonly measured on a one-item scale (e.g., Kordrostami & Rahmani, 2020; Zhuang et al., 2018). As choice intention precedes the purchase intention in the customer

---

<sup>3</sup> ASIN: B07J2Z5DBM ([www.amazon.com/dp/B07J2Z5DBM](http://www.amazon.com/dp/B07J2Z5DBM)).

journey and asks for a similar behaviour, it appears reasonable to adapt the formulation and the scale of prevailing purchase-intention-questions. Using a Likert-Scale of 6 categories ranging from 1 (*Very unlikely*) to 6 (*Very likely*) the participant is asked to indicate the likelihood of taking the product into further consideration. An even number of categories is used to force the participant to decide for or against a choice (Chyung et al., 2017). This replicates a real-life decision, which also only provides the option to choose an item and consider it further or not.

EWOM valence is included into the model by the MODV. Similar to the IV of eWOM volume, this variable has two levels. The decision for the control condition is based on the crawler data. The average rating is 4.3 stars. This value is shown as 4.5 stars in the review overview section of the product listing within the SERP and therefore likewise in the experiment. For the second level an extremely positive 5-star-rating is chosen providing a suspicious cue as consumers are sceptical of too positive reviews (Maslowska et al., 2017a).

Finally, the MEDV measures the perceived suspicion of manipulation. It is based on a construct used by Kirmani & Zhu (2007) to measure deceptiveness of an advertisement, that was later also used to measure suspicion about reviews (Zhuang et al., 2018). The semantic differential scale contains three items: “*Believable - Unbelievable*”, “*Truthful - Not Truthful*”, “*Non-Deceptive - Deceptive*”. It is a 7-point scale and participants can be indifferent between the two descriptors. Scales are coded reversely so that suspicion is expressed with a higher value.

#### **4.1.4 Procedure**

The experiment starts with an introduction stating the approximate duration, the confidential use of the responses given and the broad topic, shopping on Amazon (Appendix 5). Concrete information about the study are not provided avoiding participants being biased.

Derived from existing research, expertise moderates the effect of suspicion on purchase intention (Chapter 2.3.3). To control for this effect, participants are required to indicate their shopping frequency on Amazon – as a proxy for expertise – based on six available options ranging from “*Every Week*” to “*Never*”. Participants are then redirected to the next question, even if they have never purchased anything on Amazon. It is assumed that participants unfamiliar with Amazon are also able to give valuable answers.

Afterwards, participants are introduced to the specific situation which asks them to imagine they are currently in the consideration phase of the purchase decision process: They should imagine scrolling through a SERP and think about which product to click on to get more information. This stage was chosen as here reviews play a significant role for brands to be in- or

excluded in the consideration set (Court et al., 2009) and are even more relevant than in the final choice stage (Jang et al., 2012).

**Figure 2:** Experimental Groups

|                            |  |   |
|----------------------------|--|---|
|                            | <b>Group A</b><br>MO-logic                     |   |
| <b>Group B</b><br>OM-logic | <b>G1</b>                                      | <b>G2</b>                                       |
|                            | Control value volume<br>Control value valence  | Extremely high volume<br>Control value valence  |
|                            | <b>G3</b>                                      | <b>G4</b>                                       |
|                            | Control value volume<br>Extremely high valence | Extremely high volume<br>Extremely high valence |

Once participants are familiar with the situation, they are randomly assigned to an experimental group. As previously stated, the structure follows a 2x2 factorial between-subjects design as it can be seen in Figure 2. In each group, participants are faced with a product screenshot with either zero (G1), one (G2, G3) or two suspicious cues (G4).

Simultaneously, another randomization regarding question order takes place (Appendix 6): Group A follows a Mediator-Outcome (MO) logic, meaning that the MEDV-question is placed previous to the DV. Group B sorts the questions regarding the Outcome-Mediator (OM) logic and presents the DV-question in the very beginning. The downsides of the MO-logic is the interactive testing effect. Participants are sensitised or primed by the MEDV question regarding their suspicion about the reviews when responding to the DV question of choice intention (Malhotra et al., 2017). The OM-logic instead controls for this bias and increases external validity of the experiment, but it also reduces statistical power as sample sizes per group are smaller than if only one design would have been conducted (Chaudoin, 2020). In existing research, scholars often ignore these details and chose only one path while stating their decision as a limitation. However, the author’s expectations of this experiment are to be able to understand if this effect prevails and if so, to control for it. Consequently, the number of experimental groups is doubled from four to eight.

After being randomly assigned, all participants are confronted with the SERP-screenshot including the product with the manipulated reviews. The following questions are all answered based on this screenshot. Interaction with the product listing, e.g., clicking on the reviews or reading the content, is impossible. This is in line with the experiment’s focus on the heuristic, quantitative cues of reviews during consideration set formation. The effects of the review content on variables like choice intention are not part of this experiment.

For group A (MO) the first questions are two manipulation check (MC) questions to measure the effectiveness of the two manipulations. These two MC questions ask how the participant evaluates the review's volume/rating on a 5-item scale ranging from 1 ("very low") to 5 ("very high"). As the participant has actively considered the review information, which may influence the following questions concerning the impression of the reviews, a distraction question is inserted. It asks to indicate how important the prime badge and the shipping date is. These two features were chosen as they are independent from the product itself (compared to price, design, etc.) and thus most likely to not influence further questions. Subsequently, a short attention check (AC) follows, that asks participants to tick the option "other" and write the word "red".

Afterwards, the questions to measure the MEDV and the DV follow. Questions are set up as stated in the previous chapter. To check whether participants took the reviews into account when answering the questions, a second AC based on memory is inserted. It asks the participant to remember how many reviews the presented product had. There are three randomized options to choose from, which are different enough for an attentive participant to remember the correct one: "179", "3,852" and "136,571" reviews. An additional option "I don't remember" avoids participants guessing the answer.

The experiment closes with demographic questions about gender, age, country of residence, and education level. Thus, these factors can be controlled for and are used to describe the sample. Finally, participants have the opportunity to leave comments and are debriefed.

For participants of group B (OM) everything is kept unchanged, except for the position of the choice intention question (DV). It is inserted in the very beginning, immediately with the product screenshot. Thus, the question is responded without any influence from previous questions. However, it is also important to mention that participants are part of an experiment and might not actually be in the purchase decision process for headphones. This makes it harder for them to answer a choice intention question without any anchor they can base their decision on except for the product picture provided.

During the experiment participants are not able to go back to previous questions preventing them from changing answers. This setup is especially important for the OM-path that requires participants to answer the choice intention question independently. To increase completion rates a progress bar is shown. Further, participants can choose between English and German.

#### 4.1.5 Sampling Method

The study presented above collects responses from participants to draw causal conclusions about the research question. Necessary data is gathered using a non-probability sampling technique, precisely convenience sampling. Thereby, participants are not randomly chosen from the population and, as such, the sample might not be perfectly representative. However, the advantage is that participants can be accessed easily, fast, and inexpensively (Taherdoost, 2016).

Participants are recruited with the help of an online crowdsourcing platform – Prolific – that enables fast and easy access to participants in exchange for a payment (Palan & Schitter, 2018). Only if participants conscientiously take part in a study, without failing any AC, they are paid by the researcher. Thus, the payment serves as an incentive while increasing data quality (Goodman & Paolacci, 2017). Additionally, studies about Prolific in comparison to other platforms confirm this: Prolific’s participants are found to be similarly attentive, but less prone towards cheating and more diverse and naiver than on Amazon Mechanical Turk, a comparable platform (Peer et al., 2017). Further, participants of such platforms are most likely an adequate representation of the internet population, which is an important fact for a study about the online marketplace Amazon (Crump et al., 2013).

Prolific-participants are pre-selected based on five criteria to increase data quality. First, participants are required to shop at least once in a few months online. Additionally, only those living in the countries with the biggest Amazon marketplaces – USA, Germany, or UK (Amazon.com Inc., 2019) – are invited to participate. They should be maximum 60 years old. All these criteria intend to increase the likelihood of participants being familiar with online shopping. Finally, only participants speaking English or German and with a Prolific approval rate of 100% in previous studies are accepted to take part in the study. Participants matching these criteria and passing the first AC<sup>4</sup> are paid 0,29 £.

To estimate the total required sample size, the amount experimental groups (=8) is taken into account. If a study aims to compare groups and measure the differences a minimum sample size of 30 participants per cell is often used as a rule of thumb (Van Voorhis & Morgan, 2007). In an attempt to increase statistical power, the number of participants was increased to 45 per cell.

---

<sup>4</sup> The second AC based on memory is not accepted by Prolific. All participants that fail AC 1 are replaced automatically.

## 4.2 Data Analysis and Results

### 4.2.1 Data Cleaning and Final Sample

The uncleaned sample size ( $n$ ) encompasses 659 participants. However, to examine high quality data some adjustments have to be made. This process of data cleaning and the following analyses were conducted with the statistics software *IBM SPSS Version 26*.

First, all participants that failed both ACs or only AC2 ( $n = 17$ ; 3% of total sample) were eliminated by a listwise deletion, meaning that their whole dataset is removed from the sample. Although this approach decreases statistical power, it increases data quality (Tsikriktsis, 2005). Failing only AC1 is not a deletion criterion as most of the participants who failed AC1 ( $n = 190$ , 29% of total sample) still succeeded in the AC2 ( $n = 181$ ; 95% of AC1 failures). It proves that participants actually paid attention to the reviews, which is a more relevant factor than managing AC1. A possible explanation for the high failure rate of AC1 is that it begins with a question to which suitable answers are presented that encourage to click immediately while distracting from further reading the instructions. For details see Appendix 7.

Further, two cases were dropped from the sample due to univariate outliers detected by using z-scores that outrange the standardized value of 3,29 ( $p < .001$ ; Tabachnick & Fidell, 2013). These are two participants who never bought anything on Amazon.

The final sample size includes 640 participants, 61% females, 38% males and 1% non-binary/without indication. They are on average 36 years old ( $MIN = 20$ ;  $MAX = 60$ ;  $SD = 9.6$ ) and 72% live in the UK, 19% in Germany and 9% in the USA. The majority has a Bachelor's degree (46%; Appendix 8). Most participants shop at least once per month on Amazon (81%; Appendix 9). Further, the 640 participants are equally and mainly randomly distributed between the eight experimental groups (Appendix 10). Consequently, all groups by far exceed the required size of 45 participants (Appendix 11).

### 4.2.2 Scale Reliability and Manipulation Checks

The variable for suspicion is measured on a multi-item scale. Although it was already used in previous studies (e.g., Kirmani & Zhu, 2007; Zhuang et al., 2018) a scale reliability test is conducted to measure if the three items are related. A common approach of examining internal consistency is the calculation of Cronbach's Alpha (CA; Malhotra et al., 2017). The analysis reveals  $\alpha = .934$ , which can be classified as an excellent score (George & Mallery, 2003;

Appendix 12). As the data also proves that no item has to be deleted to increase CA, the items are aggregated by calculating the mean.

Additionally, to examine whether the manipulations of the variables volume and valence worked as intended, two one-way analyses of variance (ANOVA) were conducted with the MC variables for volume and valence as the DVs and volume (IV of CMMM) and valence (MODV) as the IVs. To do so, the required assumptions have to be confirmed: First, the groups are independent as participants were randomly assigned to independent groups. Second, there are no outliers (Chapter 4.2.1). Third, homogeneity of variances is confirmed due to a non-significant Levene's test for volume (the null hypothesis of homogenous variances cannot be rejected). For valence, homogeneity is proven by a significant Welch test, as an alternative to an unsuccessful Levene's test (Appendix 13; Delacre et al., 2017). Fourth, normality of the DVs are examined by a Shapiro-Wilk test (Razali & Wah, 2011), which results in a violation of that assumption (Appendix 13): The MC variables for volume and valence are not significantly normally distributed ( $p < .00$ ). Yet, because the sample size is large enough ( $n > 30$ ), the central limit theorem holds and this violation does not hinder from using an ANOVA (Shirota & Suzuki, 2014). Not only the descriptive statistics, but also the ANOVA results suggest the manipulations were effective (Appendix 14). For volume and valence the control group did evaluate the volume/valence lower ( $M_{\text{Control\_VOL}} = 3.98$ ;  $M_{\text{Control\_VAL}} = 4.13$ ) than the group exposed to an extreme value ( $M_{\text{High\_VOL}} = 4.40$ ;  $M_{\text{High\_VAL}} = 4.49$ ). Further, these differences in the means are found significant at a 5%-level with  $F(1) = 37.68$ ,  $p < .001$  for volume and  $F(1) = 37.72$ ,  $p < .001$  for valence.

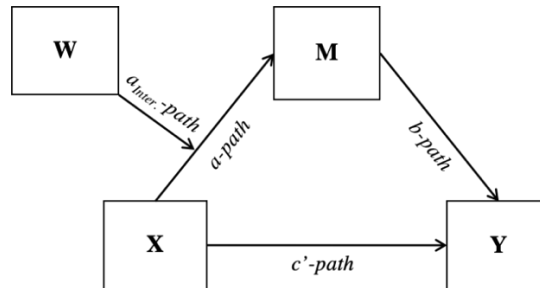
### 4.2.3 Hypotheses Testing

Before the hypotheses are tested, by using an ANOVA it is examined whether the order of the MEDV and DV has an impact on the participants' answers in regards to suspicion and choice intention (Chapter 4.1.4). After checking the assumptions (Appendix 15), the ANOVA-results show that there are no significant differences between the two groups (OM, MO) in regard to suspicion and choice intention (Appendix 16). Therefore, the following analyses will not differentiate between the two groups, but aggregate them instead. Thereby not only effect size is increased but also randomness of all demographic variables is given (education and shopping frequency were not perfectly random for OM/MO; Appendix 10).

To validate or reject the four hypotheses of the CMMM, a statistical analysis using the SPSS macro *PROCESS Version 3.5* was performed. This macro is based on regression-path analyses that uncover moderated mediation effects by estimating the regression coefficients with a

bootstrapping approach (Hayes, 2018). PROCESS offers different models depending on the variables' composition. For this experiment, the appropriate model is number 7 (Figure 3) as it represents the study's research design (Chapter 4.1.1).

