ANALYSIS ON USER-GENERATED PICTURES IN ONLINE REVIEWS

by

Youngeui Kim

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Management Science

at

The University of Wisconsin-Milwaukee

May 2023

ABSTRACT

ANALYSIS ON USER-GENERATED PICTURES IN ONLINE REVIEWS

by

Youngeui Kim

The University of Wisconsin-Milwaukee, 2023
Under the Supervision of Professor Huimin Zhao

Since the early-2010s, with the technology development, user-generated pictures in online reviews have flourished. As large volumes of user-generated pictures are being generated along with online reviews, online businesses want to learn whether the pictures can improve the information quality (helpfulness) of reviews and what features of the pictures would be meaningful for consumers in purchasing a product online. Nevertheless, in the academic area, little is known about the effects of visual information, i.e., a review picture, while text-mined features have been widely investigated. In this dissertation research, we strive to tackle the economical role of user-generated pictures posted along with online reviews and to see how the helpfulness of review pictures is perceived differently across national cultures.

Essay 1 explores how visual features in an online review posted along with review texts improve review helpfulness and reduce product returns. We mainly extract a feature of situational cues from review pictures based on two types of information: environmental cues for product usage and human activity-driven aspects. As counterparts of the visual information, we text-mine a functionality feature as well. To control for other visual impacts on review helpfulness and product returns, we additionally image-mine such attributes as a malfunction of a product, image saturation, clarity, and sharpness. The main findings are that the visual feature, which better illustrates product functionality, is perceived as being helpful by consumers and

helps decrease product returns by assisting consumers in making an informed decision. Furthermore, we found a picture superiority effect, that is, shoppers prefer to use image context rather than reading review texts when they have higher information processing costs and a helpful review with a picture helps reduce product returns as well whereas one without does not.

Essay 2 broadens our perspective of online reviews beyond the local e-commerce market. In online review systems, consumers vote for a review fundamentally because they trust the review information and/or reviewer and thus are willing to rely on the given information. Yet, consumers may build trust based on trust antecedents differentially across national cultures– some consumers may depend more on the information quality of a review whereas others rest on the source (reviewer) credibility–according to the elaboration likelihood model theory. Thus, this essay takes a cross-cultural look at online reviews. In this research, both image and text mining techniques are employed. We mainly extract a human appearance feature from review pictures (i.e., information quality) and profile photos (i.e., source credibility) for comparison as counterparts. From review texts, we derive topic diversity (i.e., information quality) using topic modeling and measure review sentiment through sentiment analysis. We found that the perception of review helpfulness tends to be discerned across national cultures. More interestingly, the same content in a different layout, here human appearance in a profile photo versus in a review picture, is differentially helpful for consumers across national cultures. This study can benefit online businesses in terms of global review system management.

To
my parents and
especially my husband

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I thank my advisor, Dr. Huimin Zhao, for his invaluable guidance and support throughout my PhD study and research processes. I appreciate his expertise and encouragement which have helped shape my research principles. I am deeply grateful for his mentorship during my doctoral journey.

I would like to extend my thanks to Dr. Mark Srite, Dr. Yang Wang, and Dr. Abhijeet Ghoshal. Their priceless feedback and insightful comments always allow me to step further. Thanks to their help, I was able to initiate my research projects and successfully develop research ideas. Their support has greatly enhanced the quality of my dissertation.

I am deeply grateful to my husband, Junyong Kim, for all his endless support and practical advice. He, as an experienced PhD and husband, always helps me reduce trial and error in research, teaching, and PhD life. He has been my constant source of energy for the last twelve years. I would also like to express my gratitude to my parents and younger brother in South Korea. Their solid faith in me has given me the strength to pursue my dreams and overcome every challenge that came my way. I cannot thank them enough for their unconditional love and support.

Last but not least, I would like to thank all my friends and colleagues at UWM. We have gone through challenges and shared fun all the time. You have been a source of my joy at UWM. In addition, I would like to express my appreciation to UWM faculty and staff members who have helped me complete my PhD journey smoothly. I am grateful to have them in my life.

CHAPTER 1

Introduction

In recent years, increasingly easy access to mobile devices has led to an explosive growth of visual content on the Internet, and online review platforms are not an exception. User-generated images have become more and more popular, and many websites highlight consumer-posted images by displaying them on top of textual reviews. However, the overabundance of pictures in an online platform can be distracting for visitors as they do not know which images indeed provide helpful information, and not every image has the same value for shoppers regarding helping them make purchase decisions. Even more, the perception of helpfulness may be different across countries due to cultural effects; one visual review feature that is considered helpful in one country may not be in a different country. Hence, it is crucial to know what makes the customer images valuable and how the helpful features are perceived differently across national cultures.

In this dissertation, we explore the role of user-generated pictures in online consumer reviews. The first essay investigates which visualized product information is perceived as being helpful to online consumers. In addition, we look into how the pictures in reviews facilitate consumers to make an informed purchase decision by examining whether the helpful visual features and picture-included reviews reduce product returns. In the second essay, we research how review features are considered helpful differently across national cultures. In the research,

we incorporate a profile photo, another type of visual information, and compare it to the user appearance in review pictures.

Essay 1: Is a Picture Always Worth a Thousand Words? An Exploration of the Informational Role of Images in Online Product Reviews

The picture superiority effect suggests that pictures (compared to text) convey more information and are remembered more. Many product review platforms thus encourage users to post images. Yet, it is unclear whether these images indeed benefit businesses. We apply image and text mining to investigate how consumer-posted images and review texts help online shoppers improve their purchase decision quality and decision efficiency. We measure the decision quality using review helpfulness votes and product returns. The results indicate that situational cues–the contextual information–in consumer-posted images assist shoppers in better evaluating product functionality. However, visual situational cues cannot illustrate durability–a feature with a temporal dimension that can be better described in the text. We further unveil the economic value of review helpfulness by showing that helpful review content reduces product returns. Every extra consumer-posted image with situational cues decreases the return probability by 20.69% in our context. Regarding decision efficiency, we demonstrate that when shoppers have a higher information processing cost, they prefer to use the image content rather than the text to make purchase decisions. Overall, we suggest that online businesses selectively highlight user-generated content to help consumers make informed purchase decisions that are less likely to be reversed.

Essay 2: Trust Antecedents in Online Reviews Across National Cultures

When reading an online review, consumers rely on the implied trust-building factors within it. Trust antecedents have been extensively researched and categorized into argument quality and source credibility. However, there is still a need to investigate how their help for consumers differs across national cultures. Furthermore, while a profile photo has been found to enhance the helpfulness of a review by improving source credibility, little is known about how consumers perceive user appearance in a review picture and no research has compared these two types of visual information. To bridge these gaps, we examine how trust antecedents and visual information help consumers' decision making differently across national cultures, using reviews collected from Amazon in the US and Germany, respectively. Our results show that information quality is more helpful for consumers in an individualistic and low uncertainty avoidance culture, whereas source credibility is more valued in the opposite culture. Interestingly, the effects of the two types of visual information, user appearance in a profile photo versus in a review picture, also differ across cultures. This research contributes to the online review literature by introducing visual information and national culture with new perspectives of trust. Our findings suggest that multinational e-commerce businesses adopt culture-specific strategies in review systems to better help consumers make informed purchase decisions.

CHAPTER 2

Essay 1: Is a Picture Always Worth a Thousand Words? An Exploration of the Informational Role of Images in Online Product Reviews

2.1 Introduction

The increasingly easy access to smartphones and mobile devices for individuals in the last decade has led to an explosive growth of visual content on the Internet. According to Deloitte Consumer Review (2016), about 2.5 trillion pictures were shared and stored online in 2016. A recent report of consumer surveys by eMarketer indicates that 83% of US online shoppers consult images prior to purchase (King 2018). Perhaps such a staggering number motivates product review platforms to deploy tools to help consumers use and post pictures[1]. For example, Amazon.com prioritizes consumer-posted images on top of the review section (John 2020), and Google has allowed users to add pictures alongside reviews in Google Shopping since 2019 (Bond 2019).

The efforts of the online retail industry to encourage consumers to post images alongside review text should benefit them for reasons grounded in theoretical research. Duan et al. (2009) highlighted the information overload caused by the overwhelming volume of review texts on the Internet. In fact, online shoppers only consider a small set of textual reviews during the purchase process. A 2021 industry survey by Bizrate Insights found that 76.1% of consumers read *less than six* review posts before purchasing (Kavanagh 2021). This number may indicate a typical

[1] The Merriam-Webster dictionary defines a *photo* as a picture obtained by photography and an *image* as a picture produced or modified by a computer. We use *consumer-posted images* or *pictures* to refer to the visual content in product reviews, and we use *photos* only when referring to a picture taken by a camera.

online shopper's limitation in processing text. By contrast, researchers have demonstrated human beings' exceptional ability to detect the meaning of a picture in 13 milliseconds (Potter et al. 2014) and memorize the visual content (Standing et al. 1970). Given these findings, consumers may prefer to use images rather than review texts to gather information for product evaluation.

Consider an example that illustrates how images may better explain product functionality. Figure 2.1 shows a picture posted by a reviewer of a snow jacket. The snow-covered mountain in the background implies a potentially cold environment, which, to some extent, helps assess the warmth of the coat. Also, the user's extreme sport-action posture indicates that the jacket is sufficiently flexible to allow full mobility. A shopper can easily capture these visual cues and use them to evaluate the product's functionality. By comparison, describing these attributes through text often requires exquisite word articulation to offer the detail of the context (e.g., the activity that the user performs) and the evaluation of the attribute (e.g., the user's exact feeling). Such a review not only is a tedious writing exercise for the reviewer but also poses a considerable cognitive challenge for information processing to the reader. Hence, images may have substantial advantages to offer in this scenario.

Figure 2.1. An Example of a Consumer-posted Image

However, not every product attribute can be appropriately depicted by an image. For example, consumers returning online orders often complain that the color of the product that they

receive differs from the one shown on the website (Walk-Morris 2019). Color representation in the images suffers from distortion due to technical issues, including the conditions in which the photo was taken, the quality of the device used to take the shot, the setting of the monitor displaying the image, etc. A dark grey may look like charcoal black if the picture is taken in a dim environment or shown on a screen with a high contrast theme.

Also, images may not be able to illustrate product durability, an attribute with a temporal dimension. Evaluations on durability can be effectively conveyed through text that vividly describes a user's previous and current product experiences and the changes in the product's state over time. An image being still cannot adequately delineate this attribute. In short, the informational role of consumer-posted images (versus that of review texts) depends on the product attributes the shoppers assess. To our knowledge, the extant literature lacks a thorough study on how the image content, vis-à-vis the textual comments, helps consumers evaluate product attributes. We aim to fill this gap in the literature through this study.

Our second goal is to unveil the economic value of consumer-posted images regarding how they affect product returns. Product returns have already been recognized as a significant problem for online retailers costing billions of dollars annually (Sahoo et al. 2018, Wang et al. 2021). The Wall Street Journal estimated that consumers returned $70.5 billion worth of goods purchased during the 2020 Christmas season alone (Petro 2021). Prior research suggests that the visual content may not help increase net sales. According to De et al. (2013), shoppers actively using professionally created photos listed by the retailers' websites on product pages for purchasing are more likely to return the products later. They ascribe this finding to consumers forming unrealistic product expectations by using the impression-based information in the image. Pictures on product pages are displayed mainly to highlight the strengths of product attributes. In

contrast, the images posted alongside reviews are taken by ordinary users who intend to emphasize the actual state and utility of the product by presenting it in a more realistic setting with very limited enhancement and editing to improve the attractiveness of the picture. It is not clear whether these images would have the same effect as the images provided by the sellers. As far as we know, no empirical work has examined the impact of consumer-posted images on product returns. For managers of platforms that promote pictures of products taken by reviewers, it is crucial to understand the role of these images in affecting product returns. Promoting such images may do more harm than good if the impact is positive (i.e., increases returns). On the other hand, if the effect is negative, a proper understanding of the value of consumer-posted images should help design optimal policies and incentives.

We gathered consumer-posted image and product review data from an online outdoor gear and clothing specialty retailer in the US and used them to study the issues raised above. Based on the constructivist theory of visual perception (Boring 1946, Epstein 1973), we extracted *situational cues* – contextual information such as outdoor environment and activity – from consumer-posted images. We expected that such content would help shoppers evaluate the product. For comparison, we also mined situational cues from the review texts. In addition, we identified textual comments on product *functionality*, *aesthetics*, and *durability* in line with the categorization of product attributes by Garvin (1984). We then compared the informational role of consumer-posted images with that of review texts by examining the following: 1) when do shoppers rely on the situational cues in an image rather than those in the text to assess a product attribute (a *substitutive* relationship), 2) when do image and text *complement* each other to facilitate the evaluation of a product attribute, and 3) what product attribute may not be adequately illustrated by an image?

To analyze consumers' utilization of the review content in making purchase decisions, we connected the image and text-mined results through the *review helpfulness votes*. The "helpful review" voting function is widely adopted by product review platforms. Although the review helpfulness is not a direct indicator of sales, it does represent the extent to which the review content facilitates online shoppers' *purchase* decision process (Mudambi and Schuff 2010, Yin et al. 2014). Our findings suggest that situational cues in consumer-posted images attract helpfulness votes. More importantly, they seem to replace the role of situational cues in the text in contextualizing product functionality, e.g., the warmth and flexibility of a jacket. In contrast, when assessing aesthetic attributes, e.g., the color and design of running pants, shoppers need to use situational cues in *both* image and text. Finally, the visual information does not help evaluate product durability, e.g., a tent's long-term performance and reliability.

More importantly, the economic value of review helpfulness is critical for practitioners. Yet, it is understudied. Prior research usually assumes a persuasive role of helpful reviews, i.e., they should positively influence consumers' purchase intention. Yin et al. (2021) pointed out that this is not always the case. While positive comments in helpful reviews might improve shoppers' product attitude, helpful negative content may prevent them from a potentially unpleasant product experience. Our analysis of the impact of review helpfulness on product returns is new and unique in the academic literature. We argue that helpful review content, whether positive or negative, should facilitate consumers to make *informed long-run purchase decisions* – decisions less likely to be reversed. When a shopper buys a product by considering more helpful reviews, she is more likely to figure out that the product fits her requirement. Likewise, when a shopper finds reviews helpful and decides not to place an order for a product, she avoids buying the product that might be unfit for her. In other words, helpful reviews should improve the purchase

decision quality by reducing purchase errors, which are measured on the dimension of product returns.

By modeling product returns as a function of reviews, we found that review helpfulness can negatively affect product returns. A 10% increase in the proportion of helpful reviews with an image is associated with a 6% decrease in the product return probability. Also, every additional picture containing situational cues can reduce the return probability by 20.69%. In contrast, we discovered a positive effect of the sheer number of consumer-posted images on product returns. It means that the visual review content not voted as helpful (i.e., after we partial out the helpful content) might increase purchase errors, thus offering further evidence of the value of review helpfulness.

Our empirical results offer several insightful business implications. First, not all consumer-posted images are beneficial. Review platforms should encourage buyers to post pictures with situational cues, which help contextualize product functionality. Moreover, in addition to the images with situational cues, comments on product aesthetics are helpful for shoppers. Presenting a summary of product durability in the review texts may also help them make informed purchase decisions. Overall, instead of having a blanket policy of suggesting users post pictures, companies should encourage images with important visual content along with the text to improve the shoppers' decision quality. Platforms like Tripadvisor and Yelp provide guidance for posting high-quality reviews (e.g., https://www.tripadvisor.com/Trust-lvBd3L1aU38Y-Review_posting_guidelines.html). Our study is a step toward giving advice on writing good reviews that benefit shoppers and companies.

Theoretically, we unveiled a scenario when the picture superiority effect does not work: images may not adequately delineate product durability. This finding suggests researchers take

the characteristics of product attributes into account when comparing the informational role between image and text. A single framework of visual information processing may not fully explain online shoppers' utilization of user-generated content. Another significant contribution of our study is the advancement of the knowledge about product returns, which cost billions of dollars to online retailers every year. This is especially the case during the pandemic – the return rates during Covid related to stay-at-home orders have skyrocketed and can be as high as 40% (Petro 2021). Helpful review content should reduce returns if it indeed facilitates consumers to make informed purchase decisions.

The remainder of this chapter is organized as follows. We present the literature review and hypotheses development in the subsequent two sections. Next, we describe the research setting and data processing. We then demonstrate the estimations of how the image and textual review content facilitate online shoppers to make purchase decisions, followed by robustness checks and mechanism explorations. The analyses of product returns are reported next. Finally, we discuss the business implications and limitations and point out future directions.

2.2 Literature Review

2.2.1 Text and Image in Online Product Reviews

Research in the literature of product reviews has explored the role of various types of content in the text. Ghose and Ipeirotis (2011) mined linguistic cues such as subjectivity and readability from reviews and found their differential impacts on the helpfulness votes and sales. The emotional expressions can also impact consumers' perceived review helpfulness (Yin et al. 2014, 2017) and purchase intentions (Yin et al. 2021). A more relevant topic to our study is the reviewers' comments on product attributes. Archak et al. (2011) text-mined product feature evaluations of digital cameras, and Liu et al. (2019) applied deep learning models to extract

comments on quality dimensions. Both papers unveiled that the positive valence of the product evaluations can improve sales.

On the other hand, the role of images in the online market has not grabbed enough attention in the business area until recently. De et al. (2013) found that online shoppers' use of seller-offered pictures increases product returns. They ascribed the result to the impression-based information in the professional photos, which would lead to unrealistically high product expectations. Zhang et al. (2021) quantified image attributes (e.g., composition and color) of seller-produced pictures on Airbnb and estimated how they affect property demand. Shin et al. (2020) is the only paper we know that mines both visual and textual content in online posts; they investigated how the content of company blogs on Tumblr affects users' reblogs and likes.

Unlike the extant studies, we extracted a new type of information – *situational cues* – from *consumer-posted images*. We inspected how such contextual cues (e.g., environmental stimuli such as mountain and river) assist online shoppers' product evaluations. Besides, consumer-posted images are likely to provide more authentic information that shoppers seek compared to the pictures posted on the product page by the sellers. Thus, the findings would differ from those in the abovementioned papers that analyze seller-offered images. Finally, instead of analyzing images or text only, we compared the informational role of pictures with that of review texts.

2.2.2 Review Helpfulness and Product Returns

Product review helpfulness, measured as the number or proportion of votes by the consumers who regard a review as a facilitator for their purchase decision-making (Mudambi and Schuff 2010, Yin et al. 2014), has received considerable attention in the information systems field. Mudambi and Schuff (2010) studied how review features affect helpfulness votes and found that

longer reviews tend to be voted as more helpful. Yin et al. (2016) showed that consumers treat a review as helpful if its rating score confirms their initial beliefs on the product. Studies have also analyzed how reviewers' identity disclosure (Forman et al. 2008) and attentional focus (Lei et al. 2021) affect the helpfulness votes. A recent paper by Zhou and Guo (2017) showed that reviews having an image tend to be voted as helpful. Yet, the authors did not extract the content in the pictures and investigate how it assists online shoppers' decision-making, which is the main focus of our study.

Although our analyses also use review helpfulness votes as the outcome, we emphasize *comparing the role of the image content versus review text* in improving consumers' *decision quality* and *decision efficiency*. For decision quality, we demonstrate that the image review content voted as helpful also reduces product returns. On the other hand, when the information processing is costly, consumers may need to use either the image or the text, whichever helps them make decisions more efficiently. Our mechanism exploration tested this conjecture and found shoppers' preference to use the image rather than text to evaluate product functionality under such a condition.

Last but not least, we addressed an understudied research question: *What is the economic value of review helpfulness?* Researchers have found that helpful review content may decrease sales (e.g., Ghose and Ipeirotis 2011, Yin et al. 2021). Hence, the actual benefit of highlighting helpful reviews, as a common practice by review websites, is unclear. We argue that the economic value of review helpfulness cannot be aptly estimated if we focus on its sales impact. Instead, we should analyze how helpful content impacts *product returns*. Consumers making purchases under the influence of helpful reviews are likely to obtain valuable information about certain product attributes' (un)fitness to their preferences. Whether they decide to buy or not, we

should expect a lower return probability in such a scenario since the decision quality increases (Sahoo et al. 2018).

Recent studies have analyzed the effect of product reviews on product returns (e.g., Wang et al. 2021). While positive ratings can decrease the return probability, the number of reviews has a diminishing negative effect on the outcome (Sahoo et al. 2018). We contribute to this growing literature by estimating how the review content voted as helpful affects product returns.

2.3 Hypotheses Development

Our empirical probe builds upon the constructive theory of visual perception, which explains how human beings use the ventral system to process objects in an image using environmental stimuli (Boring 1946, Epstein 1973). Visual object processing differentiates between the *core*, the primary sensory excitation that originates from the focal object the viewer perceives, and the *context*, which includes all the other perceptual cues that can modify the data from the core input (Holway and Boring 1941). In a typical consumer-posted image, the core is the product the reviewer evaluates, and the context contains the background cues in the picture.

2.3.1 Situational Cues in Consumer-posted Images

A reviewer may picture a product in several ways. To illustrate a snow jacket, she may display the product only in front of a plain background (e.g., Figure 2a). Alternatively, she can demonstrate her product experience in a context, e.g., the outdoor environment in Figure 2b. For a shopper, the first picture is unlikely to offer information beyond what she can find from the images by sellers. In contrast, the second picture provides additional details for contextualization, i.e., the placement of the product in a context to simulate how it can be used (Daugherty et al. 2005). For instance, if the shopper is curious about the jacket's warmth, she may find the snow-covered mountain (i.e., the context) especially helpful. We refer to such

contextual information as *situational cues* because they represent the situations (i.e., environment or activity) in which the product was performed.

2a. Without Contextual Cues **2b. With Contextual Cues**

Figure 2.2. Examples of User-Generated Product Images

Online shoppers do not get direct product exposure. In such a setting, situational cues may provide valuable information for product evaluation (Berger and Fritzsimons 2008). On the one hand, research in the psychology literature has shown that situational cues help a person activate associated representations in memory and make them more accessible (Higgins et al. 1977). On the other hand, these cues may create a predictive context that makes information processing easier (Lee and Labroo 2004). A shopper observing the snow-covered mountain in Figure 2b may readily recall her similar experiences in the past and use this cue in her memory to assess the warmth of the jacket. The shopper might also imagine a potential situation in which she wears the coat in a cold environment, thus better understanding how the jacket can keep her warm.

In contrast, shoppers can hardly find reliable contextual information in product descriptions in which sellers usually display pictures of a product in front of a plain background or an artificial setting that may induce unrealistic product expectations (De et al. 2013). Although

they may also find situational cues in the review texts, processing such information is more cognitively costly than interpreting the perceptual cues in the image, a point we will detail next. Overall, the visual situational cues should help online shoppers make more informed purchase decisions.

H1: Consumer-posted images with situational cues are more likely to facilitate consumers' purchase decision process than those without.

2.3.2 The Informational Role of Image Versus Text

Following Garvin's quality[2] categorization (1984), we group the product characteristics into three categories. Based on the nature of these attributes, we hypothesize how the image content, in contrast to the text, helps consumers conduct pre-purchase product evaluations.

The first group of attributes includes performance and features defined by Garvin (1984). Performance consists of the primary traits or functions set by the industry, e.g., the cruising speed of a car. Features are the secondary characteristics that supplement the product's basic functioning, e.g., the number of USB chargers in a vehicle. Both attributes are measured using an objective scale, i.e., improvements in these measures are viewed as translating into quality gains. For example, in our study, most consumers would agree that a higher value of packability of a tent implies better quality. We denote this group of attributes as *functionality*.

Product functionality can be described in both image and text. Yet, verbal expressions have the disadvantage in concretizing this type of attribute through context. Figure 3a shows an example where the reviewer evaluated a tent's functionality by a simple statement: *This is a great all-around tent suitable for extreme conditions and absolute security*. While the comment indicates that the tent functions well even in challenging scenarios, its key message is abstract. A

[2] Essentially, "quality" means attribute by Garvin (1984).

shopper is unlikely to know *what exact functionality attribute(s)* the reviewer was talking about and *in what extreme conditions* the tent performed well. Without a concrete understanding of the context, it is not easy to assess the product. By comparison, the shopper can quickly figure out the extreme conditions in the picture: the tent was raised on dense snow with crusted layers. Even an inexperienced tent shopper could imagine how difficult it is to push the stakes into such a tough ground. Also, the situational cues in the image offer a vivid predictive context in which the tent can hold, thus helping the shopper infer the functionality attributes such as firmness and stability. In short, the visual context communicates more concrete information for functionality assessment.