**Figure 3: Model 7 (Hayes, 2018)**



The assumptions for this analysis encompass the linearity of the variables and homoscedasticity, independency, and normality of error terms (Hayes, 2018). First, the linearity assumption is fulfilled. For the a-path it is given due to the dummy nature of the IV and MODV – the same is valid for the c'-path. For the b-path linearity is assessed by a visual analysis of scatterplots using LOESS smoothing (Jacoby, 2000). To examine the homoscedasticity assumption, a Breusch-Pagan test is among others an appropriate test, because its null hypothesis states that homoscedasticity prevails (Uyanto, 2019). As the test results are all non-significant with  $p > .05$  the null hypotheses cannot be rejected and thus the assumption is met. Finally, to check for normality of error terms a Durbin-Watson test was used (L'Esperance & Taylor, 1975). Results show that this assumption is fulfilled, as the values are close to 2, which is the target value to confirm independence (Watson & Teelucksingh, 2002). For details see Appendix 17.

For the following hypotheses tests, a 5%-significance-level (95% confidence interval) with 5,000 bootstrap replicates was chosen. Further, the model controls for the variables of age, country, education, gender and Amazon shopping frequency by including them as covariates.<sup>5</sup>

To test H<sub>1</sub> and H<sub>2</sub>, the a-path for a simple moderation is observed. The model is significant at a 5%-level ( $F(8, 631) = 4.05, p < .00$ ). However, the coefficient of determination ( $R^2$ ) is low, indicating that a much variance is left unexplained ( $R^2 = 5.5%$ ). Despite the model-significance, no significant direct effect of volume on suspicion can be found ( $t(631) = 2.42, p = .18$ ). Thus, H<sub>1</sub> is not supported. Interestingly, a significant effect of valence can be uncovered ( $t(631) =$

<sup>5</sup> Results show that none of these variables do have a significant effect within the model. The outputs without the covariates are presented in Appendix 19. However, the results are unchanged.

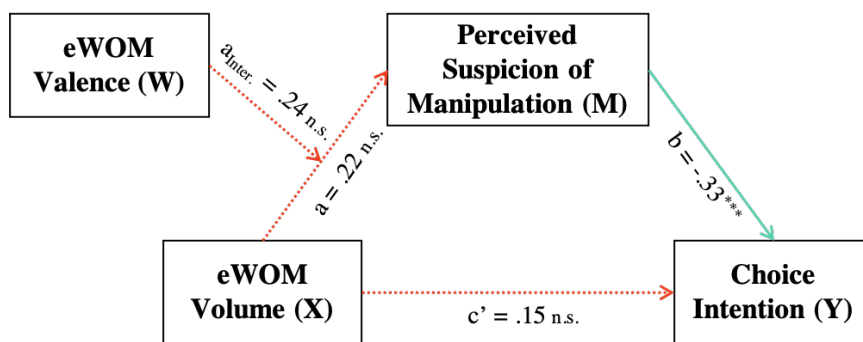
2.42,  $p = .016$ ). Although it is not in the direct focus of this thesis, it supports existing studies which showed that consumers react suspicious on overly positive ratings.

Regarding  $H_2$  the interaction term (VOL\*VAL) is considered to understand the impact of two suspicious cues on perceived suspicion. However, there is no evidence for a significant interaction at the 5%-significance-level ( $t(3631) = 1.01, p = .315$ ), leading to a rejection of  $H_2$ .

To understand the relationship between suspicion and choice ( $H_3$ ), the PROCESS's output data for the b-path is analysed. Model significance is given on a 5%-significance level ( $F(7, 632) = 10.53, p < .00$ ). Similar to the previous model, still only a small amount of variation is explained due to a low  $R^2$ -value ( $R^2 = 10.8\%$ ). Yet, there is a significant, negative relationship between suspicion and choice, as choice intention decreases with increased suspicion ( $b = -.33, t(632) = -8.09, p < .00$ ). Thus,  $H_3$  is supported. The same model can be used to examine the direct effect between volume and choice intention ( $H_4$ ): The data does not give any evidence for a significant, direct effect of the IV (volume) on the DV (choice intention) ( $t(632) = 1.30, p = .20$ ).

Summing it up, the whole CMMM is taken into account and checked for the conditional indirect effects and the index of moderation. This reveals further insights for  $H_2$ : Only if the moderator takes the value 1 (extremely high valence) a significant indirect negative effect of volume on choice through the mediator suspicion is identified as the confidence interval does not include zero (VAL = 1; 95% CI [-.28, -.04]). However, this fact has to be interpreted cautiously as the index of moderated mediation is not significant (Index = -.08, 95% CI [-.25, .08]). Consequently, there are no significant differences between the indirect effects between the different levels of the moderator (VAL = 0 / 1) and there is no sufficient evidence for a moderated mediation. For the summarized results see Figure 4, for the PROCESS output see Appendix 18.

Figure 4: Moderated Mediation Model



\*\*\*  $p < .001$

## 5 Final Discussion

### 5.1 Summary of Research Findings

This dissertation aimed to answer the research question of whether extreme levels of eWOM volume represent a suspicious cue decreasing choice intention in the consideration set formation on Amazon. The theoretical part of this thesis unveiled a research gap which led to four hypotheses that were tested in an experiment.

H<sub>1</sub> proposed to find a relationship between volume and suspicion. However, it was rejected leading to the conclusion that volume by itself is not perceived as a suspicious cue in the given scenario. The reason might be that consumers are already used to and expect high review volumes on Amazon and in this particular category (Chapter 3 & 4.1.2) and therefore do not put attention on and scrutinize these values anymore (Summerfield & Egner, 2009). Further, it cannot be precluded that the research design which only considered two, already quite high levels of volume (3,852 and 136,571) impacted the results. Interestingly, it was found that extreme valence ratings do serve as such a cue. This supports existing research stating that users become sceptical if they recognize a cue that is extremely positive (Maslowska, 2017a; Yang, 2018).

Furthermore, H<sub>2</sub> was not supported by the interaction analysis: The impact of extreme volume on suspicion is not enhanced when paired with extreme valence. As an extreme volume itself is not perceived as a suspicious cue (H<sub>1</sub>) it can be derived that it might have an opposite effect. Instead of serving as an additional suspicious cue to the extreme valence rating and thereby increasing suspicion, it might be rather a supporting cue emphasizing the rating's credibility: The more users give a positive rating, the more correct it is perceived (Salganik & Watts, 2008; Khare et al., 2011). Important to mention are the conditional indirect effects which support H<sub>2</sub>: In the case of extreme high valence, there is a mediation between volume and choice through suspicion. Based on that, but with the non-significant index of moderated mediation in mind, one can only cautiously derive that suspicion is activated in the combined presence of two extreme cues as initially hypothesized.

H<sub>3</sub> was supported, which claimed a negative relation between suspicion and choice intention. The study was able to show this relationship. The interpretation of this result is based on the fact that consumers chose those products that they trust, while those arising suspicion rather deter them from a choice (Shareef, 2013).

Finally, there is no significant evidence supporting H<sub>4</sub>. There might be many different mediators explaining the relation between volume and choice intention. In this study only one, suspicion, was covered, and was not proven to be significant. While providing a first explanation for the non-significance, as the impact of other possible mediator(s) is unknown, the interpretation of this result becomes particularly hard.

Further reasons explaining the outcomes are limitations, which are elaborated in Chapter 5.3.

## **5.2 Academic and Managerial Implications**

This study results in valuable implications for scholars and managers. To start with, the academic implications, the author gave an overview about the current state of research by combining the topics of consumer suspicion, eWOM – in particular eWOM volume – fake reviews and their impact on consumer decisions while taking a heuristic perspective. By doing so, a new, rarely examined research field of extremely high values of volume was unveiled (Appendix 3), which was the focus of the presented experiment. The results neither found evidence for extreme eWOM volume being a suspicious cue (H<sub>1</sub>) nor leading to lower choice intention (H<sub>4</sub>). While puzzling, it is a valuable first contribution to this new research field taking into account the consequences of publication bias, which happen if scholars do not publish insignificant results, thus arising an imbalance in favour of positive results (Dickersin & Min, 1993). Scholars can follow up on these results by replicating them or uncovering further effects by including other explanatory variables than suspicion and using other designs (Chapter 5.3).

Another implication is that certain effects that were already discovered by previous scholars in various contexts could be replicated, which adds to the confidence in their findings. On the one hand, the study was able to support the statement that overly positive ratings make consumers suspicious (Dholakiya, 2014). On the other hand, there is additional evidence for the negative relation between suspicion and choice intention (e.g., Augusto et al, 2019; Zhuang et al., 2018).

Regardless of the thesis' main focus on eWOM, another academic implication regarding the experiment's design can be drawn. Although no differences between the MO- and OM-group were found one cannot derive that this is true for all studies. As still many scholars neither indicate and explain which design was chosen nor control at all for the order of outcome- and mediator-variable (Chaudoin, 2020), the author encourages researchers to think carefully about which design to select and give explanations.

Regarding managers, several recommendations are derived from this paper's findings, too. First, this paper and, in particular, its summary of existing research, emphasizes the risks and consequences of engaging in fake reviews and consequently strongly discourages doing so. As elaborated in the theoretical part, fake reviews are already prevailing in the consumers' minds. They turn the reviews' actual positive effects into negative ones as consumers stop trusting and become sceptical. This fact is underlined by the study's qualitative results: Many participants voluntarily provided comments highlighting the fact that consumers not only lose trust but are even aware of or engage themselves in the tactics firms use (Appendix 20 a-f).

Based on the quantitative results of only this study, sellers on Amazon should not worry too much if any product counts a high review volume. However, this research field is not fully explored and not all effects are uncovered yet. So, the assumed fact of extreme volume being a suspicious cue might still be true. Therefore, the recommendation to managers is to track sales closely and build up a PDP that conveys trust, e.g., by high quality and accurate content. So, consumers who might be sceptical first when seeing the product within the SERP, become convinced at second when reading the details on the PDP. For scholars this also opens a further research field of examining which elements of a PDP actually convey trust.

With respect to managers of shopping platforms like in this case Amazon, it is highly recommended to regain consumer's trust in reviews as it is a valuable asset for e-commerce (Kim et al., 2008). Thus, Amazon should not stop improving systems to detect fake reviews. Furthermore, as a high review volume and valence is not necessarily caused by fake reviews but due to valid reasons, Amazon should find a way to communicate these numbers in a trusting way. A possible measure to increase transparency is to display a badge next to the review section on the SERP indicating that this product's reviews were checked and verified as trustable.

### **5.3 Limitations and Future Research**

The study's results are subject to certain limitations which are presented now, while recommendations for future research will be derived afterwards.

First, to follow up on the academic implications, the study included suspicion as the only mediator. However, as stated earlier, there might be more variables explaining the perception of extreme volume and ultimately its impact on choice intention, such as perceived (dis)trust (Wu et al., 2020), credibility (Zhang et al., 2014), product quality or popularity (De Langhe et al., 2015; Eslami & Ghasemaghaei, 2018). Researchers should integrate them and understand if their effects vary at different levels of volume.

Further, the thesis' scope is limited to the consideration phase within the purchase decision making process. This comes along with two constraints: First, only the heuristic perspective of whether the extreme eWOM metrics work as heuristic cues that make users suspicious and associate it with fake reviews, was examined. Thus, participants were only exposed to the SERP, which just presented the review overview with quantitative numbers, while the PDP with the qualitative review content – which is processed when choosing a systematic instead of a heuristic route (Sinaceur, 2010) – was out of scope. This setup also restricted participants in their evaluation of perceived suspicion to the quantitative review metrics. Based on the comments left by participants, some would have preferred the systematic path and extend the information search by checking the content and investigating it for any suspicious cues before indicating whether they perceive suspiciousness (e.g., Appendix 20 j-k). Additionally, based on the available research the author assumed that heuristic processing takes place in the consideration phase when the user is only faced with the quantitative eWOM metrics. However, no proof is given how each individual processed the information and eventually got to his/her decision.

The second limitation arising from the focus on the consideration phase is that the DV of the experiment was choice intention, meaning the inclusion or exclusion of a certain brand into the consumer's consideration set. Although being part of the consideration set is a relevant step within the user journey for a brand, the actual sales are also important (measured e.g., by purchase intention or the Amazon Sales Rank (Floyd et al., 2014)).

The participants' involvement in the experimental setting, which can take a moderating role for any effects of eWOM (Doh & Hwang, 2009), is another limitation. Participants were not in a real purchase scenario, so imagination about how they would react in a real situation was demanded. Although observing a consumer's intention is a dominant and commonly used predictor for the actual behaviour (Ajzen & Fishbein, 1980; Kim et al., 2018), discrepancies – the so called intention-behaviour gap – cannot be avoided (Sheeran, 2002).

In addition, although the author attempted to increase external validity by using real product listings without manipulating them to a large extent, the setup was still artificial. In reality, users were able to compare options, especially the quantitative metrics. Then an extreme volume might be relativised if all surrounding products have a similar volume level, while it will grab a user's attention if it is the only one by creating a contrast (Florack et al., 2020).

Concluding from these limitations, it is recommended to develop a study that includes the systematic perspective, examines the effects of extreme eWOM volume within the whole purchase decision making process, which also shows other products to compare, and eventually considers

the impact on sales. There are two different approaches to achieve this: First, the experiment should be extended by showing multiple products, each with a different combination of volume and valence values, enabling a comparison. Further, participants should be able to click on the reviews, explore their qualitative content and finally indicate their actual purchase intention. Researchers could actively ask for perceived suspicion of each product or instead indirectly derive results about the effects of the volume-valence-combination based on the product chosen by the participant. Second, to overcome the still existing artificiality of this design an AB-testing on the live website is proposed as it allows to compare the effects on actual sales (e.g., conversion rates) within a realistic scenario. Nevertheless, it is more complex due to multiple extraneous variables and the requirement to manipulate an actual product listing (e.g., adjusting the review volume for one group), which is only possible in cooperation with Amazon.

As heuristic processing was only assumed, future studies should test which processing route is actually taken. A suggested approach is to manipulate cognitive load of the participants, e.g., by requiring them to remember an eight-digit number during the experiment, and observe whether decision making changes (Gilbert & Hixon, 1991). If decisions are unaffected even under increased cognitive load one can derive that not a lot of cognitive resources are required and therefore heuristic processing takes place, otherwise the systematic route can be concluded.

Furthermore, the study focused on Amazon only. Although it is reasonable to use Amazon for a first experiment due to its popularity, market share and especially the prevalence of high-volume levels, generalization for other marketplaces is limited. Additionally, the experiment only included the product category of headphones. However, as discussed in Chapter 2.3.1, eWOM plays a different role depending on the category. Therefore, further research should be extended in regard to other platforms and further product categories.