	3a. 4-Season Tents	**3b. Performance Tights**	**3c. Ski Pants**
Consumer-posted			
Review Text	"This is a great all-around tent suitable for extreme conditions and absolute security…"	"… I would definitely recommend these pants to especially for yoga! …The picture shown is actually a blue space dye, so they're a little lighter than the black/white.…"	"…I was concerned that they would be a bit too tight in the thigh, but I have total mobility, and don't feel like they're going to rip with each step I take. …"

Figure 2.3. Examples of Consumer-posted Images and Textual Comments on Product Attributes

On the reviewer's end, she could have done a better job explaining the tent's functionality in the text. Yet, it often requires exquisite word articulation to provide detailed context (e.g., the dense snow caused by consecutive snowfalls, etc.) and the evaluation of the attribute (e.g., how the pegs and stakes hold the tent). Such a review would not only be a tedious writing exercise for the reviewer but also pose a considerable cognitive challenge for information

processing to the reader. Table 2.1 presents part of the textual comments in a well-written review that describes the product experience of a tent. The reviewer discussed how the tent performs in different weather conditions. While the text effectively demonstrates the product's functionality, the entire review is lengthy (528 words). It might cost a shopper a significant amount of time to process such content.

Table 2.1. An Example of Textual Comments on A Tent

"I took this on a Scout Winter Campout in Northern Michigan. Temperatures were in the teens with Wind-chill close to Zero and some snow. The tent performed well. No condensation on the inside of the tent and It was roomy for myself and gear. The tent breathes extremely well, although it would have been nice if the outer tent went all the way to the ground. Still, I stayed warm and dry. I saw this tent for a ridiculously low price and had to re-evaluate the use for my 4 season tent as I primarily winter camp with my scout troop and will not be doing any alpine camping. That being said, this is a huge winter tent and would be best suited for late Fall and Early Spring camping. I believe it could handle some snow, but not sure about heavy snow or ice unless you shake the tent periodically. I would not recommend taking this into the back country during winter, but on weekend camping at a scout camp or similar. I took it weekend camping in late October where temperatures ranged from 50 degrees to 20 degrees with some wind and this tent kept me warm and dry. There was also no trouble with condensation with everything closed except for the vents in the rainfly."

By contrast, shoppers are likely to process the visual information more efficiently than interpreting the text. Research has shown that human beings have a remarkable ability to extract the concepts of a picture in 13 milliseconds (Potter et al. 2014) and memorize a large amount of image content at an accuracy of 90% (Standing et al. 1970). Considering all these factors, consumers should prefer to use the situational cues in the image to evaluate product functionality.

H2a: The situational cues in the consumer-posted images are likely to replace the contextual information in the review text to help consumers assess product functionality.

The second type of attribute – *aesthetics* – represents how a product looks and feels (Garvin 1984). In our study, the color, size, and design of outdoor products belong to this category. Unlike functionality, these attributes are based on a subjective dimension, i.e.,

consumers use idiosyncratic evaluation criteria. When products are differentiated by aesthetics, there is an ideal point that offers the best value to each consumer (Lancaster 1971). Any movement away from this attribute level will sharply drop the value of the product (Ulrich 2011). For example, a person who likes navy blue may have no interest in a sky-blue shirt. Also, a pair of round-toe pumps would be too tight for a lady who prefers to wear shoes with square toes. These subtle differences of an attribute, if not clarified in the review, may prevent consumers from making accurate purchase decisions.

How do consumer-posted images help shoppers assess product aesthetics? Let's use color as an example. Consumers returning online orders often complain that the color of the product they receive differs from the one shown on the website. Situational cues in the consumer-posted images may help contextualize this attribute in a more realistic environment – in contrast to the artificial setting in the seller-provided photos. Yet, the color representation in the user-generated images may still suffer from distortion due to technical issues. A dark grey might look like charcoal black if the picture was taken in a dim environment or displayed on a screen with high contrast theme. Hence, an elucidation of the attribute in the text, in addition to the visual illustration, could be helpful. Figure 3b shows a user wearing a pair of performance tights doing yoga on a beach. In the text, the reviewer explains the nuance between the actual color she observed and the one on the website, i.e., *(the color) the picture shown is actually a blue space dye, so they're a little lighter than the black/white*, which might help a future buyer better understand this attribute. As another example, Figure 3c illustrates the overall fit of a pair of ski pants to the user while skiing. When evaluating the product design, the reviewer discussed her concern about the tight fit in the thigh (i.e., her feeling about the product, not about the functionality since she is successfully using it). Such a subtle point cannot be accurately depicted

in the picture. Yet, it might be a critical feature for many shoppers who would otherwise be unable to find the information in the image.

The above examples indicate that situational cues in consumer-posted images may not fully depict product aesthetics in that shoppers have various unique requirements. While the visual content can partially illustrate the attributes (e.g., the color of a product in an outdoor setting), consumers might search for additional information in the text to understand whether the nuance (e.g., ruby red versus turkey red) implies a match or mismatch. Thus, we expect a complementary relationship between the image and the review text in helping online shoppers assess product aesthetics.

H2b: The situational cues in the consumer-posted images and the contextual information in the review text are likely to complement each other to help consumers assess product aesthetics.

Our final categorization of product attributes is *durability*, which includes reliability, conformance, and durability described in Garvin (1984). Reliability is the failing of an attribute within a specified period. Conformance is about how a product meets the pre-established standard and is usually proxied as the number of incidences of defeats. Lastly, durability represents the amount of use of a product before it fails to function. While these are also objective attributes, they are measured on a temporal scale. To describe this type of product characteristic, a reviewer needs to verbally discuss the previous and current states of a product, how long the standard quality lasts, what events led to the failure of performance, etc.

An image cannot aptly illustrate the temporal feature since it is still. For example, a user evaluating an ice screw mentioned that *the teeth are still razor sharp after two seasons of climbing on it*. The key message is the product survives after *two seasons* of usage. Although this reviewer posted a picture, it does not help depict the durability. Overall, consumer-posted images

cannot adequately delineate product durability[3]. Instead, shoppers may obtain valuable information in the review text.

H2c: Textual comments on product durability are likely to facilitate consumers' purchase decisions.

2.3.3 Helpful Review Content and Product Returns

A product review is considered helpful if its content facilitates consumers' purchase decision process (Yin et al. 2014). As a result, the quality of the decision improves. In other words, shoppers are likely to obtain valuable information in the helpful reviews and better understand the fit of the product to their preferences. Sahoo et al. (2018) empirically tested this conjecture and revealed a negative effect of helpful reviews on product returns – an outcome that indicates the decision quality. We step further and analyze the impact of helpful review content on product returns.

De et al. (2013) found that online shoppers' use of alternative product pictures offered by the clothing retailer is associated with a higher product return probability. The authors ascribed the result to the impression-based information in the photos – a beautiful model wearing the product in front of a scenic yet artificial setting, which would lead the consumers to form an unrealistic product expectation. In contrast, consumer-posted images are taken by buyers who share their real product experiences with future consumers. The quality of these pictures (e.g., the level of diagonal dominance, rule of thirds, visual balance, etc., based on Zhang et al. 2021) should be much lower than that of the seller-produced images. Thus, the amount of impression-based information in consumer-posted images, if any, should be minimal. On the other hand, as we have illustrated, the situational cues in the consumer-posted images contextualize the

[3] One exception is that an image shows a product malfunction. We will control for this factor in our estimations.

products in more realistic settings. Online shoppers should obtain more factual information from these images and are less likely to form unrealistic product expectations.

Moreover, the informational role of situational cues in consumer-posted images, as we theoretically hypothesized in the previous sections, is to assist consumers in evaluating the product attributes. We argue that the helpful image content should facilitate shoppers to make more *informed long-run purchase decisions*, i.e., decisions that are less likely to be reversed. When the contextual information in an image indicates an excellent performance of a product attribute, the consumer is expected to get persuaded to buy the product. Contrarily, when the visual situational cues imply that a product characteristic does not meet the shopper's needs, she will avoid making a potentially disappointing purchase (which may lead to a return). Both scenarios should lead to lower purchase errors. Accordingly, we propose the following hypothesis.

H3: Consumer-posted images having situational cues negatively affect product returns.

We do not formulate hypotheses about how the helpful content in the review texts affects product returns; yet, in general, we expect its impact on product returns is lower than that of the helpful image content. First, consumers only consider a small set of product reviews before making a purchase decision (Yin et al. 2014). The helpful textual comments, if not highlighted, are unlikely to be exposed to each shopper who is often overwhelmed by the product information on the Internet (Duan et al. 2009). Hence, the helpful content in the text may not have a salient effect on product returns. In contrast, images are more likely to grab viewers' attention (Peschel and Orquin 2013). Besides, the picture superiority effect suggests that pictures communicate more information and are remembered more (Childers and Houston 1984). Such advantages may overcome the high information search or processing cost (e.g., the cost of searching for older

posts in Ghose et al. 2013), making consumers prioritize the images and use the visual content to conduct pre-purchase product evaluation. Overall, we expect to observe that *helpful image content is more influential than helpful textual content on product returns.*

2.4 Data Processing

2.4.1 Data Collection

We procured the data from an outdoor specialty online retailer in the US. The products sold on the website are clothing (e.g., snow jackets, running pants, etc.), footwear (e.g., climbing shoes), and gears (e.g., snowboards, tents, etc.) mainly designed and used for outdoor activities. The firm provided us with the original product review data on March 21, 2016. On March 3, 2019, we supplemented the data by collecting the consumer-posted images from its review platform. Altogether, we obtained data for 2,541 products available on the website across these three years. We then extracted corresponding 9,435 reviews for this study.

Each product review in the raw data is associated with a unique review ID, a consumer ID, and a product ID. Besides, each post has a review title, review text, rating score, the posting timestamp, the badge of the reviewer (if the reviewer is an employee at the retailer or a vendor representative), and the cumulative number of helpfulness votes up until the time of data collection. We saved the unstructured image data in a separate file. Because the platform allows users to upload *only one* picture when posting a review, each image is associated with a unique product review. Among the 9,435 reviews, 857 contain an image.

The review platform did not add the image uploading function until 2010. Thus, we dropped the data before 2010 to avoid biased estimation caused by the absence of visual information during this period. In addition, we excluded the category of accessories (e.g., bike nutrition, ski waxes) and underwear in that these products are not likely to be aptly pictured for

contextualization. The final sample for the main estimations contains 7,499 reviews associated with 2,123 unique products. These reviews were posted between January 2010 and March 2016. About 10.29% of them have an image.

2.4.2 Image Mining Procedure

Our image mining employed the *IBM Watson Visual Recognition*[4], which offers built-in models pre-trained on tens of millions of pictures by deep learning algorithms. Past research has also used Watson computers for feature extraction (e.g., Song et al. 2021). We applied the object classification function to extract the background information that represents the situational cues. The output of the algorithm is multi-level. For example, for the picture in Figure 3c, the model returns three distinct class labels for the detected objects. First, it extracts the user and labels it as *person/contestant/athlete/skier*. Second, it identifies the person's activity and outputs *sport/skiing*. Lastly, the downhill in the background was captured and classified as *nature/slope/ski slope*. The extracted content meets our research requirement.

Because the products in our data are mainly designed for outdoor activities, we extracted four groups of high-level objects that may contain situational cues for product contextualization in the outdoor environment[5]: *nature*, *plant*, *sport*, and *defensive structure*. The first two classes are likely to offer a natural environment context. For instance, the appearance of creeks or hills, forests, or grass in the background would help an observer understand the surrounding conditions in which the product is consumed. Shoppers can use these perceptual cues to assess attributes such as the waterproofing of a pair of running shoes and the color of a shirt (in contrast to the grass). The third class represents the outdoor activities a user carries out while using the product,

[4] https://www.ibm.com/watson/services/visual-recognition

[5] Most products presented indoor have a plain background that does not help contextualize products. We will conduct a robustness check in the mechanism exploration and show that these pictures do not help consumers make purchases.

e.g., jogging and kayaking. Consumers can rely on this type of situational cues to evaluate product features, such as the flexibility of a training shirt and the ease of using a paddle. Lastly, defensive structure refers to objects like a cliff when the user is climbing, which may help the shoppers assess attributes such as the firmness of a climbing rope. Appendix A provides a complete list of the detected objects.

For each user-generated image, as long as the objects detected by the visual recognition function belong to one of the four classes, we used an indicator *Image: Situational Cues* to denote it as having contextual information. To evaluate the classification performance, we conducted manual labeling. In particular, we randomly selected 100 images and assigned them to three coders; each was asked to do the labeling independently by identifying situational cues according to the definitions of the four categories (i.e., nature, plant, sport, and defensive structure). We then computed the inter-rater agreement. The pairwise Cohen's kappa values are around 98%, suggesting a high consistency among the coders. We used the majority voted results to estimate the model performance and obtained a precision of 96.15% and a recall of 78.90%.

2.4.3 Text Mining Procedure

To compare the role of image content with that of the review text, we text-mined the contextual information and the comments on specific product attributes.

2.4.3.1 Text Processing and Situational Cue Extraction

First, we conducted a part-of-speech tagging in R using the package *UDPipe*. By parsing the raw text (e.g., removing stopwords and transforming letters to lower case), we built a dataset comprising a review ID, a Sentence ID, a Token ID, the token, and the lemma. We then included tokens that appear at least five times in the review text and obtained a list of 4,179 unique terms.

Next, we scrutinized the tokens and found terms that imply situational cues for product contextualization. Analogous to the process for image mining, if a word refers to the natural environment (e.g., mountain, rainstorm), or the user-engaged activities (e.g., hiking, camping), we treated it as a situational cue. We accounted for all different parts of speech (i.e., nouns, verbs, adjectives, and adverbs) as long as the meaning of the word meets the definition of situational cues. For example, both the verb *climb* and the noun *climbing* are considered an indication of outdoor activity. After this procedure, we produced a list of 283 unique terms (detailed in Appendix B) that imply the context of product consumption. We then created a variable by counting the number of sentences containing situational cues and naming it *Text: Situational Cues*.

2.4.3.2 Comments on Product Attributes

Following our categorization of the product attributes based on the seminal work by Garvin (1984) in the previous section, we extracted the textual comments on specific product characteristics.

For product functionality (*Text: Functionality*), we identified the keywords related to an outdoor product's primary and secondary functions. For example, if a review text includes the discussions on the *breathability* of an insulated jacket or the *extendibility* of a climbing rope, we will treat it as having comments on product functionality. Product aesthetics (*Text: Aesthetics*) are about the looks and feels of the products. In line with this definition, we built a word list, including the terms such as *fit*, *snug*, *fashionable*, *color*, etc. The attribute *Text: Durability* refers to how long the expected quality lasts and if any malfunction occurs during the consumption. Accordingly, we extracted keywords such as *abrasion*, *dent*, *reliable*, etc. All of these variables

measure the number of sentences containing the relevant keywords in a review. Appendix B provides a complete wordlist and additional notes for this text mining task.

We checked the accuracy of the text-mined results by randomly selecting 1,000 sentences from the reviews and requesting the coders to do the manual labeling. To make the findings comparable to those of the image mining tasks (i.e., the binary outcome), we computed the accuracy at the level of an expression, i.e., whether a detected term belongs to the pre-defined category. Based on the calculations, we obtained the following precision and recall statistics: 85.71% and 97.18% for situational cues; 96.25% and 98.09% for functionality; 90.05% and 96.88% for aesthetics; and 98.15% and 89.83% for durability.

2.5 Facilitator of Purchase Decision Making

An informative product review should facilitate consumers' purchase decision process from two perspectives. First, the content should help them evaluate products effectively, i.e., it helps produce desirable outcomes. In this section, we use the review helpfulness votes to analyze the decision effectiveness because this measure indicates how consumers treat the review content as assisting their decision making (Mudumbi and Schuff 2010, Yin et al. 2014). Second, the content should ameliorate decision-making efficiency, e.g., by reducing the burden caused by information overload. To examine this dimension of the informational role, we will study how consumers vote for the review content in two scenarios associated with a higher information processing cost.

2.5.1 Model Specification

We model the effects of image and text in a review *r* issued for a product *i* on its votes for review helpfulness by applying equation (1).

$$\ln(Helpfulness_{ri}) = \beta^{X} X_{ri} + \gamma^{Z} Z_{ri} + \delta^{R} R_{ri} + \varphi_i + \varepsilon_{ri}, \quad (1)$$

where X_{ri} is a vector that includes the variable of interest – *Image: Situational Cues* – and a dummy variable that indicates the *Presence of Image* (i.e., whether a review has an image or not); Z_{ri} includes the textual counterparts – *Text: Situational Cues*, as well as the measures of comments on product attributes, i.e., *Text: Functionality*, *Text: Aesthetics*, and *Text: Durability*. β^X and γ^Z are the vectors of the estimates for X_{ri} and Z_{ri}. R_{ri} consists of the basic product review features, including *Rating* (rating score of the review), *Days Elapsed* (number of days since the review was posted), *Existing Reviews* (number of reviews on the product webpage before the focal review was posted), *Review Length* (number of words), and *Sentiment* (average sentiment score). These control variables were derived based on extant studies (e.g., Mudambi and Schuff 2010, Cao et al. 2011, Yin et al. 2014). δ^R is a vector of the estimates that correspond to the vector R_{ri}. Product fixed effects are denoted by φ_i. Lastly, ε_{ri} represents the unobserved error term. In addition, we include calendar year dummies to account for the temporal trend of consumers' voting behavior on the platform. Table 2.2 details all the variables in this study.

Since consumers in our research setting can vote for a review as helpful *only*, the dependent variable is a count measure, i.e., the cumulative number of votes for review helpfulness up until the time of data collection. Our initial model testing indicated that the outcome is slightly under-dispersed. Accordingly, we apply a Poisson estimation. Also, we adopt the robust standard errors to control for mild violation of the distribution assumption that the variance equals the mean (Cameron and Trivedi 2013).

Table 2.2. Variable Explanation and Descriptive Statistics

	Names	Explanations	Min.	Max.	Mean	Std. Dev.
Main Estimations of Review Helpfulness	Helpfulness	Cumulative number of votes for review helpfulness up until the date of data collection	0	11	0.44	0.89
	Rating	The rating score of the review	1	5	4.45	0.93
	Days Elapsed	Number of days between the review posting date and the date of our data collection	0	2271	628.60	553.22
	Existing Reviews	Number of existing reviews on the product webpage prior to the arrival of the focal review	0	17	2.85	3.05
	Review Length	Number of words in the review	1	573	64.86	58.41
	Sentiment	The average sentiment score of the review	-0.78	2.06	0.21	0.18
	Presence of Image	Whether there is an image in the product review or not	0	1	0.10	0.30
	Image: Situational Cues	Whether the image contains situational cues or not	0	1	0.06	0.23
	Text: Situational Cues	Number of sentences that contain the situational cues	0	30	2.00	2.64
	Text: Functionality	Number of sentences that comment on the product functionality	0	7	0.35	0.69
	Text: Aesthetics	Number of sentences that comment on the product aesthetics	0	12	0.69	1.10
	Text: Durability	Number of sentences that comment on the product durability	0	6	0.04	0.24
Robustness Checks and Mechanism Explorations	Helpfulness After16	Number of helpful votes between March 21, 2016, and March 3, 2019	0	8	0.19	0.57
	Non-Consumer Badge	Whether the reviewer has a specific identity, i.e., an employer or a vendor representative	0	1	0.20	0.40
	Image: Appearing of User	Whether the user appears in the image or not	0	1	0.05	0.22
	Image: Product Malfunction	Whether the image shows product malfunction	0	1	0.003	0.06
	Image: Indoor Illustration	Whether the image shows the product in an indoor environment	0	1	0.02	0.14
	Text: User Appearance	Number of sentences that contain the identity disclosure cues (e.g., age, gender, etc.)	0	5	0.09	0.36
	Past Reviews	Number of reviews that the reviewer has posted on the website up until the focal review	0	27	1.26	2.75
	Past Helpful Votes	The average number of helpful votes that the reviewer's past reviews had gathered	0	10	0.20	0.59
	Consumer Tenure	Number of days since the consumers first purchase on the website	0	2794	120	298.64
	Readability	Review readability (SMOG Index)	3.13	29.31	8.44	2.45
	Subjectivity	Review subjectivity	0	37	4.6	3.74
	Spelling Errors	Number of spelling errors in the review	0	36	1.2	1.82

Notes: 1) Following the literature (e.g., Mudambi and Schuff 2010), we take the natural logarithm for review length. 2) For the review text measures, we do not take logarithms in order to directly compare the estimates of these count variables with those of the image measures. We have tested models using the logarithm of these variables and obtained qualitatively the same results.

2.5.2 Main Results

Table 2.3 reports the main results. Model (1) has all the basic review features and the image content measures. Consistent with the finding in Zhou and Guo (2017), the significant estimate of *Presence of Image* suggests that this factor can positively affect review helpfulness votes, which seems to indicate consumers' preference to use visual information to make purchases. The coefficient of *Image: Situational Cues* in model (1) is also positive and significant, indicating that the visual situational cues attract helpfulness votes. Yet, such a result does not tell us how the visual content helps consumers evaluate product attributes. Also, we are not clear whether this image variable truly represents the contextual information we aim to measure. We investigate the first question in this section.

Table 2.3. Review Content that Facilitates Consumers' Purchase Decision Process

	(1) Image Measures	(2) Image and Text Measures	(3) Texts on Product Attributes	(4) Texts with Situ. Cues	(5) Texts Only	(6) Image × Texts
Rating	0.0849** (0.0283)	0.0846** (0.0285)	0.0830** (0.0284)	0.0812** (0.0285)	0.1138*** (0.0289)	0.0868** (0.0287)
Days Elapsed	0.0007** (0.0002)	0.0007** (0.0002)	0.0007** (0.0002)	0.0007** (0.0002)	0.0006* (0.0002)	0.0006** (0.0002)
Existing Reviews	0.0167 (0.0182)	0.0168 (0.0181)	0.0155 (0.0181)	0.0171 (0.0181)	0.0153 (0.0182)	0.0180 (0.0180)
Review Length	0.5586*** (0.0356)	0.5562*** (0.0425)	0.5209*** (0.0457)	0.5156*** (0.0457)	0.5482*** (0.0470)	0.5239*** (0.0457)
Sentiment	-0.5022** (0.1668)	-0.5021** (0.1667)	-0.5206** (0.1667)	-0.5104** (0.1666)	-0.4912** (0.1704)	-0.5198** (0.1659)
Presence of Image	0.7661*** (0.0575)	0.7661*** (0.0575)	0.7647*** (0.0574)	0.7550*** (0.0575)		0.5992*** (0.0851)
Image: Situational Cues	0.2307* (0.1050)	0.2301* (0.1050)	0.2201* (0.1048)	0.2176* (0.1045)		0.2172* (0.1087)
Text: Situational Cues		0.0011 (0.0083)	-0.0005 (0.0083)	-0.0015 (0.0084)	0.0001 (0.0085)	-0.0030 (0.0084)
Text: Aesthetics			0.0233 (0.0213)			
Text: Aesthetics with Situ. Cues				0.0772 (0.1040)	-0.0204 (0.1026)	0.0015 (0.1038)
Text: Aesthetics without Situ. Cues				0.0224 (0.0213)	0.0268 (0.0212)	0.0229 (0.0212)
Image: Situ. Cues × Text: Aes. with Situ. Cues						1.2285** (0.4697)
Text: Functionality			0.0454 (0.0304)			

Text: Functionality with Situ. Cues				0.1542** (0.0550)	0.2280*** (0.0583)	0.1368* (0.0640)
Text: Functionality without Situ. Cues				0.0202 (0.0327)	0.0210 (0.0333)	0.0221 (0.0325)
Image: Situ. Cues × Text: Fun. with Situ. Cues						0.0527 (0.1196)
Text: Durability			0.1264^{+} (0.0758)	0.1305^{+} (0.0750)	0.1308^{+} (0.0734)	0.1241^{+} (0.0739)
Review Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
Product Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	7,499	7,499	7,499	7,499	7,499	7,499
AIC	13272	13274	13274	13272	13443	13270
Note: Robust standard errors are shown in parentheses. $^{+}$ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001.						

After we include the situational cues in the review text in model (2), we find that such information in the written format does not affect the outcome. Yet, presenting the contextual information in the text without discussing product attributes may not be very meaningful. To better understand the role of verbal comments, we add the text measures of the three groups of product attributes – *functionality*, *aesthetics*, and *durability* – to the estimation. As shown in model (3), only the coefficient of *Text: Durability* is marginally significant. A possible explanation of these estimates is that consumers need to read a complete sentence to cognitively assess the product using the combination of situational cues and product feature evaluations. Hence, separately measuring the two types of verbal content may not be appropriate.