## **6 Conclusion**

Online reviews have become a crucial element in the e-commerce landscape as they are an additional source of information during the decision making process. Consequently, scholars are increasingly interested in studying the various effects of eWOM. This dissertation has contributed to this important research field by emphasizing the theory of eWOM, its perception, the prevalence and consumers' awareness of fake reviews, and by identifying a relevant research gap regarding the unknown effects of extreme eWOM volume. The author hopes that her findings are a first contribution to close this research gap and other scholars are encouraged to explore further effects of extreme values of eWOM volume.

## References

- Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior* (1st ed.). Prentice-Hall, Inc.
- Almana, A., & Mirza, A. (2013). The Impact of Electronic Word of Mouth on Consumers' Purchasing Decisions. *International Journal of Computer Applications*, 82(9), 23-31. <https://doi.org/10.5120/14145-2286>
- Amazon.com Inc. (2019). *Amazon.com Annual Report 2019*. Retrieved from [https://s2.q4cdn.com/299287126/files/doc\\_financials/2020/ar/2019-Annual-Report.pdf](https://s2.q4cdn.com/299287126/files/doc_financials/2020/ar/2019-Annual-Report.pdf)
- Ambler, N., & Bui, T. (2011). Harnessing the Influence of Social Proof in Online Shopping: The Effect of Electronic Word of Mouth on Sales of Digital Microproducts. *International Journal Of Electronic Commerce*, 16(2), 91-114. <https://doi.org/10.2753/jec1086-4415160205>
- Ante, S. (2009). *Amazon: Turning Consumer Opinions into Gold*. Bloomberg.com. Retrieved 8 November 2020, from <https://www.bloomberg.com/news/articles/2009-10-15/amazon-turning-consumer-opinions-into-gold>.
- Aslam, S., Jadoon, E., Zaman, K., & Gondal, S. (2011). Effect of Word of Mouth on Consumer Buying Behavior. *Mediterranean Journal Of Social Sciences*. <https://doi.org/10.5901/mjss.2011.v2n3p497>
- Augusto, M., Torres, J., & Torres, P. (2019). The Impact of Trust and Electronic Word-Of-Mouth Reviews on Purchasing Intention. *International Journal Of Entrepreneurship And Small Business*, 37(1), 136. <https://doi.org/10.1504/ijesb.2019.10021496>
- Ba, S., & Pavlou, P. (2002). Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly*, 26(3), 243. <https://doi.org/10.2307/4132332>
- Bailey, A. (2005). Consumer Awareness and Use of Product Review Websites. *Journal Of Interactive Advertising*, 6(1), 68-81. <https://doi.org/10.1080/15252019.2005.10722109>
- Banerjee, S., & Chua, A. (2017). Theorizing the Textual Differences between Authentic and Fictitious Reviews. *Internet Research*, 27(2), 321-337. <https://doi.org/10.1108/intr-11-2015-0309>
- Bobko, P., Barelka, A., & Hirshfield, L. (2013). The Construct of State-Level Suspicion: A Model and Research Agenda for Automated and Information Technology (IT) Contexts. *Human Factors: The Journal Of The Human Factors And Ergonomics Society*, 56(3), 489-508. <https://doi.org/10.1177/0018720813497052>
- Bryman, A., & Bell, E. (2011). *Business Research Methods* (3rd ed.). Oxford University Press Inc.
- Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source versus Message Cues in Persuasion. *Journal Of Personality and Social Psychology*, 39(5), 752.

- Charness, G., Gneezy, U., & Kuhn, M. (2012). Experimental Methods: Between-Subject and Within-Subject Design. *Journal Of Economic Behavior & Organization*, 81(1), 1-8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- Chaudoin, S., Gaines, B., & Livny, A. (2020). *Survey Design, Order Effects, and Causal Mediation Analysis*. [Submitted for publication]. University of Illinois.
- Chen, P., Wu, S., & Yoon, J. (2004). The Impact of Online Recommendations and Consumer Feedback on Sales. *The International Conference on Information Systems*. Washington, DC.; ICIS 2004.
- Cheung, C., & Thadani, D. (2012). The Impact of Electronic Word-Of-Mouth Communication: A Literature Analysis and Integrative Model. *Decision Support Systems*, 54(1), 461-470. <https://doi.org/10.1016/j.dss.2012.06.008>
- Cheung, M., Luo, C., Sia, C., & Chen, H. (2009). Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of Online Consumer Recommendations. *International Journal Of Electronic Commerce*, 13(4), 9-38. <https://doi.org/10.2753/jec1086-4415130402>
- Chevalier, J., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal Of Marketing Research*, 43(3), 345-354. <https://doi.org/10.1509/jmkr.43.3.345>
- Chintagunta, P., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box-Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29(5), 944-957. <https://doi.org/10.2139/ssrn.1331124>
- Chyung, S., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement*, 56(10), 15-23. <https://doi.org/10.1002/pfi.21727>
- Clemons, E., Gao, G., & Hitt, L. (2006). When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry. *Journal Of Management Information Systems*, 23(2), 149-171. <https://doi.org/10.2753/mis0742-1222230207>
- Court, D., Elzinga, D., Mulder, S., & Vetvik, O. J. (2009). *The Consumer Decision Journey*. McKinsey Quarterly, 3, 96–107.
- Crump, M., McDonnell, J., & Gureckis, T. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *Plos ONE*, 8(3), e57410. <https://doi.org/10.1371/journal.pone.0057410>
- Cui, G., Lui, H., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal Of Electronic Commerce*, 17(1), 39-58. <https://doi.org/10.2753/jec1086-4415170102>
- Dai, H., Chan, C., & Mogilner, C. (2019). People Rely Less on Consumer Reviews for Experiential than Material Purchases. *Journal Of Consumer Research*, 46(6), 1052-1075. <https://doi.org/10.1093/jcr/ucz042>

- De Langhe, B., Fernbach, P., & Lichtenstein, D. (2015). Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings. *Journal Of Consumer Research*, 42(6), 817-833. <https://doi.org/10.1093/jcr/ucv047>
- DeAndrea, D., Van Der Heide, B., Vendemia, M., & Vang, M. (2018). How People Evaluate Online Reviews. *Communication Research*, 45(5), 719-736. <https://doi.org/https://doi.org/10.1177%2F0093650215573862>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default use Welch's t-test nstead of Student's t-test. *International Review of Social Psychology*, 30(1), 92-101. <https://doi.org/10.5334/irsp.82>
- Demand for Fast and Innovative Delivery Persists amid the Pandemic.* eMarketer. (2020). Retrieved 8 November 2020, from <https://www.emarketer.com/content/demand-for-fast-and-innovative-delivery-persists-amid-the-pandemic>.
- Dholakiya, P. (2014). *Are Excessively Positive Customer Reviews Hurting Your Conversion Rates? [Study]*. Unbounce.com. Retrieved 14 March 2021, from <https://unbounce.com/conversion-rate-optimization/customer-reviews-conversion-rates/>.
- Dickersin, K., & Min, Y. (1993). Publication Bias: The Problem That Won't Go Away. *Analys Of The New York Academy Of Science*, 703(1), 135-148. <https://doi.org/https://doi.org/10.1111/j.1749-6632.1993.tb26343.x>
- Doh, S., & Hwang, J. (2009). How Consumers Evaluate eWOM (Electronic Word-of-Mouth) Messages. *Cyberpsychology & Behavior*, 12(2), 193-197. <https://doi.org/10.1089/cpb.2008.0109>
- Duan, W., Gu, B., & Whinston, A. (2008). The Dynamics of Online Word-Of-Mouth and Product Sales - An Empirical Investigation of the Movie Industry. *Journal Of Retailing*, 84(2), 233-242. <https://doi.org/10.1016/j.jretai.2008.04.005>
- Elwalda, A., & Lü, K. (2016). The Impact of Online Customer Reviews (OCRs) on Customers' Purchase Decision: An Exploration of the Main Dimensions of OCRs. *Journal Of Customer Behaviour*, 15(2), 123-152. <https://doi.org/10.1362/147539216X14594362873695>
- eMarketer. (2019) *eMarketer Report: The Future of Retail of 2019*. p.5. Retrieved 29 January 2020, from <https://on.emarketer.com/Report-2019110-Download.html>
- Eslami, S., & Ghasemaghaei, M. (2018). Effects of Online Review Positiveness and Review Score Inconsistency on Sales: A Comparison by Product Involvement. *Journal Of Retailing And Consumer Services*, 45, 74-80. <https://doi.org/10.1016/j.jretconser.2018.08.003>
- Evans, J., & Mathur, A. (2005). The Value of Online Surveys. *Internet Research*, 15(2), 195-219. <https://doi.org/10.1108/10662240510590360>
- Filieri, R. (2016). What Makes an Online Consumer Review Trustworthy?. *Annals Of Tourism Research*, 58, 46-64. <https://doi.org/10.1016/j.annals.2015.12.019>

- Filieri, R., & McLeay, F. (2013). E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews. *Journal Of Travel Research*, 53(1), 44-57. <https://doi.org/10.1177/0047287513481274>
- Florack, A., Egger, M., & Hübner, R. (2020). When Products Compete for Consumers Attention: How Selective Attention Affects Preferences. *Journal Of Business Research*, 111, 117-127. <https://doi.org/10.1016/j.jbusres.2019.05.009>
- Fretwell, L., Stine, J., Sethi, H., & Noronha, A. (2013). *'Catch and Keep' Digital Shoppers How To Deliver Retail Their Way* [Ebook] (p. 4). Cisco IBSG (Cisco Internet Business Solutions Group). Retrieved 22 October 2020, from [https://www.cisco.com/c/dam/en\\_us/about/ac79/docs/retail/Catch-and-Keep-the-Digital-Shopper\\_PoV.pdf](https://www.cisco.com/c/dam/en_us/about/ac79/docs/retail/Catch-and-Keep-the-Digital-Shopper_PoV.pdf).
- Gani, A. (2015). *Amazon sues 1,000 'fake reviewers'*. the Guardian. Retrieved 9 November 2020, from <https://www.theguardian.com/technology/2015/oct/18/amazon-sues-1000-fake-reviewers>.
- George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference* (11th ed., p. 231). Allyn & Bacon.
- Ghose, A., & Ipeirotis, P. (2011). Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498-1512. <https://doi.org/10.1109/tkde.2010.188>
- Gibbs, S. (2015). *Computer Says No: Amazon Uses AI to Combat Fake Reviews*. The Guardian. Retrieved 9 November 2020, from <https://www.theguardian.com/technology/2015/jun/22/amazon-ai-fake-reviews-star-ratings-astrourfing>.
- Gilbert, D., & Hixon, J. (1991). The Trouble of Thinking: Activation and Application of Stereotypic Beliefs. *Journal Of Personality And Social Psychology*, 60(4), 509-517. <https://doi.org/10.1037/0022-3514.60.4.509>
- Global Consumer Survey 2020 - Drivers of online purchases in Germany*. Statista. (2020). Retrieved 8 November 2020, from <https://www.statista.com/forecasts/998692/drivers-of-online-purchases-in-germany>.
- Goodman, J., & Paolacci, G. (2017). Crowdsourcing Consumer Research. *Journal Of Consumer Research*, 44(1), 196-210. <https://doi.org/10.1093/jcr/ucx047>
- Gopinath, S., Thomas, J., & Krishnamurthi, L. (2014). Investigating the Relationship Between the Content of Online Word of Mouth, Advertising, and Brand Performance. *Marketing Science*, 33(2), 241-258. <https://doi.org/10.1287/mksc.2013.0820>
- Gross, R. (2014). A Theoretical Consumer Decision Making Model: The Influence of Interactivity and Information Overload on Consumers Intent to Purchase Online. *International Journal Of Business Management & Economic Research*, 5(4), 64-70.
- Gu, B., Park, J., & Konana, P. (2012). Research Note—The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1), 182-196. <https://doi.org/10.1287/isre.1100.0343>

- Hao, Y., Ye, Q., Li, Y. & Cheng, Z. (2010). How Does the Valence of Online Consumer Reviews Matter in Consumer Decision Making? Differences Between Search Goods and Experience Goods. *Proceedings of the 43rd Hawaii International Conference on System Sciences*. 10.1109/HICSS.2010.455
- Hayes, A. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis - A Regression-Based Approach* (2nd ed., pp. 68-73). The Guilford Press.
- Hong, I., & Cha, H. (2013). The Mediating Role of Consumer Trust in an Online Merchant in Predicting Purchase Intention. *International Journal Of Information Management*, 33(6), 927-939. <https://doi.org/10.1016/j.ijinfomgt.2013.08.007>
- Hu, N., Bose, I., Koh, N., & Liu, L. (2012). Manipulation of Online Reviews: An Analysis of Ratings, Readability and Sentiments. *Decision Support Systems*, 52(3), 674-684. <https://doi.org/10.1016/j.dss.2011.11.002>
- Hu, N., Koh, N., & Reddy, S. (2014). Ratings Lead You to the Product, Reviews Help You Clinch it? The Mediating Role of Online Review Sentiments on Product Sales. *Decision Support Systems*, 57, 42-53. <https://doi.org/10.1016/j.dss.2013.07.009>
- Hu, N., Liu, L., & Zhang, J. (2008). Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects. *Information Technology and Management*, 9(3), 201-214. <https://doi.org/10.1007/s10799-008-0041-2>
- Idealo. (2020). *E-Commerce Trends 2020*. Retrieved 8 November 2020, from: <https://partner.idealo.com/de/e-commerce-trends-2020-whitepaper/>
- Jacoby, W. (2021). Loess: A Nonparametric, Graphical Tool for Depicting Relationships Between Variables. *Electoral Studies*, 19, 577-613. [https://doi.org/https://doi.org/10.1016/S0261-3794\(99\)00028-1](https://doi.org/https://doi.org/10.1016/S0261-3794(99)00028-1)
- Jin Ma, Y., & Lee, H. (2014). Consumer Responses Toward Online Review Manipulation. *Journal Of Research In Interactive Marketing*, 8(3), 224-244. <https://doi.org/10.1108/jrim-04-2013-0022>
- Jourdan, P. (2001). Search Or Experience Products: An Empirical Investigation of Services, Durable and Non-Durable Goods. *Asia Pacific Advances In Consumer Research*, 4, 167-174.
- Kahnemann, D., & Frederick, S. (2004). Attribute Substitution in Intuitive Judgement. In H. Simon, M. Augier & G. James, *Models of a man: Essays in memory of Herbert A. Simon* (pp. 411-432). MIT press.
- Kahnemann, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291. <https://doi.org/https://doi.org/10.2307/1914185>
- Kats, R. (2018). *Surprise! Most Consumers Look at Reviews Before a Purchase*. eMarketer. Retrieved 23 October 2020, from <https://www.emarketer.com/content/surprise-most-consumers-look-at-reviews-before-a-purchase>