To address this issue, we split the review text measures of each attribute into two parts: one part counts the number of sentences having comments on both product features and situation cues; the other part counts the number of sentences containing opinions on product attributes only (i.e., without situational cues). For example, for text on aesthetic attributes, we replace the original measure with two variables in which *Text: Aesthetics with Situ. Cues* represents the number of sentences including keywords of both product aesthetics and situational cues, and *Text: Aesthetics without Situ. Cues* is the number of sentences having comments on product

aesthetics without situational cues. The same rule applies to the measures of functionality. Very few evaluations on durability are associated with situational cues, probably because this attribute does not need to be explained in a context (but rather with proper temporal cues). Thus, we keep the original measure (i.e., *Text: Durability*) due to the lack of variation when we split it. Model (4) includes these newly created variables. The results suggest that textual comments on functionality with situational cues can positively affect the outcome. It means when a functionality attribute (e.g., the warmth of a jacket) is verbally described in a context (e.g., the cold weather), shoppers are likely to find such an evaluation helpful and thus make informed purchase decisions. The estimate of review text about durability remains significant in model (4). Lastly, the coefficients of comments on the aesthetic features (with or without situational cues) are still insignificant.

In model (5), we exclude all image measures. The estimate of *Text: Functionality with Situ. Cues* in this model is 0.2280. In contrast, the inclusion of image variables in model (4) makes the coefficient drop to 0.1542 (by 32.27%). Albeit inconclusive, this decrease likely implies that the informational role of visual situational cues substitutes that of the review text in describing product functionality. Shoppers prefer to use contextual information in the images rather than their textual counterparts to evaluate functionality attributes.

In model (6), we include the interaction term between *Image: Situational Cues* and *Text: Aesthetics with Situ. Cues* to examine whether the two types of content complement each other and help consumers make purchases. The estimate is positive and significant, thus supporting our hypothesis of the complementary relationship between the visual and textual content in illustrating product aesthetics. We also take the interaction between *Image: Situational Cues* and *Text: Functionality with Situ Cues* and obtain an insignificant estimate. It means such a

complementary relationship does not work for evaluating product functionality. We do not take the interaction between the visual situational cues and the durability text measures because, as we explained before, these two types of content are not related. The estimates of *Text: Durability* in the last three models in Table 2.3 are about the same, indicating that comments on product durability are not affected by the image content. They have an independent effect on the helpfulness votes.

In summary, the results across all models in Table 2.3 support H1. On the other hand, the comparison between the estimates of the image and text measures in models (4) and (5) shows that the visual situational cues seem to replace the verbal comments on product functionality, thus supporting H2a. The estimates of the interaction terms in model (6) suggest a complementary relationship between the situational cues in the image and the textual comments on product aesthetics, which is consistent with H2b. Finally, the estimates of *Text: Durability* across the last three models in Table 2.3 support H2c. We step further in the following two sections to consolidate the results regarding all these hypotheses.

2.5.3 Robustness Checks

The above estimations may suffer from endogeneity because of 1) measurement errors of the image and text-mined results (Yang et al. 2018) and 2) unobservables that affect both the variables of interest and the outcome. In the second case, for example, an experienced reviewer may tend to post images; in the meantime, her reviews may receive more votes due to her writing skills. The impact of the image content on the outcome might be overestimated in such a scenario.

To mitigate the concern of potential endogeneity, we start by including a few additional control variables. First, using the reviewer's badge information, we denote whether a review was

posted by a non-consumer (i.e., an employee at the retailer or a vendor representative). Second, we follow Ghose and Ipeirotis (2011) and include linguistic cues – *Readability*, *Subjectivity*, and the number of *Spelling Errors*. Third, we add variables that proxy reviewers' characteristics. For each reviewer, we compute the number of *Past Reviews* she has posted before the focal review, the average number of helpfulness votes per review that she has written in the past (*Past Helpful Votes*), and the number of days since her first purchase at the online retailer (*Consumer Tenure*). In addition, we create two image measures – whether the user appears in the image (*Image: Appearing of User*) and if the image shows product malfunction (*Image: Product Malfunction*). The first measure proxies identity disclosure in the pictures, increasing the review's credibility and obtaining more review helpfulness votes (Forman et al. 2008). The second variable indicates a particular type of illustration of product durability – the product defect (e.g., a broken helmet or a cloth that pilled). Lastly, we include a text measure that denotes the user's identity disclosure (*Text: User Appearance*). Appendix C details these image and text analyses.

We then apply two approaches to robustly address this concern – propensity score matching (PSM) and simulation extrapolation (SIMEX). First, we employ the propensity score matching method to mitigate the systematic difference between reviews with an image and reviews without images. Specifically, we run a logit model in which the outcome is whether a product review contains a picture or not. The explanatory variables are the features of previous reviews aggregated at the product level when a reviewer posts the focal review, e.g., the average rating of all the existing reviews before the current post[6]. We calculate the propensity score for each review post based on this estimation. Next, we match *product reviews with an image* using the nearest neighbor algorithm to search for the *reviews with no image* that have the most similar

[6] We cannot include measures of the focal review because these covariates can be affected by the presence of image.

propensity scores. Such a procedure helps us create an experimental-like setting in which the review with a picture appears randomly, regardless of the features of the existing reviews on the product web page. To maximize the potentially matched units, we set the ratio as 1:6, i.e., a product review with an image is paired with the six most similar reviews without pictures. Besides, to get better-matched results, we allow replication (i.e., a treated unit can be matched with multiple control units) and discard instances if no suitable matched units can be found. After these steps, we obtain a sample of 5,152 observations; 736 are reviews with an image, while 4,416 are reviews without images. We will use this matched sample for the remaining analyses of review helpfulness.

Second, we conduct measurement error correction. Because the number of pictures in our data is moderate, we request the coders to label all of them and use the true values for robustness checks. To account for the error of the text measures, we adopt the SIMEX model proposed by Cook and Stefanski (1994). This approach estimates the error variance using a sample of manually labeled true values and extrapolates the simulated errors. Using the 1,000 coded instances to assess the error variance, we run 1,000 bootstrapping estimations and set the fitting method as linear[7].

The estimations reported in Table 2.4 are the SIMEX estimations using the matched sample. Models (1) and (2) in Table 2.4 have the same specifications as model (5) and model (4) in Table 2.3, respectively, except for the inclusion of additional control variables. Comparing these two models, we find that the coefficient of *Text: Functionality with Situ. Cues* becomes insignificant after including the image measures, indicating that the visual content fully

[7] Alternative specifications using increasing labeled samples, bootstrapping loops, and other fitting methods produce qualitatively the same results as the proposed model specification.

substitutes the text on product functionality. Such a result further supports H2a. Model (3) is analogous to model (6) in Table 2.3, in which we include the interaction terms. The estimate of the interaction between *Image: Situational Cues* and *Text: Aesthetics with Situ. Cues* remains positive and significant, thus further supporting H2b. Moreover, after we control for the illustration of product malfunction in the image and apply the SIMEX and PSM methods, the significance level of the positive estimate of *Text: Durability* increases, suggesting a more substantial impact of product durability evaluations in the review text on the review helpfulness votes. Hence, we consolidate the finding regarding H2c.

Table 2.4. SIMEX Estimation Using the PSM Method

	(1) Texts Only	(2) Full Spec.	(3) Interaction
Rating	0.1345*** (0.0263)	0.1068*** (0.0263)	0.1069*** (0.0263)
Days Elapsed	0.0007** (0.0002)	0.0007** (0.0002)	0.0006** (0.0002)
Existing Reviews	-0.0270** (0.0086)	-0.0292*** (0.0085)	-0.0286*** (0.0085)
Review Length	0.4157*** (0.0548)	0.4394*** (0.0547)	0.4389*** (0.0547)
Sentiment	-0.7362*** (0.1562)	-0.7830*** (0.1558)	-0.7893*** (0.1559)
Presence of Image		0.2430* (0.1029)	0.2337* (0.1030)
Image: Situational Cues		0.4480*** (0.1134)	0.4415*** (0.1121)
Text: Situational Cues	0.0054 (0.0080)	-0.0090 (0.0079)	-0.0157 (0.0106)
Text: Aesthetics with Situ. Cues	-0.0997 (0.1277)	0.0293 (0.1294)	-0.2673 (0.1827)
Text: Aesthetics without Situ. Cues	0.0055 (0.0171)	0.0068 (0.0170)	0.0189 (0.0228)
Image: Situ. Cues × Text: Aes. with Situ. Cues			0.8984* (0.3800)
Text: Functionality with Situ. Cues	0.1382* (0.0656)	0.0043 (0.0519)	0.0474 (0.0869)
Text: Functionality without Situ. Cues	0.0132 (0.0241)	0.0291 (0.0242)	0.0253 (0.0305)
Image: Situ. Cues × Text: Fun. with Situ. Cues			0.0138 (0.1221)
Text: Durability	0.1501** (0.0531)	0.1326* (0.0536)	0.1321* (0.0587)

Non-Consumer Badge	-0.2001** (0.0690)	-0.2276*** (0.0687)	-0.2200** (0.0687)
Readability	0.0274** (0.0105)	0.0337** (0.0105)	0.0315** (0.0105)
Subjectivity	0.0131+ (0.0073)	0.0114 (0.0073)	0.0120+ (0.0073)
Spelling Errors	0.0048 (0.0113)	-0.0214* (0.0101)	-0.0209* (0.0101)
Past Reviews	0.0137 (0.0087)	0.0171* (0.0087)	0.0172* (0.0087)
Past Helpful Votes	0.3822*** (0.0178)	0.3681*** (0.0178)	0.3648*** (0.0179)
Consumer Tenure	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)
Image: Appearing of User		0.0439 (0.0929)	0.0495 (0.0934)
Image: Product Malfunction		0.7758*** (0.2014)	0.7878*** (0.2016)
Text: User Appearance	-0.0670 (0.0482)	-0.0749 (0.0479)	-0.0637 (0.0659)
Review Year Effects	Yes	Yes	Yes
Product Fixed Effects	Yes	Yes	Yes
N	5,152	5,152	5,152
$^{+}$ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001.			

With respect to the newly added control variables, we find that the more votes a reviewer has gathered in the past, the more likely her current post would be voted as helpful. It means that a good review writer would keep producing quality posts. Consistent with the results in Ghose and Ipeirotis (2011), the readability score of a review can positively affect the outcome, while the number of spelling errors has an opposite impact. Besides, pictures with product malfunction are more likely to be voted helpful. Shoppers would notice the low quality of these products and avoid purchasing them. Finally, reviews posted by employees and vendors (i.e., *Non-Consumer Badge*) receive fewer helpfulness votes, probably because the posts from the sellers are promotional (e.g., they only highlight the good part of a product without showing its drawbacks). Consumers tend to have lower trust in these reviews. The other covariates do not seem to affect the outcome.

2.5.4 Mechanism Exploration

In this section, we explore three questions related to the mechanism of the main findings. First, in what situations would the image content help consumers make decisions more efficiently (in terms of easing the burden of information overload)? Second, are there any alternative mechanisms that better explain the complementary relationship between image and text in helping consumers assess product attributes? Third, does our image-related variable (i.e., *Image: Situational Cues*) measure the contextual information in the pictures it is supposed to do?

2.5.4.1 Picture Superiority Effect and Decision Efficiency

The votes for review helpfulness tell us that the consumers find the review, which includes image content, helpful in making effective purchases, i.e., successfully deciding whether to purchase or not. Yet, we were unable to test how the visual information improves decision efficiency, notably achieving the goal with minimum effort. According to the picture superiority effect (Childers and Houston 1984), online shoppers should prioritize using the information in the consumer-posted images to evaluate the products since they can process the visual content with a lower effort. Such cost-saving decision-making should especially be the case when information processing is demanding. We aim to test this conjecture in two situations when online shoppers have a higher cost than usual to process the review content.

In scenario A, we set a minimum number of *Existing Reviews* prior to the focal review. Shoppers may experience information overload since they are overwhelmed by the review posts (Duan et al. 2009). In scenario B, we change the outcome to the number of helpfulness votes collected after March 21, 2016 (*Helpfulness After16*). The reviews in the previous analyses became old to the consumers who viewed the webpage after our first data collection period. Considering that our data provider *only* displays reviews anti-chronologically, the shoppers are

less likely to read these early posts since they need to scroll down the webpage to find them, implying a high information search cost (Ghose et al. 2013). If the image content becomes more influential in these two scenarios, it means such information can improve shoppers' decision efficiency. The first two columns in Table 2.5 report the proposed analyses. Note that we apply SIMEX and PSM approaches to all the estimations in this Table.

Table 2.5. Mechanism Explorations

	Substitution		Complement		Falsification Checks	
	(1) More Posts	(2) Older Posts	(3) Attention	(4) Supplement	(5) Accessories	(6) Indoor Context
Image: Situational Cues	0.6610*** (0.1601)	0.5508** (0.1837)	0.5021*** (0.1122)	0.4255*** (0.1203)	-0.6591 (0.5089)	0.4486*** (0.1276)
Image: Indoor Illustration						-0.3786* (0.1886)
Text: Situational Cues	0.0244 (0.0165)	0.0006 (0.0112)	-0.0095 (0.0078)	-0.0094 (0.0078)	0.0186 (0.0286)	-0.0091 (0.0079)
Text: Aesthetics			-0.0088 (0.0194)	-0.0029 (0.0188)		
Text: Aesthetics with Situ. Cues	0.0779 (0.3092)	0.0459 (0.2048)			0.0112 (0.3717)	0.0275 (0.1295)
Text: Aesthetics without Situ. Cues	-0.0085 (0.0344)	0.0177 (0.0234)			-0.0452 (0.0559)	0.0058 (0.0170)
Text: Functionality			0.0352 (0.0280)	0.0075 (0.0279)		
Text: Functionality with Situ. Cues	0.0847 (0.1179)	0.1086 (0.0803)			-0.0340 (0.1659)	0.0050 (0.0518)
Text: Functionality without Situ. Cues	0.0177 (0.0204)	0.0267 (0.0378)			-0.0334 (0.1077)	0.0285 (0.0242)
Text: Durability	0.2120** (0.0751)	0.3295*** (0.0795)	0.1820** (0.0617)	0.1353* (0.0538)	-0.0803 (0.2630)	0.1312* (0.0536)
Presence of Image × Text: Aesthetics			0.0514 (0.0391)			
Presence of Image × Text: Functionality			-0.0561 (0.0744)			
Presence of Image × Text: Durability			-0.1751 (0.1227)			
Image: Situational Cues × Text: Aesthetics				0.0033 (0.0338)		
Image: Situational Cues × Text: Functionality				0.0029 (0.0420)		
Past Helpfulness Votes		0.2137*** (0.0245)				
Review Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
Product Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	2,799	5,152	5,152	5,152	1,371	5,152

Note: Due to space limitations, we report the estimates of the image and text content measures only. All the control variables in Table 2.3 are included in these estimations.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Model (1) tests scenario A, in which we only include the observations with a minimum of three existing posts – the median value of this variable in the data. The results confirm our conjecture. Compared to the estimates in model (2) in Table 2.4, the coefficient of *Image: Situational Cues* is much larger (i.e., 0.6610 versus 0.4480). We have also run a model using a sample of reviews with at least five existing reviews and obtained an even more substantial coefficient change, which is logical as the information processing cost increases. Model (2) in Table 2.5 examines scenario B. In addition to the existing variables, we control for *Past Helpfulness Votes* – the number of helpfulness votes the focal review received up until the first time of data collection. This covariate helps us rule out the concern that helpful reviews attract more votes in the future. The findings are similar to those in the first model, i.e., consumers tend to process (and vote) an older post if it contains visual situational cues. Overall, these findings demonstrate that the situational cues in consumer-posted images not only help shoppers make purchase decisions more effectively but also improve their decision efficiency.

2.5.4.2 Alternative Mechanisms of the Interaction between Image and Text

We then investigate a couple of alternative explanations of a complementary relationship between image and text. First, Peschel and Orquin (2013) proposed the attentional mechanism and found that pictures draw consumers' attention to the center of an ad before they read the verbal content. If this is the case in our study, consumers facing a group of posts may tend to read reviews with an image. As a result, the textual comments in these posts, in contrast to reviews with text only, should gather more votes. We examine this assumption by interacting *Presence of Image* with the measures of textual comments on functionality, aesthetics, and durability. The estimates of all interaction terms in model (3) in Table 2.5 are insignificant, thus failing to support this conjecture.

Second, consumers might use the overall evaluations on a product attribute in the review text and connect them with the situational cues in the images to make purchase decisions. In other words, we should not interact the visual situational cues with the textual comments that also contain the contextual information. This idea is consistent with Adaval and Wyer (1998) and Adaval et al. (2007), in which these authors revealed that the overall narrative description supplements the pictorial illustration in an ad. To examine such a complementary relationship, we take interactions between *Image: Situational Cues* and the textual comments on functionality and aesthetics without splitting them into the part with situational cues and the part without situational cues. As before, we do not interact visual situational cues with comments on durability. Neither of the two interaction terms in model (4) of Table 2.5 obtains significant results.

Comparing the findings in these two models with the significant estimate of the interaction between *Image: Situational Cues* and *Text: Aesthetics with Situ. Cues* in model (2), Table 2.4, we may conclude that shoppers vote a review as helpful only when they find the visual situational cues and the verbal comments on aesthetic attributes are relevant.

2.5.4.3 Validity of the Image Measurement

Our image mining extracts the objects in the pictures. Yet, we are not sure whether the pre-defined classes (i.e., *nature*, *plant*, *sport*, and *defensive structure*) indeed represent the situational cues – the construct we operationalize. We conduct two falsification checks to validate the measurement.

First, we replicate model (2) in Table 2.4 using the excluded product reviews of accessories (e.g., nutrition bars, ski waxes) and underwear. We expect that the situational cues of the outdoor setting should not help contextualize features for these products. If the extracted

objects from the images are proper operations of the concept for the outdoor goods, we should not get a significant estimate of *Image: Situational Cues* in this estimation. The result in model (5), Table 2.5, supports this conjecture. Second, we use a manually labeled variable *Image: Indoor Illustration* to indicate that a picture displays the product indoor. These images do not seem to offer additional value for product contextualization since shoppers can find the same information in the product description. The negative coefficient of this variable in model (6), Table 2.5, indicates that reviews with such image content are less likely to be voted as helpful, perhaps because consumers feel the information in these pictures is redundant. In short, the falsification checks help us validate that 1) situational cues of a setting do not help contextualize products that do not fit such a context, and 2) illustrating products in an inadequate context does not facilitate consumers' decision-making.

2.6 Review Content and Product Returns

The estimations of review helpfulness demonstrate how the helpful product review content assists online shoppers in making effective purchase decisions. Yet, there are two major concerns in these analyses. Methodologically, only self-selected consumers choose to vote for a review. If the silent majority (i.e., those who do not vote) consider the review as not helpful, the information value of the review content in the review helpfulness analyses might be exaggerated. Practically, from the businesses' perspective, the economic value of helpful reviews is still unclear. We aim to address these questions in this section by investigating how the (helpful) review content affects product returns. On the one hand, the return outcome (i.e., whether a purchased product is eventually returned) is based on all shoppers who have bought a product, regardless of whether they voted reviews as helpful or not. A consistent finding in such analyses (e.g., the helpful review content decreases product returns) would validate the informational role

of the review content in helping shoppers make quality purchase decisions. On the other hand, the estimates in the product return models should have real economic implications.

2.6.1 Model Specification

Following the extant studies examining the effect of product reviews on product returns (Sahoo et al. 2018, Wang et al. 2021), we model the return outcome of product *i* by consumer *j* at time *t* when she bought the product by equation (2).

$$Return_{ijt} = \frac{1}{1 + \exp\{-(\beta^X X_{it} + \gamma^Z Z_{it} + \delta^R R_{it} + \eta^T T_{ijt} + \varphi_i + \tau_t + \varepsilon_{it})\}}. \quad (2)$$

The first three elements in equation (2) are similar to those in equation (1), except that they are time-variant review measures (i.e., X_{it} – image measures, Z_{it} – text measures, and R_{it} – other product review features) at the product level; the vectors β^X, γ^Z, and δ^R include the corresponding estimates. The vector R_{it} consists of an indicator of the *Absence of Review*, *Average Rating*, *Rating Volume*, *Rating Variance*, *Average Review Length*, and *Average Sentiment* of product *i* at time *t*. The vector T_{ijt} includes the transaction-level control variables, namely *Payment*, *Discount* (dollar amount off the original price), *Shipping Fee*, and *Product Age* (number of days since the product's first sales). We put the notation *j* in this vector because the value of *Payment*, *Discount*, and *Shipping Fee* is personalized and thus varies across buyers. Lastly, we denote the product fixed effects as φ_i and time dummies as τ_t, and use ε_{it} to represent the unobserved error.

We procured the sales and product returns data in 2014 from our data provider. Next, we extracted the products that appear in our review helpfulness analyses and connected the product review information with the return outcome. We excluded repeat purchases because, in these cases, consumers have already fully inspected the product and are unlikely to use product reviews to make decisions. Our return records end in October 2015, providing enough time for

all the purchases in the sales data. The final sample for the product return analyses includes 40,075 transactions of 2,081 products placed by 39,358 consumers. Table 2.6 presents the descriptive statistics of the variables in this sample. Since the outcome is binary, we apply a logit specification. Moreover, we use the manually labeled results for the image content and the SIMEX method to account for the measurement error in the review text measures.

Table 2.6. Descriptive Statistics of the Variables in the Product Return Estimations

Variables	Explanations	Min.	Max.	Mean	Std. Dev.
Returns	Whether a purchased product is returned or not	0	1	0.08	0.27
Payment	Payment per Item	0	8.39	4.09	1.06
Discount	Dollar amounts off the original price	0	5.35	0.01	0.19
Shipping Fee	Payment for the Shipment	0	5.02	0.33	0.79
Product Age	Number of days since the product's first sales	0	5115	806	872.54
Absence of Review	Whether reviews are absent on the product webpage	0	1	0.37	0.48
Average Rating	Average rating scores of all reviews	1	5	4.47	0.68
Rating Volume	Total number of reviews	0	2.94	0.88	0.81
Rating Variance	Standard deviation of all ratings	0	3.60	0.33	0.49
Average Review Length	Average number of words in a review	0.69	6.39	3.96	0.63
Average Sentiment	Average sentiment score of all reviews	-0.65	1.29	0.21	0.14
Number of Images	Number of reviews having a consumer-posted image	0	7	0.25	0.52
Image: Situational Cues	Number of images that contain situational cues	0	5	0.12	0.36
Image: Appearing of User	Number of images in which the user appears	0	6	0.15	0.37
Image: Product Malfunction	Number of images showing product malfunction	0	1	0.01	0.08
Text: Situational Cues	Number of sentences having situational cues	0	71	5.98	6.73
Text: Aesthetics	Number of sentences evaluating product aesthetics	0	20	2.68	3.12
Text: Functionality	Number of sentences evaluating product functionality	0	16	1	1.52
Text: Durability	Number of sentences evaluating product durability	0	9	0.08	0.33
Text: User Appearance	Number of sentences having identity disclosure	0	14	0.46	1.03
Helpful Reviews with an Image%	Proportion of helpful reviews with a consumer-posted image	0	1	0.05	0.17
Helpful Reviews without an Image%	Proportion of helpful reviews without a consumer image	0	1	0.32	0.35

2.6.2 Results

Table 2.7 shows the results of the returns estimations. Model (1) includes all the basic review features, the transactional variables, and the image measures. Our focal variable, *Image: Situational Cues*, has a negative and significant coefficient, implying that the contextual cues in the consumer-posted images help reduce product return probability, supporting H3. In Model (2), we remove the image content and include the review text measures. None of them obtain a

significant estimate. Model (3) has both image and text measures. The result of situational cues in the images remains significant.