- Kaziukenas, J. (2019). *Amazon Replaces "Reviews" with "Ratings"*. Marketplace Pulse. Retrieved 9 November 2020, from <https://www.marketplacepulse.com/articles/amazon-replaces-reviews-with-ratings>.
- Khammash, M. (2008). Electronic Word-Of-Mouth: Antecedents of Reading Customer Reviews in Online Opinion Platforms: A Quantitative Study from the UK Market. In *IADIS International Conference WWW/Internet 2008* (p. 81). Freiburg, Germany.
- Khare, A., Labrecque, L., & Asare, A. (2011). The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings. *Journal Of Retailing*, 87(1), 111-126. <https://doi.org/10.1016/j.jretai.2011.01.005>
- Kim, D., Ferrin, D., & Rao, H. (2008). A Trust-Based Consumer Decision-Making Model in Electronic Commerce: The Role of Trust, Perceived Risk, and their Antecedents. *Decision Support Systems*, 44, 544-564. <https://doi.org/10.1016/j.dss.2007.07.001>
- Kirmani, A., & Zhu, R. (2007). Vigilant against Manipulation: The Effect of Regulatory Focus on the Use of Persuasion Knowledge. *Journal Of Marketing Research*, 44(4), 688-701. <https://doi.org/10.1509/jmkr.44.4.688>
- Kordrostami, E., & Rahmani, V. (2020). Investigating Conflicting Online Review Information: Evidence from Amazon.com. *Journal Of Retailing and Consumer Services*, 55, 102125. <https://doi.org/10.1016/j.jretconser.2020.102125>
- Kostyra, D., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the Effects of Online Customer Reviews on Brand, Price, and Product Attributes. *International Journal Of Research In Marketing*, 33(1), 11-26. <https://doi.org/10.1016/j.ijresmar.2014.12.004>
- Kotler, P., Armstrong, G., Saunders, J., & Wong, V. (2008). *Principles of Marketing* (5th ed., pp. 265-271). Pearson Education Limited.
- Kotler, P., Keller, K., Brady, M., Goodman, M., & Hansen, T. (2009). *Marketing Management* (13th ed., p. 248). Pearson Education Limited.
- Kozinets, R., de Valck, K., Wojnicki, A., & Wilner, S. (2010). Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities. *Journal Of Marketing*, 74(2), 71-89. <https://doi.org/10.1509/jmkg.74.2.71>
- L'Esperance, W., & Taylor, D. (1975). The Power of Four Tests of Autocorrelation in the Linear Regression Model. *Journal Of Econometrics*, 3(1), 1-21. [https://doi.org/10.1016/0304-4076\(75\)90062-7](https://doi.org/10.1016/0304-4076(75)90062-7)
- Lee, D., & Murphy, H. (2020). *Facebook Groups Trading Fake Amazon Reviews Remain Rampant*. Ft.com. Retrieved 9 November 2020, from <https://www.ft.com/content/d4af6504-924e-4f94-b82e-0f02671faa12>.
- Li, K., Chen, Y., & Zhang, L. (2020). Exploring the Influence of Online Reviews and Motivating Factors on Sales: A Meta-Analytic Study and the Moderating Role of Product Category. *Journal Of Retailing and Consumer Services*, 55, 102107. <https://doi.org/10.1016/j.jretconser.2020.102107>

- Li, X., Wu, C., & Mai, F. (2019). The Effect of Online Reviews on Product Sales: A Joint Sentiment-topic Analysis. *Information & Management*, 56(2), 172-184. <https://doi.org/10.1016/j.im.2018.04.007>
- Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal Of Marketing*, 70(3), 74-89. <https://doi.org/10.1509/jmkg.70.3.074>
- MacKinnon, D. (2011). Integrating Mediators and Moderators in Research Design. *Research On Social Work Practice*, 21(6), 675-681. <https://doi.org/10.1177/1049731511414148>
- Magnani, M. (2020). The Economic and Behavioral Consequences of Online User Reviews. *Journal Of Economic Surveys*, 34(2), 263-292. <https://doi.org/10.1111/joes.12357>
- Malhotra, N., Nunan, D., & Birks, D. (2017). *Marketing Research - An Applied Approach* (5th ed., pp. 302-360). Pearson Education Limited.
- Maslowska, E., Malthouse, E., & Bernritter, S. (2017a). Too Good to be True: The Role of Online Reviews' Features in Probability to Buy. *International Journal Of Advertising*, 36(1), 142-163. <https://doi.org/10.1080/02650487.2016.1195622>
- Maslowska, E., Malthouse, E., & Viswanathan, V. (2017b). Do Customer Reviews Drive Purchase Decisions? The Moderating Roles of Review Exposure and Price. *Decision Support Systems*, 98, 1-9. <https://doi.org/10.1016/j.dss.2017.03.010>
- Mathews Hunt, K. (2015). Gaming the System: Fake Online Reviews v. Consumer Law. *Computer Law & Security Review*, 31(1), 3-25. <https://doi.org/10.1016/j.clsr.2014.11.003>
- Mavlanova, T., Benbunan-Fich, R., & Koufaris, M. (2012). Signaling Theory and Information Asymmetry in Online Commerce. *Journal Of Information & Management*, 49(5), 240-247. <https://doi.org/10.1016/j.im.2012.05.004>
- Mudambi, S., & Schuff, D. (2010). What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, 34(1), 185-200.
- Murphy, R. (2018). *Local Consumer Review Survey 2018*. BrightLocal. Retrieved 8 November 2020, from <https://www.brightlocal.com/research/local-consumer-review-survey-2018/#Q7>.
- Nelson, P. (1974). Advertising as Information. *Journal Of Political Economy*, 82(4), 729-754.
- Palan, S., & Schitter, C. (2018). Prolific.ac—A Subject Pool for Online Experiments. *Journal Of Behavioral And Experimental Finance*, 17, 22-27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Park, D., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal Of Electronic Commerce*, 11(4), 125-148. <https://doi.org/10.2753/jec1086-4415110405>

- Park, D.H. & Lee, J.(2008). eWOM Overload and its Effect on Consumer Behavioural Intention Depending on Consumer Involvement. *Electronic Commerce Research and Applications*, 7(4): 386-398. <http://doi.org/10.1016/j.elerap.2007.11.004>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research. *Journal Of Experimental Social Psychology*, 70, 153-163. <https://doi.org/https://doi.org/10.1016/j.jesp.2017.01.006>.
- Petrescu, M., O'Leary, K., Goldring, D., & Mrad, S. (2018). Incentivized Reviews: Promising the Moon for a Few Stars. *Journal Of Retailing And Consumer Services*, 41, 288-295. <https://doi.org/https://doi.org/10.1016/j.jretconser.2017.04.005>
- Ramadani, V., Demiri, A., & Demiri, S. (2014). Social Media Channels: The Factors that Influence the Behavioural Intention of Customers. *International Journal Of Business and Globalisation*, 12(3), 297. <https://doi.org/10.1504/ijbg.2014.060214>
- Razali, N., & Wah, Y. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal Of Statistical Modeling And Analytics*, 2(1), 21-33.
- Rosario, A. B., Sotgiu, F., De Valck, K., & Bijmolt, T. (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(3), 297-318. <https://doi.org/10.1509/jmr.14.0380>
- Salganik, M., & Watts, D. (2008). Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market. *Social Psychology Quarterly*, 71(4), 338-355. <https://doi.org/10.1177/019027250807100404>
- Schiffer, Z. (2020). *Amazon is Trying to Crack Down on Fraudulent Reviews. They're Thriving in Facebook Groups*. The Verge. Retrieved 9 November 2020, from <https://www.theverge.com/2020/10/2/21497416/amazon-crack-down-fraudulent-reviews-facebook-wechat-groups>.
- Schoolov, K. (2020). *Amazon is filled with fake reviews and it's getting harder to spot them*. CNBC. Retrieved 9 November 2020, from <https://www.cnbc.com/2020/09/06/amazon-reviews-thousands-are-fake-heres-how-to-spot-them.html>.
- Shareef, M. (2013). Online Buying Behavior and Perceived Trustworthiness. *British Journal of Applied Science & Technology*, 3(4), 662-683. <https://doi.org/10.9734/bjast/2013/2394>
- Sheeran, P. (2002). Intention - Behavior Relations: A Conceptual and Empirical Review. *European Review Of Social Psychology*, 12(1), 1-36. <https://doi.org/10.1080/14792772143000003>
- Shirota, Y., & Suzuki, S. (2014). Visualization of the Central Limit Theorem and 95 Percent Confidence Intervals. *Gakushuin Economics Papers*, 50(4), 125-136.
- Sinaceur, M. (2010). Suspending Judgment to Create Value: Suspicion and Trust in Negotiation. *Journal Of Experimental Social Psychology*, 46(3), 543-550. <https://doi.org/10.1016/j.jesp.2009.11.002>

- Summerfield, C., & Egner, T. (2009). Expectation (and Attention) in Visual Cognition. *Trends In Cognitive Sciences*, 13(9), 403-409. <https://doi.org/10.1016/j.tics.2009.06.003>
- Sun, M. (2012). How Does the Variance of Product Ratings Matter?. *Management Science*, 58(4), 696-707. <https://doi.org/10.1287/mnsc.1110.1458>
- Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics* (6th ed., p. 73). Pearson.
- Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3205035>
- Tata, S., Prashar, S., & Gupta, S. (2020). An Examination of the Role of Review Valence and Review Source in Varying Consumption Contexts on Purchase Decision. *Journal Of Retailing And Consumer Services*, 52, 1-10. <https://doi.org/10.1016/j.jretconser.2019.01.003>
- Thakur, R., Hale, D., & Summey, J. (2018). What Motivates Consumers to Take Part in Cyber Shilling?. *Journal Of Marketing Theory And Practice*, 26(1-2), 181-195. <https://doi.org/10.1080/10696679.2017.1389236>
- Tinuiti (2020). *The 2020 Amazon Consumer Shopping Survey*. Retrieved 8 November 2020, from: <https://tinuiti.com/content/guides/2020-amazon-shopper-survey/>
- Traylor, M. B. (1981). Product Involvement and Brand Commitment. *Journal Of Advertising Research*, 21(6), 51-56.
- Tsikriktsis, N. (2005). A Review of Techniques for Treating Missing Data in OM Survey Research. *Journal Of Operations Management*, 24, 53-62.
- Uyanto, S. (2019). Monte Carlo Power Comparison of Seven Most Commonly Used Heteroscedasticity Tests. *Communications In Statistics - Simulation And Computation*, 1-19. <https://doi.org/https://doi.org/10.1080/03610918.2019.1692031>
- Van Boven, L., & Gilovich, T. (2003). To Do or to Have? That is the Question. *Journal Of Personality And Social Psychology*, 85(6), 1192-1202.
- Van Voorhis, C., & Morgan, B. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials In Quantitative Methods For Psychology*, 3(2), 43-50. <https://doi.org/10.20982/tqmp.03.2.p043>
- Wang, C., Li, Y., Luo, X., Ma, Q., Fu, W., & Fu, H. (2018a). The Effects of Money on Fake Rating Behavior in E-Commerce: Electrophysiological Time Course Evidence From Consumers. *Frontiers In Neuroscience*, 12, 1-9. <https://doi.org/10.3389/fnins.2018.00156>
- Wang, Y., Wang, J., & Yao, T. (2018b). What Makes a Helpful Online Review? A Meta-Analysis of Review Characteristics. *Electronic Commerce Research*, 19(2), 257-284. <https://doi.org/10.1007/s10660-018-9310-2>

- Watson, P., & Teelucksingh, S. (2002). *A Practical Introduction to Econometric Methods: Classical and Modern* (1st ed., p. 72). University of the West Indies Press.
- Wilcox, R. (2001). *Fundamentals of Modern Statistical Methods - Substantially Improving Power and Accuracy* (1st ed., p. 70). Springer.
- Wu, A., & Zumbo, B. (2007). Understanding and Using Mediators and Moderators. *Social Indicators Research*, 87(3), 367-392. <https://doi.org/10.1007/s11205-007-9143-1>
- Wu, Y., Ngai, E., Wu, P., & Wu, C. (2020). Fake Online Reviews: Literature Review, Synthesis, and Directions for Future Research. *Decision Support Systems*, 132, 1-15. 113280. <https://doi.org/10.1016/j.dss.2020.113280>
- Yang, L. (2018). More Is Less: Only Moderate Polarized Online Product Reviews can Affect Sales. *International Journal Of Business And Management*, 13(4), 192-200. <https://doi.org/10.5539/ijbm.v13n4p192>
- Zhang, L., Ma, B., & Cartwright, D. (2013). The Impact of Online User Reviews on Cameras Sales. *European Journal Of Marketing*, 47(7), 1115-1128. <https://doi.org/10.1108/03090561311324237>
- Zhang, X., & Barnes, C. (2019). The Suspicious Factors in Electronic Word-of-Mouth Communication. *Journal Of Marketing Development And Competitiveness*, 13(2), 101-115. <https://doi.org/10.33423/jmdc.v13i2.2015>
- Zhang, X., Ko, M., & Carpenter, D. (2016). Development of a Scale to Measure Skepticism toward Electronic Word-Of-Mouth. *Computers In Human Behaviour*, 56, 198-208. <https://doi.org/http://dx.doi.org/10.1016/j.chb.2015.11.042>
- Zhang, X., Wu, Y., & Li, Y. (2019). The Tendency of Trust in A Distrustful Environment: The Mediation Role of Contextual Perceptions in eWOM. *Journal Of Marketing Development And Competitiveness*, 13(5), 45-63. <https://doi.org/10.33423/jmdc.v13i5.2641>
- Zhao, C., & Zhang, Y. (2020). Impacts of Uncertain Online Reviews on Pricing and Profits of Competitive Retailers. *Kybernetes*. <https://doi.org/10.1108/k-05-2019-0308>
- Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling Consumer Learning from Online Product Reviews. *Journal Of Marketing Science*, 32(1), 153-169. <https://doi.org/10.1287/mksc.1120.0755>
- Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal Of Marketing*, 74(2), 133-148. <https://doi.org/10.1509/jmkg.74.2.133>
- Zhuang, M., Cui, G., & Peng, L. (2018). Manufactured Opinions: The Effect of Manipulating Online Product Reviews. *Journal Of Business Research*, 87, 24-35. <https://doi.org/10.1016/j.jbusres.2018.02.016>