Table 2.7. The Impact of Review Content on Product Returns

	(1) Image	(2) Text	(3) Image & Text	(4) Helpfulness	(5) Interaction
Payment	0.9087***	0.9109***	0.9104***	0.9663***	0.9031***
	(0.1005)	(0.1390)	(0.1391)	(0.1468)	(0.1393)
Discount	-0.1958*	-0.0860	-0.0873	-0.0907	-0.0849
	(0.0968)	(0.1082)	(0.1083)	(0.1090)	(0.1084)
Shipping Fee	-0.1618***	-0.1397***	-0.1395***	-0.1325***	-0.1385***
	(0.0335)	(0.0299)	(0.0300)	(0.0305)	(0.0299)
Product Age	-0.0005	-0.0005	-0.0003	-0.0003	-0.0003
	(0.0003)	(0.0004)	(0.0002)	(0.0002)	(0.0003)
Absence of Review	-0.1187	0.0780	0.0108	-0.0065	-0.0284
	(0.1505)	(0.1956)	(0.2006)	(0.2075)	(0.2044)
Average Rating	0.1198	0.0670	0.0566	0.1109	0.0998
	(0.1694)	(0.1044)	(0.1050)	(0.1098)	(0.1082)
Rating Volume	-0.1284	-0.0218	-0.0763	-0.0595	-0.1142
	(0.1159)	(0.1623)	(0.1665)	(0.1718)	(0.1683)
Rating Variance	0.0009	0.0089	0.0012	-0.0048	0.0105
	(0.0462)	(0.0639)	(0.0642)	(0.0652)	(0.0647)
Average Review Length	0.0178	-0.0032	-0.0167	0.0360	0.0611
	(0.0740)	(0.1397)	(0.1401)	(0.1507)	(0.1492)
Average Sentiment	-0.3762	-0.2479	-0.2612	-0.4253	-0.4429
	(0.4810)	(0.5915)	(0.6033)	(0.6232)	(0.6161)
Number of Images	0.3617**		0.3380*	0.3352*	0.3496*
	(0.1304)		(0.1469)	(0.1687)	(0.1759)
Image: Situational Cues	-0.8546***		-0.8691*	-0.9291**	-0.8844**
	(0.2243)		(0.3403)	(0.3459)	(0.3248)
Image: Appearing of User	-0.0695		0.1534	0.3198	0.3403
	(0.2200)		(0.3234)	(0.3530)	(0.3529)
Image: Product Malfunction	-0.1440		-0.1216	-0.0608	-0.1831
	(0.4657)		(0.6582)	(0.6562)	(0.6490)
Text: Situational Cues		-0.0042	-0.0045	-0.0052	-0.0052
		(0.0134)	(0.0135)	(0.0137)	(0.0136)
Text: Aesthetics with Situ. Cues		-0.4268	-0.4222	-0.4509	-0.3625
		(0.2943)	(0.2945)	(0.3025)	(0.3019)
Text: Aesthetics without Situ. Cues		-0.0244	-0.0228	-0.0226	-0.0164
		(0.0270)	(0.0272)	(0.0275)	(0.0273)
Image: Situ. Cues × Text: Aes. with Situ. Cues					-0.9501*
					(0.4702)
Text: Functionality with Situ. Cues		0.0806	0.0896	0.1564	-0.0107
		(0.1334)	(0.1337)	(0.1365)	(0.1383)
Text: Functionality without Situ. Cues		-0.0117	0.0054	-0.0129	0.0038
		(0.0598)	(0.0604)	(0.0620)	(0.0612)
Image: Situ. Cues × Text: Fun. with Situ. Cues					0.4453
					(0.2795)
Text: Durability		-0.1122	-0.0913	-0.0889	-0.1672
		(0.2453)	(0.2501)	(0.2549)	(0.2553)

Text: User Appearance		0.0429 (0.0574)	0.0311 (0.0577)	0.0218 (0.0584)	0.0161 (0.0581)
Helpful Reviews with an Image%				-0.9412* (0.4567)	-0.8988* (0.4553)
Helpful Reviews without an Image%				-0.1999 (0.2343)	-0.2152 (0.2322)
Time Dummies	Yes	Yes	Yes	Yes	Yes
Product Fixed Effects	Yes	Yes	Yes	Yes	Yes
N	40,075	40,075	40,075	40,075	40,075
AIC	22,727	22,740	22,737	22,737	22,737
$^{+}$ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001.					

In Model (4), we add two measures derived from the review helpfulness data to the estimation. Specifically, we calculate the proportion of helpful reviews with an image (*Helpful Reviews with an Image %*) and the proportion of helpful reviews with text only (*Helpful Reviews without an Image %*). Similar to Sahoo et al. (2018), we treat a review as deemed helpful if it receives at least one vote by the end of our data collection. While both variables obtain a negative coefficient, only the estimate of *Helpful Reviews with an Image %* is significant. Overall, helpful reviews with a consumer-posted image have a stronger negative impact on product returns than helpful reviews with text only. Such a finding might be another indication of the picture superiority effect, i.e., consumers are more likely to rely on helpful image content (rather than helpful textual comments) to make informed purchases.

Notably, the sheer volume of consumer-posted images (i.e., *Number of Images*) obtains a positive coefficient. It means the mere increase of the pictures without considering the content in the image may only increase product returns. Hence, not all visual information helps shoppers make informed purchases. Like the seller-provided pictures (De et al. 2013), consumer-posted images may also contain impression-based information that might incur a higher return cost.

Model (5) includes the interaction terms between the image measures and the textual comments on product attributes. Consistent with the review helpfulness analyses, the interaction

between *Image: Situational Cues* and *Text: Aesthetics with Situ. Cues* appears to improve consumers' purchase decision quality, i.e., reduce product returns. Surprisingly, the estimate of comments on product durability is insignificant across all models in Table 2.7. In fact, all the text measures do not obtain a significant coefficient. As expected, consumers might not fully use all the helpful textual comments due to the potentially high information processing cost. They would rely more on the visual information in consumer-posted images to make effective purchase decisions.

In summary, the findings in the product return analyses are consistent with our expectations. The situational cues in the images are likely to help reduce product returns. Based on the estimates of *Number of Images* and *Image: Situational Cues* in model (4), one more picture with situational cues is associated with a 20.69% decrease in the product return probability. In addition, if the proportion of helpful reviews with an image increase by 10%, the return probability will have a 6% reduction. These estimates are substantial given the total dollar amount of returns ($608,719) in our study. For example, on average, if all products in our data received one more picture containing situational cues, it might help prevent returns of products associated with an amount of $125,944.

2.7 Business Implications, Limitations, and Future Directions

We have investigated the informational role of user-generated images in online product reviews and compared it with that of textual comments. In the analyses of review helpfulness, we find that shoppers tend to value the visual content that contextualizes the product. Moreover, the situational cues in the images can replace the relevant comments on functionality attributes (e.g., the warmth of a jacket) and facilitate consumers' purchase decision-making. By comparison, the visual situational cues and review text can complement each other to help consumers better

assess the aesthetic features (e.g., the color of a pair of running pants). For product durability (e.g., the reliability of a tent), images do not communicate adequate information regarding this characteristic because this attribute has a temporal dimension. Reviewers usually do not describe the durability features using contextual cues, and consumers tend to treat the textual comments on this attribute as more helpful. On the other hand, when information processing is costly, shoppers rely on visual information to conduct more efficient product evaluations.

In the product return estimations, we find consistent results related to the visual information. Specifically, the situational cues in consumer-posted images help reduce product returns. Also, a higher proportion of helpful reviews with an image on a product web page is associated with a lower return probability, suggesting the economic value of helpful image content. By contrast, merely increasing the number of consumer-posted images may only incur a higher return probability. Lastly, the helpful image content has a more substantial influence on product returns than the counterpart in the review text. The textual comments only work when they complement the visual situational cues in describing product aesthetics. Overall, we have advanced the literature on the informational role of consumer-posted images by revealing their strong relationships with online shoppers' purchase decision-making.

The empirical results in this research provide online review platform owners or sellers with valuable business insights. First, it is important to encourage users to post pictures that share their product experiences in a real setting. For example, suppose a reviewer posts her shot being at rest while riding a mountain bike on a hill. The background information in the image (e.g., the road condition) can offer a prospective buyer helpful perceptual cues to evaluate the bike's functionality on a typical mountain road. In contrast, picturing a bike inside a room would offer little valuable information, if any at all. Second, reviewers should be suggested to clarify

the detailed nuance of the aesthetic attributes in the text in addition to an image because the image cannot sufficiently depict this type of attribute. For example, a reviewer discussing the difference between the color of a coat in the picture she has posted (as perceived in reality) and the color visible on the screen may help future buyers better understand this attribute. Third, it could be beneficial to highlight a summary of the textual comments on product durability – a feature that cannot be aptly illustrated in an image. Online retailers often provide short highlights of review comments on places most frequented by the consumers; examples can be found in studies like Chen et al. (2018) and Wang et al. (2021). Details about product durability may be a valuable topic to cover in such areas.

While the presence of an image in a product review can positively affect helpfulness votes, it is not a wise strategy to promote all consumer-posted images because the sheer volume of pictures may increase product returns. Instead, as a priority, the review platform should display reviews with helpful image content, e.g., those with situational cues or having more helpfulness votes. The information in such reviews can help reduce the return probability, thus increasing the net sales.

Our study also makes theoretical contributions. First, we unveiled a scenario when the picture superiority effect does not work – images cannot aptly delineate product durability. Besides, the visual information also has limitations in fully illustrating the aesthetic features. Researchers should consider the characteristics of product attributes when analyzing the informational role of user-generated images in the online market. Another significant theoretical implication of our study is the economic value of helpful reviews. We study a relatively less analyzed yet important characteristic of online business – the impact of review helpfulness on product returns. Returns can be especially costly in the online market because each product is

shipped separately at the purchase and return stages, and companies often bear the cost. Indeed, we show that the review content voted as helpful can improve consumers' decision quality by reducing purchase errors, i.e., lowering the product return probability. In this study, if all products in our sample received one more picture with situational cues, it might help prevent returns of products associated with an amount of $125,944.

Our study inevitably has limitations. First, we could not inspect consumers' cognitive processing of consumer-posted images, e.g., how they perceive and interpret the visual content. A well-designed lab experiment could help explore the underlying mechanism. Second, the main product categories we focus on belong to material goods. Future studies may shed light on how image content helps contextualize services (e.g., the food delivery service). Third, the sentiments in an image may synergize with the contextual cues to affect consumers' decision-making. Assessing this dimension of visual content may help better understand the informational role of consumer-posted images from a new angle. Fourth, factors improving reviewer credibility (e.g., profile photo) can be a good source for reducing product returns. Subsequent research may include more reviewer-driven features that may build readers' trust in product reviews. Despite the limitations, we believe that this study has unveiled the unique informational role of visual information provided by consumers in the online market and hope that it opens a venue for prospective research to address the issues and benefits that images bring to various types of Internet users.

CHAPTER 3

Essay 2: Trust Antecedents in Online Reviews Across National Cultures

3.1 Introduction

According to the Federal Trade Commission, the reported loss due to online shopping fraud increased by 59% in one year, from $246 million in 2020 to $392 million in 2021, during the COVID-19 pandemic (Mayfield 2022). To mitigate such problems caused by information asymmetry between online transaction participants, consumers often consult online reviews. Yet, building trust is essential for readers on online review platforms given the existence of fake reviews (Lee and Hong 2019) and the lack of traditional face-to-face interactions between sellers and consumers (Aggarwal et al. 2021).

On the one hand, review factors that affect trust and thus consumers' purchase decisions have attracted extensive research (Filieri 2016). Awad and Ragowsky (2008) argue that online trust established by information quality facilitates consumers' online shopping while source credibility, which improves trust, increases information adoption of online reviews and helps consumers make purchase decisions, contributing to review helpfulness (Aggarwal et al. 2021, Li et al. 2013). Additionally, visual information is one of the factors fostering trust in online reviews (Park et al. 2021, Shukla and Mishra 2021) as it presents real product experience of users, thereby reducing product uncertainty (Zinko et al. 2021), and reveals reviewer identity information, hence alleviating concerns about message source and increasing information credibility (Karimi and Wang 2017). With the advance in technology, most online platforms now allow consumers to upload pictures along with text-based reviews, and there has been increasing

research on review pictures; for example, when pictures provide accurate and relevant information (e.g., kitchen scene in a picture for milk products) (Yang et al. 2023), content consistent with review texts (Ceylan et al. 2021), and details on quality (e.g., bathroom, amenities, guestrooms in hotel service) (An et al. 2020, Li et al. 2021), the reviews tend to be perceived as being helpful. Nevertheless, as Pooja and Upadhyaya (2022) point out, there is still a need to investigate how two types of visual information in online reviews–one improving information quality and the other advancing source credibility–influence consumers' perception of review helpfulness, which reflects the information's ability to assist consumers in purchase decision making.

On the other hand, the trust-building process depends on national culture because social norms and values of culture play a critical role when people decide whether and whom to trust (Doney et al. 1998, Qin and De-Juan-Vigaray 2021), so findings regarding review factors influencing trust found in one country may not be generalizable to other countries that differ in national culture values. However, across national cultures, only a few studies have investigated how trust antecedents of reviews influence information usefulness perceived by consumers and no research has empirically examined the role of visual information. Therefore, we strive to fill in these gaps by studying *a cultural effect on consumer perception of trust-building factors (hereafter, trust antecedents) in online reviews accompanied by visual information*.