## **Appendix**

|   |         |
|---|---------|
| <b>Appendix 1:</b> Transcript of the Interview with Filip Egert.....                    | XVIII   |
| <b>Appendix 2:</b> Crawler Data Amazon.com .....  | XX      |
| <b>Appendix 3:</b> Overview of Review Volume Levels Used in Past Research .....         | XXI     |
| <b>Appendix 4:</b> Review Presentation on Amazon.com.....                               | XXII    |
| <b>Appendix 5:</b> Questionnaire .....  | XXIII   |
| <b>Appendix 6:</b> Survey Flow.....   | XXVII   |
| <b>Appendix 7:</b> Attention Check Analysis.....  | XXVIII  |
| <b>Appendix 8:</b> Demographics.....  | XXIX    |
| <b>Appendix 9:</b> Amazon Shopping Frequency .....                                      | XXX     |
| <b>Appendix 10:</b> Tests of Randomness of Experimental Groups .....                    | XXXI    |
| <b>Appendix 11:</b> Sample Size and Description of Experimental Groups .....            | XXXIII  |
| <b>Appendix 12:</b> Scale Reliability Analysis .....                                    | XXXIV   |
| <b>Appendix 13:</b> Manipulation Check Analysis: Check of Assumptions.....              | XXXV    |
| <b>Appendix 14:</b> Manipulation Check Analysis: Descriptive Statistics and ANOVA ..... | XXXVI   |
| <b>Appendix 15:</b> Question Order Analysis: Check of Assumptions.....                  | XXXVII  |
| <b>Appendix 16:</b> Question Order Analysis: Descriptive Statistics and ANOVA .....     | XXXVIII |
| <b>Appendix 17:</b> CMMM Analysis: Check of Assumptions .....                           | XL      |
| <b>Appendix 18:</b> PROCESS Output (with Covariates).....                               | XLII    |
| <b>Appendix 19:</b> PROCESS Output (without Covariates) .....                           | XLIV    |
| <b>Appendix 20:</b> Qualitative Study Results (Extract).....                            | XLVI    |

## Appendix 1: Transcript of the Interview with Filip Egert

---

**Interviewee**    **Filip Egert (FE)**  
Remazing GmbH, Hamburg (Amazon Agency)  
Partner

**Date**            02<sup>nd</sup> November 2020

---

- ID:** *Hi Filip, thanks a lot for your time. As you know, I am currently working on my master thesis which is about reviews on Amazon. I thought I could get some insights from you.*
- FE:** *Hi Inna, yes sure!*
- ID:** *Before I start may you please give a brief introduction about you and Remazing?*
- FE:** *Of course. Let me start with Remazing. Remazing is a 360° Amazon agency. We are offering all services that a client needs to be successful on Amazon from advertising to content or overall product strategy. I am partner and work at Remazing since the very beginning. My focus are seller clients and those who are using a hybrid approach of a vendor and seller account.*
- ID:** *Thanks for the introduction. I guess you are the perfect person to ask all my questions. So my first question is what is your impression about the review volumes on Amazon? It seems like volumes have exploded for some product categories.*
- FE:** *Yes I've the same impression, but I think it is not unreasonable. Bestseller products often show many reviews. But there are also many other products with many thousand reviews. I think, you observed it because Amazon's algorithm give products with more and better reviews more visibility within the SERP.*
- ID:** *Makes sense, so do you think it is important to have so many reviews to be successful on Amazon?*
- FE:** *Reviews in general are very relevant. I would say if you are launching a product on Amazon without and you cannot make it to get at least a few reviews soon, you almost have no chance. Competition is just too high.*
- ID:** *But how should I get these reviews? Isn't it super hard, especially in regard to the fact that my product will not be shown on the first page of the SERP?*
- FE:** *It is. But Amazon offers two programs that help sellers, but also vendors, to generate reviews. One is called Amazon Vine, the other is Amazon Early Reviewer Program. They both work similar and focus only on the very first reviews. That means, if your product has already many reviews, you cannot apply for these programs. It works in that way, that Amazon selects users that should write a review. In return they will get the product for free. The users are selected based on criteria like, how many reviews did they write already, were they helpful etc.*
- ID:** *That's interesting. But do they need to write a positive review then? Because a negative one might not really help the seller.*
- FE:** *No, there is no regulation on that. The user should give a honest opinion, even if it means that the rating will just be 1 star. This is a clear intention from Amazon, because otherwise other consumers won't get the real evaluation of the product. But*

*yes, you are right, this might not be useful then for the seller. But if the product is bad, he rather should think of improving it. But I also have to say, based on experience, usually the reviews are quite good. So if your product is not totally disastrous, it most likely will help you. Also, the reviews are labelled if they were written as part of one program. This will increase credibility.*

**ID:** *Ok, got it. But are these the only measures to get reviews or is there anything else? I mean, it's nice to have 5 reviews, but what if my competitor has 50?*

**FE:** *It's true that you are worse off then – just in terms of reviews. Another thing you can do which is still legal is to build variations of your products.*

**ID:** *What does that mean?*

**FE:** *For example, you have a product that is available in many different styles, colours, sizes etc. Instead of having one PDP per product you can combine all of them under one PDP. The advantage is that the reviews are cumulated. So if each product only has 5 reviews, but you have 5 different styles the rating count on the PDP shows 25. This is especially helpful, when you have some products that you are selling already for a long time and a new product. The “old” product has already many reviews which it has collected over the years. The “new” product does not have any review so far, but can benefit from the old ones. Of course, it still needs to make sense to combine them in one PDP!*

**ID:** *That's a smart idea. You said this is a legal approach. I heard that there are many ways out there to get fake reviews. Do you know anything about that?*

**FE:** *Yes, indeed. It's super bad because one day the whole system might collapse if no one can trust reviews anymore. As far as I know, you can get reviews by asking random people to write a positive review for a money refund. It is often organized with Facebook or WhatsApp Groups. You buy the product, write a positive review and then get the money via Paypal. There are many thousand people who are attracted by the free products. Here, other than for the official Amazon programs, the user has to write a positive review. If not, he or she won't get the money.*

**ID:** *Yes this is what I also heard about. But is Amazon doing anything against it?*

**FE:** *Yes, they are often deleting many reviews that seem to be fake. They have technologies that try to detect fake reviews. However, it is not always easy to spot them.*

**ID:** *Ok, I see the issue. Well, I think for now I got all information I need. Thanks a lot Filip!*

**FE:** *Sure, tell me if you need anything else. Bye!*

## Appendix 2: Crawler Data Amazon.com

---

**Search Term: *Bluetooth Headphones***

---

| <b>ASIN<sup>1</sup><br/>Count</b> | <b>Average<br/>Rating</b> | <b>SD<br/>Rating</b> | <b>Average<br/>Price</b>    | <b>SD Price</b>             |
|-----------------------------------|---------------------------|----------------------|-----------------------------|-----------------------------|
| 129 <sup>2</sup>                  | 4.3                       | 0.7                  | \$ 41.53                    | \$ 53.97                    |
| <b>Average<br/>VOL</b>            | <b>Median<br/>VOL</b>     | <b>SD<br/>VOL</b>    | <b>Max. VOL<br/>(#ASIN)</b> | <b>Min. VOL<br/>(#ASIN)</b> |
| 14,988                            | 3,852                     | 33,668               | 224,998 (1)                 | 0 (3)                       |

---

Data captured on 03/12/2020.

<sup>1</sup> ASIN: Amazon Standard Identification Number, a unique identifier for products listed on Amazon marketplaces.

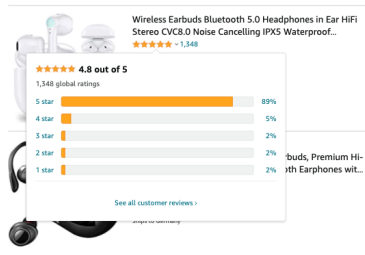
<sup>2</sup> First 5 SERPs were considered.

### Appendix 3: Overview of Review Volume Levels Used in Past Research

| Study                       | Indication of max. review volume considered | Study                   | Indication of max. review volume considered |
|-----------------------------|---|-------------------------|---|
| Amblee & Bui, 2011          | $M = 1.83$<br>$SD = 2.13$                   | Maslowska et al., 2017b | $M = 39.64$<br>$SD = 86.77$                 |
| Chen et al., 2004           | $M = 96.01$<br>$SD = 279.46$                | Park & Lee, 2008        | 27  |
| Chevalier & Mayzlin, 2006   | $M = 68.31$<br>$SD = 205.42$                | Park et al., 2007       | 6   |
| Khare et al., 2011          | 3,470                                       | Zhang et al., 2013      | $M = 48.86$ *                               |
| Kordrostami & Rahmani, 2020 | 420   | Zhao et al., 2013       | $M = 18.07$<br>$SD = 21.65$                 |
| Kostyra et al., 2016        | 200   | Zhu & Zhang, 2010       | $M = 10.30$<br>$SD = 16.29$                 |
| Liu, 2006                   | $M = 303.4$ *                               |                         |   |
| Maslowska et al., 2017a     | 6   |                         |   |

# Appendix 4: Review Presentation on Amazon.com

## Review Overview on SERP



## Review Overview on PDP (Top)

Wireless Earbuds Bluetooth 5.0 Headphones in Ear HIFI Stereo CVC8.0 Noise Cancelling IPX5 Waterproof Headsets with Fast Charging Case for Apple AirPods Pro iPhone/Android Ear Buds

Brand: Soundcore  
★★★★★ 4.8 out of 5 (1,348 ratings) | 9 answered questions

Price: \$38.99 + \$18.53 Shipping & Import Fees Deposit to Germany Details

Color: White  
Sound: Wireless  
Connections: Wireless  
Color: White  
Headphones Form Factor: In Ear  
Noise Control: Active Noise Cancellation

**About this item**

- **[THE LATEST BLUETOOTH 5.0 & 3D STEREO SOUND]** Our wireless headphones adopt the latest Bluetooth 5.0 and antenna technology to provide fast and stable transmission in 3D/3D stereo sound and add Bluetooth signal guarantees whether you're playing games, making phone calls. Striving to help you experience higher quality sound and extraordinary stereo effects.
- **[FOOT-TO-TOE PAIRING & WIRE COMPATIBILITY]** Take out the charging case automatically start the automatic pairing. The smart phone system will pop up a window to connect to the phone to complete the pairing. Easy set up for all your cellphones devices and compatible with all Bluetooth devices/speakers a variety of mobile devices.
- **[SMART TOUCH CONTROL]** One touch-control button supports a lot of functions. Hands-free hold the phone in your hands, very comfortable and convenient. Thanks to the intelligent touch design, there is no need to apply

## Review Section on PDP (Bottom)

**Customer reviews**  
★★★★★ 4.8 out of 5  
1,348 global ratings

|        |     |
|--------|-----|
| 5 star | 89% |
| 4 star | 5%  |
| 3 star | 2%  |
| 2 star | 2%  |
| 1 star | 2%  |

How are ratings calculated?

**Review this product**  
Share your thoughts with other customers

Write a customer review

10% of images

Most recent

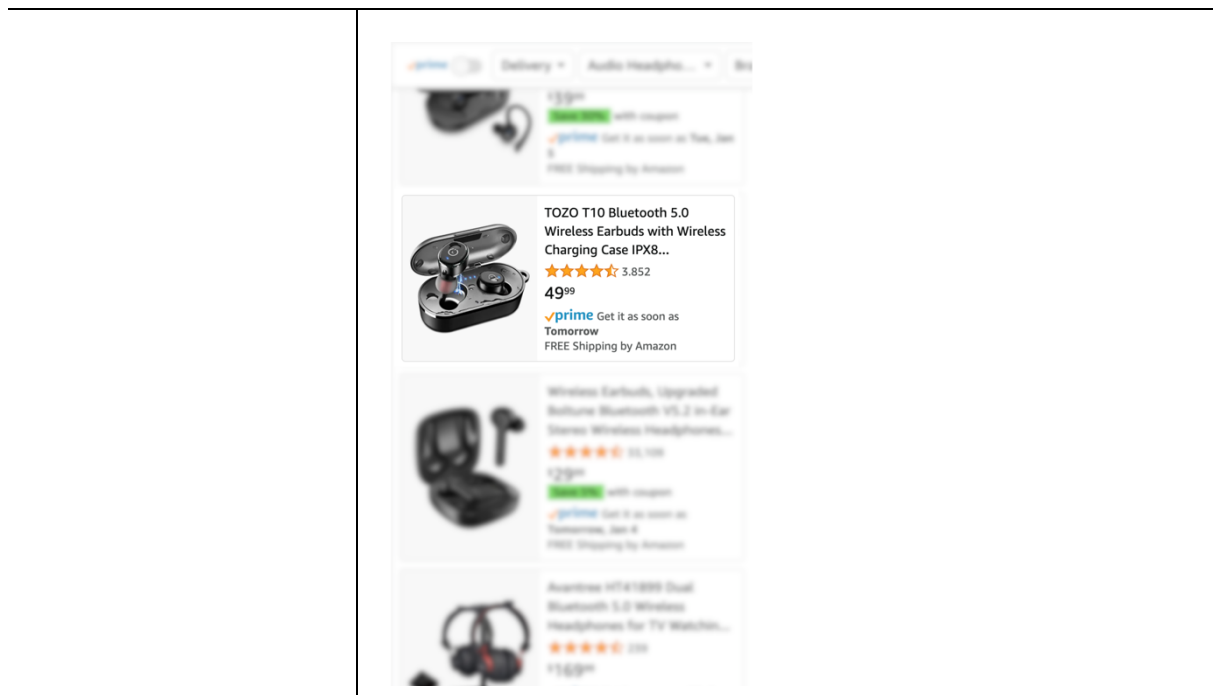
Jon Freeman  
★★★★★ Returning these as unusable  
Reviewed in the United Kingdom on November 19, 2020  
Verified Purchase  
I was very disappointed in these and cannot grasp how they have a 5\* rating. Granted, they pair easily. But it is impossible to control the volume. The sound quality is nasty. The instructions are a joke, and if I hadn't been able to understand the ones in german I would never have figured out who volume control is meant to work. Unfortunately it still didn't. You cannot touch them when in use without unintended effects, so you can't move them to get ear comfort.  
Report abuse

M Darren Pritchett  
★★★★★ Great little boy  
Reviewed in the United Kingdom on November 8, 2020  
Verified Purchase  
Nice compact box and the case is sturdy enough.  
They fit into the ear snug enough and feel comfortable. The Bluetooth connects easily.  
The sound is decent enough for the price you pay. Why spend a fortune on air pods when you could end up losing them. The volume goes high enough that it doesn't blow your ear drums off.  
Great little boy

## Appendix 5: Questionnaire

Example for **Group 1** (Control volume, Control valence);  
Screenshots are adapted according to the other groups

| Coding  | Question  |
|---|---|
| <b>Intro</b>  | <p>Dear Participant,</p> <p>Thanks for taking part in my <b>2-minute-survey about shopping on Amazon</b> as part of my Master Thesis at Católica Lisbon SBE.</p> <p>All answers will be kept strictly anonymous and confidential, so I kindly ask you to answer as <b>honestly as possible</b>. The collected data will exclusively be used for research purposes.</p> <p>If you do have any questions or doubts, please do not hesitate to reach out:</p> <p>Inna Dickmänken (152119122@alunos.lisboa.ucp.pt)</p>  |
| Page Break  |   |
| <b>FREQ</b>   | <p>How often do you shop on Amazon?</p> <ul style="list-style-type: none"> <li><input type="radio"/> Every week (<i>redirected to HF_Intro</i>)</li> <li><input type="radio"/> Several times per month (<i>redirected to HF_Intro</i>)</li> <li><input type="radio"/> Once per month (<i>redirected to HF_Intro</i>)</li> <li><input type="radio"/> Several times per year (<i>redirected to LW_Intro</i>)</li> <li><input type="radio"/> Once per year (<i>redirected to LW_Intro</i>)</li> <li><input type="radio"/> Never (<i>redirected to LW_Intro</i>)</li> </ul> |
| Page Break  |   |
| <p><b>HF_Intro</b></p> <p><i>HF = High Frequency: every week, several times per month, once per month</i></p>   | <p><b>Imagine</b> you are looking on Amazon for new headphones.</p> <p>You are <b>not ready to purchase yet</b>, but you are currently <b>scrolling through Amazon, comparing</b> the options and <b>thinking about which one to click on</b> and to consider in more detail.</p>   |
| <p><b>LF_Intro</b></p> <p><i>LF = Low Frequency: several times per year, once a year, never</i></p>   | <p>Even if you only rarely or never shop on Amazon, for this experiment <b>imagine</b> you are looking on Amazon for new headphones.</p> <p>Imagine you are <b>not ready to purchase yet</b>, but you are currently <b>scrolling through Amazon, comparing</b> the options and <b>thinking about which one to click on</b> and to consider in more detail.</p>  |
| Page Break  |   |
| <p><b>G1_CC_Intro</b><br/><i>IV &amp; MODV</i></p> <p><i>G1 = Group 1</i><br/><i>MO = M</i><br/><i>CC = indication of manipulation (Control, Control)</i></p> | <p>While scrolling through Amazon you found these <b>headphones</b>. Take a second to <b>evaluate</b> the product based on the information given.</p>   |



**G1\_OM\_CC\_CHOICE**  
*DV*  
*Question position for OM-group*

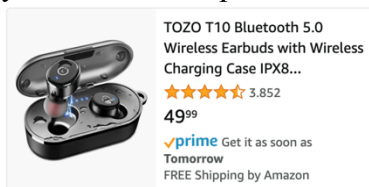
Based on your first impression, and imagining you were looking to buy new headphones, how **likely is it that you would click on the product** and take it **into further consideration for a possible purchase?**

Very unlikely             Very likely

Page Break for **OM-group**

**G1\_CC\_MC\_Vol**  
*Product image only inserted for OM-group*

Based on your previous experience with Amazon (if any), how do you evaluate the product's **amount of reviews?**




- Very low
- Low
- Moderate
- High
- Very high

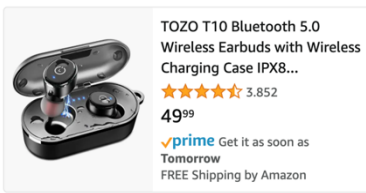
**G1\_CC\_MC\_Val**

Based on your previous experience with Amazon (if any), how do you evaluate the **product's star rating?**

- Very low
- Low
- Moderate
- High
- Very high

Page Break

|  |   |
|--|---|
| <p><b>G1_CC_DIS_Prime</b></p>  | <p>Considering your decision to click on the product to get more information, <b>how important is it</b> for you, that ...<br/> <b>... it is a Prime product?</b></p> <ul style="list-style-type: none"> <li><input type="radio"/> Not at all important</li> <li><input type="radio"/> Slightly important</li> <li><input type="radio"/> Moderately important</li> <li><input type="radio"/> Very important</li> <li><input type="radio"/> Extremely important</li> <li><input type="radio"/> <i>I don't have a Prime membership or don't know Prime.</i></li> </ul>  |
| <p><b>G1_CC_DIS_Ship</b></p>   | <p>... the <b>shipping date</b> is shown?</p> <ul style="list-style-type: none"> <li><input type="radio"/> Not at all important</li> <li><input type="radio"/> Slightly important</li> <li><input type="radio"/> Moderately important</li> <li><input type="radio"/> Very important</li> <li><input type="radio"/> Extremely important</li> </ul>   |
| <p><b>G1_CC_ATT1</b><br/><i>Multiple Answers possible</i></p>                              | <p>What else is important to you? Please show attention, tick "other" and write down the word "red" in the text box.</p> <ul style="list-style-type: none"> <li><input type="radio"/> Image</li> <li><input type="radio"/> Long Title</li> <li><input type="radio"/> Price</li> <li><input type="radio"/> Shipping Costs</li> <li><input type="radio"/> Vouchers</li> <li><input type="radio"/> Other: _____</li> </ul>   |
| <p>Page Break</p>  |   |
| <p><b>G1_CC_SUS_Bel</b><br/><i>MEDV</i></p>  | <p>What is your <b>first impression</b> about the product's reviews?</p> <div data-bbox="555 1218 927 1402" style="border: 1px solid gray; padding: 5px;">  <p>TOZO T10 Bluetooth 5.0<br/>Wireless Earbuds with Wireless<br/>Charging Case IPX8...</p> <p>★★★★☆ 3.852</p> <p>49<sup>99</sup></p> <p> Get it as soon as<br/>Tomorrow<br/>FREE Shipping by Amazon</p> </div> <p><b>The reviews seem to be ...</b><br/> Believable    <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>    Unbelievable</p> |
| <p><b>G1_CC_SUS_Truth</b><br/><i>MEDV</i></p>  | <p><b>The reviews seem to be ...</b><br/> Truthful    <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>    Not Truthful</p>   |
| <p><b>G1_LL_SUS_Dec</b><br/><i>MEDV</i></p>  | <p><b>The reviews seem to be ...</b><br/> Non-Deceptive    <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>    Deceptive</p>   |
| <p>Page Break</p>  |   |
| <p><b>G1_MO_CC_CHOICE</b><br/><i>DV</i><br/><br/><i>Question position for MO-group</i></p> | <p>Based on your first impression, and imagining you were looking to buy new headphones, how <b>likely is it that you would click on the product</b> and take it <b>into further consideration for a possible purchase?</b></p>   |

|  |  |
|--|--|
|  |  <p>Very unlikely    <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>    Very likely</p> |
|--|--|

Page Break

|  |  |
|--|--|
| <b>G1_CC_ATT2</b>  | How many reviews did the product you saw during this survey have?  |
| <i>“I don’t remember” and the wrong number is considered as fail</i> | <input type="radio"/> 3.852<br><input type="radio"/> 179<br><input type="radio"/> 136.571<br><input type="radio"/> I don’t remember. |

Page Break

|        |   |
|--------|---|
| Gender | Please indicate your <b>gender</b> .<br><input type="radio"/> Male<br><input type="radio"/> Female<br><input type="radio"/> Non-binary / third gender<br><input type="radio"/> Prefer not to say. |
|--------|---|

|     |   |
|-----|---|
| Age | Please indicate your <b>age</b> . _____ |
|-----|---|

|         |  |
|---------|--|
| Country | In which <b>country</b> do you currently live?<br><input type="radio"/> USA<br><input type="radio"/> UK<br><input type="radio"/> Germany<br><input type="radio"/> Other: _____ |
|---------|--|

|           |   |
|-----------|---|
| Education | What is you highest level of education?<br><input type="radio"/> Less than secondary school<br><input type="radio"/> Secondary school<br><input type="radio"/> Bachelor’s degree<br><input type="radio"/> Master’s degree<br><input type="radio"/> Doctorate degree<br><input type="radio"/> Other: _____ |
|-----------|---|

|         |   |
|---------|---|
| Comment | Is there anything you would like to say or comment on?<br>_____ |
|---------|---|

Page Break

|            |   |
|------------|---|
| Debriefing | <p>Thanks for participating! Your responses have been transmitted.</p> <p>The goal of this study was to measure the effects of extreme high amounts of reviews on perceived suspicion and finally choice intention. Data is gathered by means of an experiment. Participants were randomly assigned to one of four experimental groups and exposed to different levels of review amounts.</p> <p>Have a nice day!</p> |
|------------|---|

## Appendix 6: Survey Flow

Introduction

(Intro | FREQ | HF\_Intro; LF\_Intro)

Randomization into **G1**, **G2**, **G3** or **G4**

Randomization for **G1** into Group **MO** or **OM**

**G1\_MO**

(G1\_MO\_CC\_Intro | G1\_MO\_CC\_MC\_Vol | G1\_MO\_CC\_MC\_Val | G1\_MO\_CC\_DIS\_Prime | G1\_MO\_CC\_DIS\_Ship | G1\_MO\_CC\_ATT1 | G1\_MO\_CC\_SUS\_Bel | G1\_MO\_CC\_SUS\_Truth | G1\_MO\_CC\_SUS\_Dec | G1\_MO\_CC\_CHOICE | G1\_MO\_CC\_ATT2)

**G1\_OM**

(G1\_OM\_CC\_Intro | G1\_OM\_CC\_CHOICE | G1\_OM\_CC\_MC\_Vol | G1\_OM\_CC\_MC\_Val | G1\_OM\_CC\_DIS\_Prime | G1\_OM\_CC\_DIS\_Ship | G1\_OM\_CC\_ATT1 | G1\_OM\_CC\_SUS\_Bel | G1\_OM\_CC\_SUS\_Truth | G1\_OM\_CC\_SUS\_Dec | G1\_OM\_CC\_ATT2)

Randomization for **G2** into Group **MO** and **OM**

**G2\_MO**

**G2\_OM**

Randomization for **G3** into Group **MO** and **OM**

**G3\_MO**

**G3\_OM**

Randomization for **G4** into Group **MO** and **OM**

**G4\_MO**

**G4\_OM**

Demographics

(Gender | Age | Country | Comment)

End

## Appendix 7: Attention Check Analysis

### Attention Check 1

|       |         | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|---------|-----------|---------|---------------|--------------------|
| Valid | Fail    | 190       | 28.8    | 28.8          | 28.8               |
|       | Success | 469       | 71.2    | 71.2          | 100.0              |
|       | Total   | 659       | 100.0   | 100.0         |                    |

### Attention Check 2

|       |         | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|---------|-----------|---------|---------------|--------------------|
| Valid | Fail    | 17        | 2.6     | 2.6           | 2.6                |
|       | Success | 642       | 97.4    | 97.4          | 100.0              |
|       | Total   | 659       | 100.0   | 100.0         |                    |

### AC 1 \* AC 2 Crosstabulation

|       |         | AC 2          |        |         |        |
|-------|---------|---------------|--------|---------|--------|
|       |         |               | Fail   | Success | Total  |
| AC 1  | Fail    | Count         | 9      | 181     | 190    |
|       |         | % within AC 1 | 4.7%   | 95.3%   | 100.0% |
|       |         | % within AC 2 | 52.9%  | 28.2%   | 28.8%  |
|       |         | % of Total    | 1.4%   | 27.5%   | 28.8%  |
|       | Success | Count         | 8      | 461     | 469    |
|       |         | % within AC 1 | 1.7%   | 98.3%   | 100.0% |
|       |         | % within AC 2 | 47.1%  | 71.8%   | 71.2%  |
| Total |         | Count         | 17     | 642     | 659    |
|       |         | % within AC 1 | 2.6%   | 97.4%   | 100.0% |
|       |         | % within AC 2 | 100.0% | 100.0%  | 100.0% |
|       |         | % of Total    | 2.6%   | 97.4%   | 100.0% |

## Appendix 8: Demographics

### Gender

|       |                           | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|---------------------------|-----------|---------|---------------|--------------------|
| Valid | Male                      | 244       | 38.1    | 38.1          | 38.1               |
|       | Female                    | 390       | 60.9    | 60.9          | 99.1               |
|       | Non-binary / third gender | 5         | .8      | .8            | 99.8               |
|       | Prefer not to say         | 1         | .2      | .2            | 100.0              |
|       | Total                     | 640       | 100.0   | 100.0         |                    |

### Descriptive Statistics: Age

|                           | N   | Range | Minimum | Maximum | Mean  | Std. Deviation |
|---------------------------|-----|-------|---------|---------|-------|----------------|
| Please indicate your age. | 640 | 40    | 20      | 60      | 36.38 | 9.626          |
| Valid N (listwise)        | 640 |       |         |         |       |                |

### Education

|       |                            | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|----------------------------|-----------|---------|---------------|--------------------|
| Valid | Less than secondary school | 6         | .9      | .9            | .9                 |
|       | Secondary school           | 192       | 30.0    | 30.0          | 30.9               |
|       | Bachelor's degree          | 304       | 47.5    | 47.5          | 78.4               |
|       | Master's degree            | 103       | 16.1    | 16.1          | 94.5               |
|       | Doctorate Degree           | 16        | 2.5     | 2.5           | 97.0               |
|       | Other                      | 19        | 3.0     | 3.0           | 100.0              |
|       | Total                      | 640       | 100.0   | 100.0         |                    |

### Country of Residence

|       |         | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|---------|-----------|---------|---------------|--------------------|
| Valid | USA     | 60        | 9.4     | 9.4           | 9.4                |
|       | UK      | 461       | 72.0    | 72.0          | 81.4               |
|       | Germany | 119       | 18.6    | 18.6          | 100.0              |
|       | Total   | 640       | 100.0   | 100.0         |                    |

**Appendix 9: Amazon Shopping Frequency**

*Amazon Shopping Frequency*

|                         | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------------------------|-----------|---------|---------------|--------------------|
| Valid Every week        | 109       | 17.0    | 17.0          | 17.0               |
| Several times per month | 270       | 42.2    | 42.2          | 59.2               |
| Once per month          | 141       | 22.0    | 22.0          | 81.3               |
| Several times a year    | 108       | 16.9    | 16.9          | 98.1               |
| Once a year             | 12        | 1.9     | 1.9           | 100.0              |
| Total                   | 640       | 100.0   | 100.0         |                    |

## Appendix 10: Tests of Randomness of Experimental Groups

### Chi-Square Tests: Gender

|                              | 8 Groups            |    |                                   | 4 Groups (OM & MO aggregated) |    |                                   |
|------------------------------|---------------------|----|-----------------------------------|-------------------------------|----|-----------------------------------|
|                              | Value               | df | Asymptotic Significance (2-sided) | Value                         | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square           | 16.910 <sup>a</sup> | 21 | .717                              | 6.918 <sup>a</sup>            | 9  | .646                              |
| Likelihood Ratio             | 15.089              | 21 | .818                              | 7.782                         | 9  | .556                              |
| Linear-by-Linear Association | 1.014               | 1  | .314                              | .115                          | 1  | .735                              |
| N of Valid Cases             | 640                 |    |                                   | 640                           |    |                                   |

a. 16 cells (50.0%) have expected count less than 5. The minimum expected count is .12.

a. 8 cells (50.0%) have expected count less than 5. The minimum expected count is .25.

### Chi-Square Tests: Education

|                              | 8 Groups            |    |                                   | 4 Groups (OM & MO aggregated) |    |                                   |
|------------------------------|---------------------|----|-----------------------------------|-------------------------------|----|-----------------------------------|
|                              | Value               | df | Asymptotic Significance (2-sided) | Value                         | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square           | 55.060 <sup>a</sup> | 35 | .017                              | 21.569 <sup>a</sup>           | 15 | .120                              |
| Likelihood Ratio             | 53.409              | 35 | .024                              | 21.132                        | 15 | .133                              |
| Linear-by-Linear Association | .002                | 1  | .966                              | .325                          | 1  | .568                              |
| N of Valid Cases             | 640                 |    |                                   | 640                           |    |                                   |

a. 24 cells (50.0%) have expected count less than 5. The minimum expected count is .73.

a. 12 cells (50.0%) have expected count less than 5. The minimum expected count is 1.48.

### Chi-Square Tests: Country

|                              | 8 Groups           |    |                                   | 4 Groups (OM & MO aggregated) |    |                                   |
|------------------------------|--------------------|----|-----------------------------------|-------------------------------|----|-----------------------------------|
|                              | Value              | df | Asymptotic Significance (2-sided) | Value                         | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square           | 8.995 <sup>a</sup> | 14 | .831                              | 2.285 <sup>a</sup>            | 6  | .892                              |
| Likelihood Ratio             | 9.901              | 14 | .769                              | 2.260                         | 6  | .894                              |
| Linear-by-Linear Association | .365               | 1  | .546                              | .200                          | 1  | .655                              |
| N of Valid Cases             | 640                |    |                                   | 640                           |    |                                   |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.31.

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.81.

*Chi-Square Tests: Amazon Shopping Frequency*

|                              | 8 Groups            |    |                                   | 4 Groups (OM & MO aggregated) |    |                                   |
|------------------------------|---------------------|----|-----------------------------------|-------------------------------|----|-----------------------------------|
|                              | Value               | df | Asymptotic Significance (2-sided) | Value                         | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square           | 55.580 <sup>a</sup> | 28 | .001                              | 16.355 <sup>a</sup>           | 12 | .176                              |
| Likelihood Ratio             | 57.267              | 28 | .001                              | 16.499                        | 12 | .169                              |
| Linear-by-Linear Association | 3.403               | 1  | .065                              | .906                          | 1  | .341                              |
| N of Valid Cases             | 640                 |    |                                   | 640                           |    |                                   |

a. 8 cells (20.0%) have expected count less than 5. The minimum expected count is 1.46. a. 4 cells (20.0%) have expected count less than 5. The minimum expected count is 2.96.

*ANOVA: Age – 8 Groups*

|                | Sum of Squares | df  | Mean Square | F    | Sig. |
|----------------|----------------|-----|-------------|------|------|
| Between Groups | 228.527        | 7   | 32.647      | .350 | .931 |
| Within Groups  | 58986.448      | 632 | 93.333      |      |      |
| Total          | 59214.975      | 639 |             |      |      |

*ANOVA: Age – 4 Groups (OM & MO aggregated)*

|                | Sum of Squares | df  | Mean Square | F    | Sig. |
|----------------|----------------|-----|-------------|------|------|
| Between Groups | 73.903         | 3   | 24.634      | .265 | .851 |
| Within Groups  | 59141.072      | 636 | 92.989      |      |      |
| Total          | 59214.975      | 639 |             |      |      |

**Comment on the Randomness of Variables:**

There is no sig. dependence between the eight groups and gender, age, and residential country. For education ( $\chi^2(35, n = 640) = 55.1, p = .017$ ) and shopping frequency ( $\chi^2(28, n = 640) = 55.6, p = .001$ ) participants are not perfectly random distributed between the groups. When considering only the four experimental groups – ignoring MO/OM – all variables are significantly randomly distributed.

## Appendix 11: Sample Size and Description of Experimental Groups

*Groups \* Order Crosstabulation: Count*

|        |       | Order |     |       |
|--------|-------|-------|-----|-------|
|        |       | MO    | OM  | Total |
| Groups | G1_CC | 79    | 82  | 161   |
|        | G2_HC | 78    | 80  | 158   |
|        | G3_CH | 81    | 81  | 162   |
|        | G4_HH | 81    | 78  | 159   |
| Total  |       | 319   | 321 | 640   |

### **Comment on the Sample Size:**

As Prolific accepted only the first AC, many participants were rejected and replaced automatically. However, the data of those rejected participants who failed the first AC, but passed the second were still considered in the analysis leading to a sample size above plan.

## Appendix 12: Scale Reliability Analysis

### Case Processing Summary

|                       | N   | %     |
|-----------------------|-----|-------|
| Cases Valid           | 640 | 100.0 |
| Excluded <sup>a</sup> | 0   | .0    |
| Total                 | 640 | 100.0 |

a. Listwise deletion based on all variables in the procedure.

### Reliability Statistics

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|------------------|--|------------|
| .9340            | .936   | 3          |

### Item Statistics

|           | Mean   | Std. Deviation | N   |
|-----------|--------|----------------|-----|
| SUS_Bel   | 3.5563 | 1.71524        | 640 |
| SUS_Truth | 3.6031 | 1.53376        | 640 |
| SUS_Dec   | 3.7297 | 1.62831        | 640 |

### Inter-Item Correlation Matrix

|           | SUS_BEL | SUS_Truth | SUS_Dec |
|-----------|---------|-----------|---------|
| SUS_Bel   | 1.000   | .853      | .789    |
| SUS_Truth | .853    | 1.000     | .845    |
| SUS_Dec   | .789    | .845      | 1.000   |

### Item-Total Statistics

|           | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
|-----------|----------------------------|--------------------------------|----------------------------------|------------------------------|----------------------------------|
| SUS_Bel   | 7.3328                     | 9.224                          | .854                             | .744                         | .915                             |
| SUS_Truth | 7.2859                     | 9.998                          | .898                             | .806                         | .881                             |
| SUS_Dec   | 7.1594                     | 9.784                          | .847                             | .731                         | .918                             |

## Appendix 13: Manipulation Check Analysis: Check of Assumptions

### Test of Homogeneity of Variances

|        |                                      | Levene    | df1 | df2     | Sig. |
|--------|--------------------------------------|-----------|-----|---------|------|
|        |                                      | Statistic |     |         |      |
| MC_VOL | Based on Mean                        | .521      | 1   | 638     | .471 |
|        | Based on Median                      | .305      | 1   | 638     | .581 |
|        | Based on Median and with adjusted df | .305      | 1   | 623.888 | .581 |
|        | Based on trimmed mean                | .017      | 1   | 638     | .896 |

### Tests of Normality

|        |                    | Kolmogorov-Smirnov <sup>a</sup> |     |      | Shapiro-Wilk |     |      |
|--------|--------------------|---------------------------------|-----|------|--------------|-----|------|
| VOL    |                    | Statistic                       | df  | Sig. | Statistic    | df  | Sig. |
| MC_VOL | Control VOL        | .266                            | 323 | .000 | .820         | 323 | .000 |
|        | Extremely High VOL | .337                            | 317 | .000 | .732         | 317 | .000 |

a. Lilliefors Significance Correction

### Test of Homogeneity of Variances

|        |                                      | Levene    | df1 | df2     | Sig. |
|--------|--------------------------------------|-----------|-----|---------|------|
|        |                                      | Statistic |     |         |      |
| MC_VAL | Based on Mean                        | 10.436    | 1   | 638     | .001 |
|        | Based on Median                      | .576      | 1   | 638     | .448 |
|        | Based on Median and with adjusted df | .576      | 1   | 591.943 | .448 |
|        | Based on trimmed mean                | 4.213     | 1   | 638     | .041 |

### Robust Tests of Equality of Means

| MC_VAL         | Statistic <sup>a</sup> | df1 | df2     | Sig. |
|----------------|------------------------|-----|---------|------|
| Welch          | 37.730                 | 1   | 636.860 | .000 |
| Brown-Forsythe | 37.730                 | 1   | 636.860 | .000 |

a. Asymptotically F distributed.

### Tests of Normality

|        |                    | Kolmogorov-Smirnov <sup>a</sup> |     |      | Shapiro-Wilk |     |      |
|--------|--------------------|---------------------------------|-----|------|--------------|-----|------|
| VAL    |                    | Statistic                       | df  | Sig. | Statistic    | df  | Sig. |
| MC_VAL | Control VAL        | .295                            | 319 | .000 | .784         | 319 | .000 |
|        | Extremely High VAL | .375                            | 321 | .000 | .681         | 321 | .000 |

a. Lilliefors Significance Correction

## Appendix 14: Manipulation Check Analysis: Descriptive Statistics and ANOVA

### Descriptives: MC\_VOL

|                    | N   | Mean   | Std. Deviation | Std. Error | 95% Confidence Interval for Mean |             | Minimum | Maximum |
|--------------------|-----|--------|----------------|------------|----------------------------------|-------------|---------|---------|
|                    |     |        |                |            | Lower Bound                      | Upper Bound |         |         |
| Control VOL        | 323 | 3.9814 | .92898         | .05169     | 3.8797                           | 4.0831      | 1.00    | 5.00    |
| Extremely High VOL | 317 | 4.4006 | .79181         | .04447     | 4.3131                           | 4.4881      | 1.00    | 5.00    |
| Total              | 640 | 4.1891 | .88821         | .03511     | 4.1201                           | 4.2580      | 1.00    | 5.00    |

### ANOVA: MC\_VOL

|                | Sum of Squares | df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 28.115         | 1   | 28.115      | 37.683 | .000 |
| Within Groups  | 476.008        | 638 | .746        |        |      |
| Total          | 504.123        | 639 |             |        |      |

### Descriptives: MC\_VAL

|                    | N   | Mean   | Std. Deviation | Std. Error | 95% Confidence Interval for Mean |             | Minimum | Maximum |
|--------------------|-----|--------|----------------|------------|----------------------------------|-------------|---------|---------|
|                    |     |        |                |            | Lower Bound                      | Upper Bound |         |         |
| Control VAL        | 319 | 4.1285 | .73058         | .04090     | 4.0480                           | 4.2090      | 1.00    | 5.00    |
| Extremely High VAL | 321 | 4.4922 | .76696         | .04281     | 4.4080                           | 4.5764      | 1.00    | 5.00    |
| Total              | 640 | 4.3109 | .77026         | .03045     | 4.2511                           | 4.3707      | 1.00    | 5.00    |

### ANOVA: MC\_VAL

|                | Sum of Squares | df  | Mean Square | F      | Sig. |
|----------------|----------------|-----|-------------|--------|------|
| Between Groups | 21.163         | 1   | 21.163      | 37.718 | .000 |
| Within Groups  | 357.961        | 638 | .561        |        |      |
| Total          | 379.123        | 639 |             |        |      |

## Appendix 15: Question Order Analysis: Check of Assumptions

### *Check of ANOVA-assumptions*

| Assumption               | Suspicion                                 | Choice Intention                          |
|--------------------------|---|---|
| Independent Groups       | √   | √   |
| No Outliers              | √   | √   |
| Homogeneity of Variances | √   | √   |
|                          | see table below                           | see table below                           |
| Normality                | √   | √   |
|                          | see table below,<br>central limit theorem | see table below,<br>central limit theorem |

### *Levene's Test of Equality of Error Variances*

| Suspicion |     |     |      | Choice Intention |     |     |      |
|-----------|-----|-----|------|------------------|-----|-----|------|
| F         | df1 | df2 | Sig. | F                | df1 | df2 | Sig. |
| 1.804     | 7   | 632 | .084 | 1.110            | 7   | 632 | .355 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.  
Design: VOL + VAL + OM + VOL\*OM + VAL\*OM + Gender + Age + Country + Educat + AMZFreq

### *Tests of Normality*

|           | Kolmogorov-Smirnov <sup>a</sup> |     |      | Shapiro-Wilk |     |      |
|-----------|---------------------------------|-----|------|--------------|-----|------|
|           | Statistic                       | df  | Sig. | Statistic    | df  | Sig. |
| Suspicion | .095                            | 640 | .000 | .967         | 640 | .000 |
| Choice    | .203                            | 640 | .000 | .888         | 640 | .000 |

a. Lilliefors Significance Correction

## Appendix 16: Question Order Analysis: Descriptive Statistics and ANOVA

### Descriptive Statistics: Suspicion

| Order | Mean   | Std. Deviation | N   |
|-------|--------|----------------|-----|
| MO    | 3.6155 | 1.57787        | 319 |
| OM    | 3.6438 | 1.48259        | 321 |
| Total | 3.6297 | 1.52969        | 640 |

### Tests of Between-Subjects Effects

| Dependent Variable: Suspicion |                         |     |             |         |      |                     |
|-------------------------------|-------------------------|-----|-------------|---------|------|---------------------|
| Source                        | Type III Sum of Squares | df  | Mean Square | F       | Sig. | Partial Eta Squared |
| Model                         | 8514.334 <sup>a</sup>   | 11  | 774.030     | 344.643 | .000 | .858                |
| OM                            | .377                    | 1   | .377        | .168    | .682 | .000                |
| VOL                           | 4.054                   | 1   | 4.054       | 1.805   | .180 | .003                |
| VAL                           | 20.631                  | 1   | 20.631      | 9.186   | .003 | .014                |
| VOL * OM                      | 2.121                   | 1   | 2.121       | .944    | .332 | .001                |
| VAL * OM                      | .000                    | 1   | .000        | .000    | .991 | .000                |
| Gender                        | 8.281                   | 1   | 8.281       | 3.687   | .055 | .006                |
| Age                           | .263                    | 1   | .263        | .117    | .732 | .000                |
| Country                       | 6.556                   | 1   | 6.556       | 2.919   | .088 | .005                |
| Educat                        | 5.053                   | 1   | 5.053       | 2.250   | .134 | .004                |
| AMZFreq                       | 1.860                   | 1   | 1.860       | .828    | .363 | .001                |
| Error                         | 1412.666                | 629 | 2.246       |         |      |                     |
| Total                         | 9927.000                | 640 |             |         |      |                     |

a. R Squared = .858 (Adjusted R Squared = .855)

*Descriptive Statistics: Choice*

| Order | Mean   | Std. Deviation | N   |
|-------|--------|----------------|-----|
| MO    | 4.1223 | 1.59424        | 319 |
| OM    | 4.0436 | 1.52212        | 321 |
| Total | 4.0828 | 1.55776        | 640 |

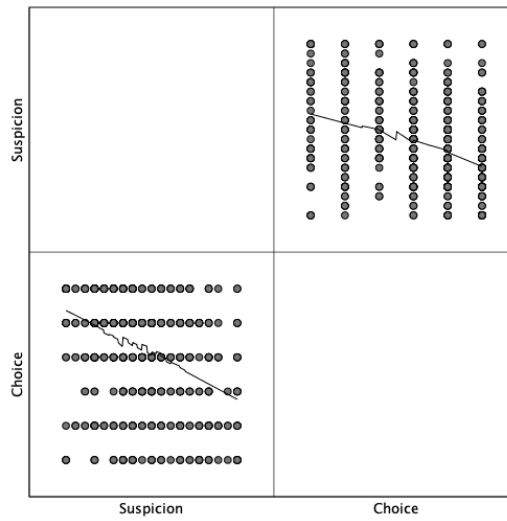
*Tests of Between-Subjects Effects*

| Dependent Variable: Choice |                         |     |             |         |      |                     |
|----------------------------|-------------------------|-----|-------------|---------|------|---------------------|
| Source                     | Type III Sum of Squares | df  | Mean Square | F       | Sig. | Partial Eta Squared |
| Model                      | 10678.423 <sup>a</sup>  | 11  | 970.766     | 396.353 | .000 | .874                |
| OM                         | .788                    | 1   | .788        | .322    | .571 | .001                |
| VOL                        | .219                    | 1   | .219        | .089    | .765 | .000                |
| VAL                        | .070                    | 1   | .070        | .028    | .866 | .000                |
| VOL * OM                   | .051                    | 1   | .051        | .021    | .886 | .000                |
| VAL * OM                   | .022                    | 1   | .022        | .009    | .924 | .000                |
| Gender                     | 1.015                   | 1   | 1.015       | .415    | .520 | .001                |
| Age                        | 4.581                   | 1   | 4.581       | 1.871   | .172 | .003                |
| Country                    | 3.246                   | 1   | 3.246       | 1.325   | .250 | .002                |
| Educat                     | 1.166                   | 1   | 1.166       | .476    | .490 | .001                |
| AMZFreq                    | .800                    | 1   | .800        | .327    | .568 | .001                |
| Error                      | 1540.577                | 629 | 2.449       |         |      |                     |
| Total                      | 12219.000               | 640 |             |         |      |                     |

a. R Squared = .874 (Adjusted R Squared = .872)

## Appendix 17: CMMM Analysis: Check of Assumptions

### Test for Linearity Assumption



### Test for Homoscedasticity Assumption

#### *Breusch-Pagan Test for Heteroskedasticity<sup>a,b,c</sup>: A-Path*

| Chi-Square | df | Sig. |
|------------|----|------|
| 2.603      | 1  | .107 |

- a. Dependent variable: Suspicion  
 b. Tests the null hypothesis that the variance of the errors does not depend on the values of the independent variables.  
 c. Predicted values from design: Intercept + Interaction\_VolVal + VOL + VAL

#### *Breusch-Pagan Test for Heteroskedasticity<sup>a,b,c</sup>: B-Path*

| Chi-Square | df | Sig. |
|------------|----|------|
| .372       | 1  | .542 |

- a. Dependent variable: Choice  
 b. Tests the null hypothesis that the variance of the errors does not depend on the values of the independent variables.  
 c. Predicted values from design: Intercept + SUSPICION

#### *Breusch-Pagan Test for Heteroskedasticity<sup>a,b,c</sup>: C-Path*

| Chi-Square | df | Sig. |
|------------|----|------|
| .005       | 1  | .942 |

- a. Dependent variable: Choice  
 b. Tests the null hypothesis that the variance of the errors does not depend on the values of the independent variables.  
 c. Predicted values from design: Intercept + VOL

## Test for Independence of Error Terms Assumption

### *Model Summary<sup>b</sup>: A-Path*

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------------------|----------|-------------------|----------------------------|---------------|
| 1     | .202 <sup>a</sup> | .041     | .036              | 1.50159                    | 1.811         |

a. Predictors: (Constant), Interaction\_VolVal, Valence, Volume

b. Dependent Variable: Suspicion

### *Model Summary<sup>b</sup>: B-Path*

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------------------|----------|-------------------|----------------------------|---------------|
| 1     | .321 <sup>a</sup> | .103     | .102              | 1.47639                    | 1.875         |

a. Predictors: (Constant), Suspicion

b. Dependent Variable: Choice

### *Model Summary<sup>b</sup>: C-Path*

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------------------|----------|-------------------|----------------------------|---------------|
| 1     | .012 <sup>a</sup> | .000     | -.001             | 1.55888                    | 1.802         |

a. Predictors: (Constant), Volume

b. Dependent Variable: Choice

## Appendix 18: PROCESS Output (with Covariates)

```

Run MATRIX procedure:

***** PROCESS Procedure for SPSS Version 3.5.2 *****

      Written by Andrew F. Hayes, Ph.D.      www.afhayes.com
      Documentation available in Hayes (2018). www.guilford.com/p/hayes3

*****
Model   : 7
  Y     : CHOICE
  X     : VOL
  M     : SUSPICI
  W     : VAL

Covariates:
  Gender  Age      Country  Educat  AMZFreq

Sample
Size: 640

*****
OUTCOME VARIABLE:
  SUSPICI

Model Summary
      R      R-sq      MSE      F      df1      df2      p
      .235      .055      2.239      4.053      8.000      631.000      .000

Model
      coeff      se      t      p      LLCI      ULCI
constant  -39.015      26.529      -1.471      .142      -91.112      13.082
VOL        .222      .165      1.341      .180      -.103      .546
VAL        .388      .160      2.422      .016      .074      .703
Int_1      .242      .240      1.006      .315      -.230      .713
Gender     -.226      .126      -1.801      .072      -.473      .020
Age        .002      .007      .348      .728      -.011      .015
Country    .215      .135      1.599      .110      -.049      .479
Educat     .100      .061      1.651      .099      -.019      .220
AMZFreq    -.062      .058      -1.067      .286      -.176      .052

Product terms key:
  Int_1      :      VOL      x      VAL

Test(s) of highest order unconditional interaction(s):
      R2-chng      F(HC4)      df1      df2      p
X*W      .002      1.013      1.000      631.000      .315
-----
      Focal predict: VOL      (X)
      Mod var: VAL      (W)

*****
OUTCOME VARIABLE:
  CHOICE

Model Summary
      R      R-sq      MSE      F      df1      df2      p
      .329      .108      2.187      10.527      7.000      632.000      .000

Model
      coeff      se      t      p      LLCI      ULCI
constant  20.941      23.837      .878      .380      -25.869      67.750
VOL        .151      .116      1.298      .195      -.077      .379
SUSPICI    -.331      .041      -8.091      .000      -.411      -.250
Gender     .006      .117      .048      .961      -.224      .235

```

|         |       |      |        |      |       |      |
|---------|-------|------|--------|------|-------|------|
| Age     | -.009 | .006 | -1.397 | .163 | -.022 | .004 |
| Country | -.078 | .121 | -.646  | .518 | -.316 | .159 |
| Educat  | -.016 | .060 | -.267  | .789 | -.133 | .101 |
| AMZFreq | .017  | .058 | .287   | .775 | -.097 | .130 |

\*\*\*\*\* DIRECT AND INDIRECT EFFECTS OF X ON Y \*\*\*\*\*

Direct effect of X on Y

| Effect | se(HC4) | t     | p    | LLCI  | ULCI |
|--------|---------|-------|------|-------|------|
| .151   | .116    | 1.298 | .195 | -.077 | .379 |

Conditional indirect effects of X on Y:

INDIRECT EFFECT:

| VAL   | Effect | BootSE | BootLLCI | BootULCI |
|-------|--------|--------|----------|----------|
| .000  | -.073  | .055   | -.183    | .034     |
| 1.000 | -.153  | .061   | -.277    | -.039    |

Index of moderated mediation (difference between conditional indirect effects):

| VAL | Index | BootSE | BootLLCI | BootULCI |
|-----|-------|--------|----------|----------|
| --- | -.080 | .081   | -.248    | .076     |

\*\*\*\*\* ANALYSIS NOTES AND ERRORS \*\*\*\*\*

Level of confidence for all confidence intervals in output:

95.0000

Number of bootstrap samples for percentile bootstrap confidence intervals:

5000

NOTE: A heteroscedasticity consistent standard error and covariance matrix estimator was used.

----- END MATRIX -----

## Appendix 19: PROCESS Output (without Covariates)

```

Run MATRIX procedure:

***** PROCESS Procedure for SPSS Version 3.5.2 *****

      Written by Andrew F. Hayes, Ph.D.      www.afhayes.com
      Documentation available in Hayes (2018). www.guilford.com/p/hayes3

*****
Model   : 7
  Y     : CHOICE
  X     : VOL
  M     : SUSPICI
  W     : VAL

Sample
Size:   640

*****
OUTCOME VARIABLE:
  SUSPICI

Model Summary
      R      R-sq      MSE      F(HC4)      df1      df2      p
      .202      .041      2.255      8.833      3.000      636.000      .000

Model
      coeff      se(HC4)      t      p      LLCI      ULCI
constant      3.257      .110      29.675      .000      3.041      3.472
VOL            .241      .164      1.466      .143      -.082      .564
VAL            .406      .159      2.554      .011      .094      .718
Int_1          .201      .238      .846      .398      -.265      .668

Product terms key:
  Int_1      :      VOL      x      VAL

Test(s) of highest order unconditional interaction(s):
      R2-chng      F(HC4)      df1      df2      p
X*W            .001      .716      1.000      636.000      .398

*****
OUTCOME VARIABLE:
  CHOICE

Model Summary
      R      R-sq      MSE      F(HC4)      df1      df2      p
      .325      .105      2.178      34.757      2.000      637.000      .000

Model
      coeff      se(HC4)      t      p      LLCI      ULCI
constant      5.216      .166      31.427      .000      4.890      5.542
VOL            .150      .116      1.290      .197      -.078      .377
SUSPICI       -.333      .040      -8.250      .000      -.412      -.253

***** DIRECT AND INDIRECT EFFECTS OF X ON Y *****

Direct effect of X on Y
      Effect      se(HC4)      t      p      LLCI      ULCI
      .150      .116      1.290      .197      -.078      .377

Conditional indirect effects of X on Y:

INDIRECT EFFECT:
  VOL      ->      SUSPICI      ->      CHOICE

```

| VAL   | Effect | BootSE | BootLLCI | BootULCI |
|-------|--------|--------|----------|----------|
| .000  | -.080  | .055   | -.190    | .028     |
| 1.000 | -.147  | .061   | -.271    | -.032    |

Index of moderated mediation (difference between conditional indirect effects):

| VAL | Index | BootSE | BootLLCI | BootULCI |
|-----|-------|--------|----------|----------|
| VAL | -.067 | .081   | -.229    | .089     |

---

\*\*\*\*\* ANALYSIS NOTES AND ERRORS \*\*\*\*\*

Level of confidence for all confidence intervals in output:  
95.0000

Number of bootstrap samples for percentile bootstrap confidence intervals:  
5000

NOTE: A heteroscedasticity consistent standard error and covariance matrix estimator was used.

----- END MATRIX -----

## Appendix 20: Qualitative Study Results (Extract)

- a) *In the past I would get paid to write fake reviews (I am holding my head in shame). I would look at the five-star reviews and see if it looked like they are fake reviews.*
- b) *I don't always trust reviews and star rating as many companies that sell via Amazon have sites/pages via Facebook that people use to get discounts on products in return for a good review. Some even get the products for free. Electronics such as ear-phones/headphones are one of the main contenders.*
- c) *I used to run a deal group on fb, and was a member of many sites where sellers would have free or nearly free products to purchase. We also had sellers directly contact us for reviews that were not on the sites. **They expected 5 star reviews in return for free or nearly free products.** This is NOT allowed! They also would pay people to like or dislike another seller's products, and also paid people to write fake reviews! This greatly harmed the integrity of our genuine reviews! **It still goes on today**, as I still receive emails from sellers asking for 5 star reviews, and in return they will pay up to \$50 for the review, and also give free high dollar products like gaming office chairs! [...] **It's shameful, and reviews on every site have lost all credibility!** [...] Seeing a NEW product, that is not a name brand company with nearly 4 thousand reviews sends lots of red flags, and 100% those reviews are not honest!*
- d) *Don't trust reviews as a lot of times it includes various styles, not only the one that you're interested in. [...]*
- e) *Because of the scams surrounding product reviews, I would be interested in the product due to the high volume of reviews but **also would be sceptical** unless I had heard first hand a review.*
- f) *There are too **many products** on Amazon with **fake/planted reviews**, so I would assume any product like that with tens of thousands of reviews (**unless it was a market leader**) was **not to be trusted**.*
- g) *There might be many reviews but these might be due to a product being online a long period of time, being sold many thousand times, being a market leader etc. This has to be checked first. **Many thousand reviews within a few days make me sceptical.***
- h) *I would not trust reviews just on the stars without knowing **how many reviews there had been and over what period of time (are they current or out of date?)**.*
- i) *Normally I trust the reviews on Amazon fully but **this seemed too good to be true**.*
- j) *On the topic of the product on Amazon, I would be **highly suspicious** of the number of reviews **BUT** I would **still click through to see if they appeared genuine or not**.*
- k) *Amazon reviews **need to be viewed cautiously**, so rather read them and not only trust on the rating/volume.*