As national culture can be defined as "the collective programming of the mind which distinguishes the members of one human group from another" (Hofstede 1983, p. 25), there may be a country-specific effect on reducing risk and elevating trust. According to Gefen and Heart (2006), the two cultural dimensions of individualism (as opposed to collectivism) and uncertainty avoidance are influential to consumers in an online transactional environment. They

argue that people in an individualistic country are more likely to rely on and trust strangers compared to people in a collectivistic country while the combination of individualism and a low level of uncertainty avoidance would let people adopt new things that may contribute to their performance. Therefore, we endeavor to answer the following research questions: 1. *How do trust antecedents of online reviews function, as helpful features, differently across national cultures*? 2. *How are two types of visual information related to online reviews–profile photos of reviewers vs. user appearance in pictures attached to reviews–perceived differently across national cultures*?

To examine the role of national culture in online reviews, we collected 61,435 product reviews from the US and 60,102 reviews from Germany. We collected data from the same online shopping platform, Amazon, instead of two local websites because trust in a website (web reputation, design, etc.) tends to affect consumer perception of trust in online reviews (Lee and Hong 2019). We chose the two countries as the US represents individualism and is marked with low uncertainty avoidance whereas Germany is relatively high in collectivism and uncertainty avoidance, according to Hofstede Insights[8]. The two countries have considerable gaps in individualism/collectivism and uncertainty avoidance but show similar levels in the other cultural dimensions, power distance and masculinity, so we can control for the potential influences of the non-focal cultural dimensions. The cultural dimensions will be more distinguished as we, for robustness, re-collected additional reviews from other countries–Canada and the UK in line with the US vs. Spain, Brazil, and Poland in line with Germany[9]. We collected review text, pictures attached to a review, time stamp, star rating, reviewer information (vine badge, verified mark, past review experience, and nickname), and helpful votes. We text-

[8] https://www.hofstede-insights.com/

[9] According to Hofstede Insight, Spain, Brazil, and Poland are more collectivistic and higher in uncertainty avoidance than Germany, so we can supplement the cultural characteristics.

mined the reviews to extract sentiment, as well as information quality, which is based on a variety of topics identified through topic modeling.

The results indicate that trust antecedents, in terms of argument quality and source credibility, facilitate consumer perception of review helpfulness as measured by helpful votes, but the effects are moderated by national culture. Specifically, online consumers in a culture high in individualism and low in uncertainty avoidance prefer to process quality information, whereas consumers in the opposite culture may want to check reviewer credibility first. Interestingly, visual information is also perceived differently depending on the culture. US consumers tend to prefer a picture in a review, whereas German consumers may like to process profile photos of reviewers.

This research contributes to the growing literature on online reviews by introducing cultural dimensions into consumer review research through text and image mining. As for business implications, we suggest that multinational e-commerce corporations consider accommodating a culture-specific way of operating review platforms to better help consumers make purchase decisions. Also, in light of our findings, businesses can develop review-writing guidelines in more detail according to national cultures.

3.2 Literature Review

3.2.1 Trust in Online Review and the Elaboration Likelihood Model

Trust can be defined as the willingness to depend upon another party with the assumption that the trusted party will deliver actions without taking advantage of the situation (Mayer et al. 1995). While trust has traditionally been studied in the context of face-to-face and repeated interactions between people (Racherla et al. 2012), researchers have found that trust can be established in a new relationship even in an online environment where people have no interaction

history and do not expect continuous relationships (e.g., Awad and Ragowsky (2008), Racherla et al. (2012)). In the context of online reviews, individual factors that boost trust, such as reviewer expertise (Lee and Hong 2019), sociodemographic information (Racherla et al. 2012), and information quality (Cheung et al. 2012), have been widely examined as trust antecedents (Dong et al. 2019), and they can be categorized based on the Elaboration Likelihood Model (ELM) (e.g., Brand and Reith (2022), Cheung et al. (2012), Qahri-Saremi and Montazemi (2019)).

ELM was originally designed for explaining how people process information, get persuaded, and change their attitudes (Petty and Cacioppo 1986). According to ELM, informational influence is exerted through two major routes, the central route and the peripheral route, each of which involves a level of elaboration (Cheung et al. 2012, Pornpitakpan and Francis 2000, Racherla et al. 2012). The central route needs a higher level of elaboration with extensive cognitive efforts to evaluate argument quality (e.g., review message content) and information trustworthiness, whereas the peripheral route entails lower elaboration where information recipients use simple heuristics cues (e.g., source credibility of a review) to assess the trustworthiness of information. Both argument quality and source credibility of online reviews are trust antecedents as they foster trust by reducing perceived risk and uncertainty in online transactions (Cheung et al. 2012, Dong et al. 2019, Filieri 2016). On the basis of ELM, we next further discuss these trust antecedents of online reviews–argument quality and source credibility.

3.2.1.1 Argument Quality

The argument quality, as the central route in ELM, of an online review can be defined as the extent to which a consumer perceives the review content as being meaningful, relevant, and

comprehensive (Cheung et al. 2012, Qahri-Saremi and Montazemi 2019). Recipients of arguments carefully process the information in the argument contents to catch issues and evaluate merits, and this takes more effort and cognitive processing from the recipients (Chou et al. 2022).

When elaborated meticulously, higher argument quality of a review improves consumer trust in the review (Park et al. 2021, Racherla et al. 2012, Zhao et al. 2020). For example, review arguments that provide persuasive information and sufficient evidence for consumers to purchase a product (Zhao et al. 2020), effectively deliver detailed product aspects by offering pictures (Park et al. 2021), show multiple sides of a product (Jensen et al. 2013), and present relevant and useful information to consumers (Shukla and Mishra 2021) improve consumer trust in the review. This tends to foster the helpfulness of the message because the provided quality information contains situational cues for product usage and helps consumers decrease ambiguity about a product when shopping online (Qahri-Saremi and Montazemi 2019, Weathers et al. 2015). Therefore, argument quality factors that improve the trust in online reviews (i.e., trust antecedents) tend to enhance review helpfulness (Li et al. 2013) and thus facilitate the purchase decision making of consumers (Sahoo et al. 2018).

One of the widely investigated components of this construct is the amount of information from the perspective of review comprehensiveness. First, review length has been widely considered. A longer review is generally considered to have in-depth information, which can reflect the reviewer's high involvement with the review (Pan and Zhang 2011) and intensify review usefulness by helping consumers reduce product uncertainty and make purchase decisions (Mudambi and Schuff 2010). Also, Filieri (2016) argues that short reviews are not perceived as being trustworthy by a consumer as they may not provide user experience in detail.

With these merits, review length has been found to improve perceived review credibility (Lantzy et al. 2021) and helpfulness (Chou et al. 2022, Karimi and Wang 2017, Kim et al. 2019, Mudambi and Schuff 2010, Pan and Zhang 2011, Zhou and Guo 2017). Second, topic diversity, as a factor to improve argument quality, has been analyzed. According to Racherla et al. (2012) and Jensen et al. (2013), a greater extent of information available in a review helps consumers assess the attributes of the product reviewed and thus enhances the trust in the review. Also, Son et al. (2019) discover that the topic diversity representing review information is positively associated with helpful votes that facilitate consumer purchase decisions, and Lee and Zhao (2020) find that the amount of information, measured by the entropy of the topic distribution, has a positive effect on review helpfulness.

Another component of the argument quality construct is pictures attached to a review. According to Walker and Olson (1997), visual images of product usage lead to concrete and vivid mental images that help consumers vicariously experience the product and evaluate the consequence of product use. The contextual information from a picture can add authentication on top of text and provide descriptive information by providing relevant situational cues for product usage, so the picture may advance argument quality. Xu et al. (2015) show that visual presentation of a review confirms verbal descriptions and reduces uncertainty about product performance, Park et al. (2021) find that review pictures are effective in delivering information and improve readers' perception on the trustworthiness of a review, Shukla and Mishra (2021) find that pictures increase review credibility, and Filieri (2016) finds that consumer pictures improve the perceived trustworthiness of a review as they provide evidence of product use experience and help consumers evaluate product quality. Through empirical research, the presence of an image in reviews is found to be perceived helpful by consumers (An et al. 2020,

Li et al. 2021, Zhou and Guo 2017). Ma et al. (2018) confirm the predictive power of review pictures on review helpfulness using a deep learning method, and Kim et al. (2019) corroborate that the presence of a review picture would reduce product uncertainty, resulting in higher helpful votes on the review.

3.2.1.2 Source Credibility

Trust based on the source credibility of a review suggests that the other party is able to perform the job in a reliable manner (Ba and Pavlou 2002). Source credibility is judged based on a perception of the expertness and reliability of the speaker (Zhu et al. 2014)–it is not concerned with the message content per se (Chaiken 1980). Different from argument quality, the credibility of the message source as a peripheral route in ELM would take less cognitive work and effort from information recipients to process the source information, because source credibility is more likely an environmental cue to people (Xu et al. 2015) and people tend to use simple heuristic cues for evaluating the trustworthiness of a message (Cheung et al. 2012). Different from a face-to-face environment where information providers are relatively well known, online platforms host product information that is likely generated by unknown strangers. Thus, credibility perception of the source has a significant impact on a consumer's attitude toward persuasiveness by reducing source uncertainty in the online review environment (Qahri-Saremi and Montazemi 2019, Zhu et al. 2014).

For this source credibility construct, various components have been analyzed as significant factors to enhance trust in a review. First, indicating the expertise of the reviewer can be a way to build up consumer trust in a review (Filieri 2016). Badges, e.g., Top 100 reviewer and vine badges, indicating the trustable social status and reputation of the reviewer on the site are found to foster trust in a review (Cheung et al. 2012, Filieri 2016, Zhu et al. 2014). Second, the mark of

a verified purchaser can help build trust by showing that the review is indeed from a real purchaser (Leong et al. 2021, Qahri-Saremi and Montazemi 2019) and by improving the perceptual affinity between reviewers and review readers (Wu 2017). Third, an experienced reviewer, based on the number of reviews the reviewer has posted, tends to be deemed more reliable and credible than a novice reviewer (Filieri 2016, Sahoo et al. 2018). In addition, disclosed reviewer information (e.g., real name, profile photo) would improve source credibility (Xu 2014).

The source credibility feature as a trust antecedent in online reviews has been found to improve review helpfulness as well. Zhou and Guo (2017) show that reviewer expertise shown as the Elite badge on Yelp has a positive association with review helpfulness, and Hong et al. (2017) and Chou et al. (2022) show that reviewer expertise measured by the total number of reviews by a reviewer on the review platform has a significant positive effect on review helpfulness. Reviewer's disclosed information, such as profile image (Karimi and Wang 2017, Zhou and Guo 2017) and self-descriptive identity (e.g., nickname, hobbies, real name) (Forman et al. 2008, Ghose and Ipeirotis 2011, Karimi and Wang 2017), leads to higher review helpfulness. Besides, the verified purchaser mark on Amazon is found to be a helpful factor (Kokkodis and Lappas 2016) but Zhu et al. (2020) also show that non-verified reviews can be helpful if they contain more words and are written in a narrative style.

Despite the abundant studies, empirical research of online reviews rather focuses on one culture (Pooja and Upadhyaya 2022), although the trust-building process is influenced by national cultures (Fukuyama 1995, Doney et al. 1998, Gefen and Heart 2006) and thus the effects of the trust antecedents may differ across cultures. Recently, a few studies have conducted multinational empirical research on online reviews; Biswas et al. (2022) discover different roles

of review features (e.g., review title length, sentiment) on helpful votes across the US and India, and Barbro et al. (2020) investigate national effects on consumer reviewing behavior and percentage of helpful reviews. Yet, to the best of our knowledge, little is known about how the trust antecedents are perceived as being helpful across national cultures. Especially, despite the increasing research on pictures of online reviews, there is a lack of studies suggesting graphic features that make a review helpful and comparing two types of visual information–profile photo and review picture–that include the same content, user appearance. To bridge this gap, we endeavor to investigate how the effects of information quality and source credibility features would differ across national cultures by employing both textual and visual features, whereas previous studies are limited to commonly used variables.

3.3 Hypotheses Development

Figure 3.1 summarizes our research design. We detail our hypotheses development below.

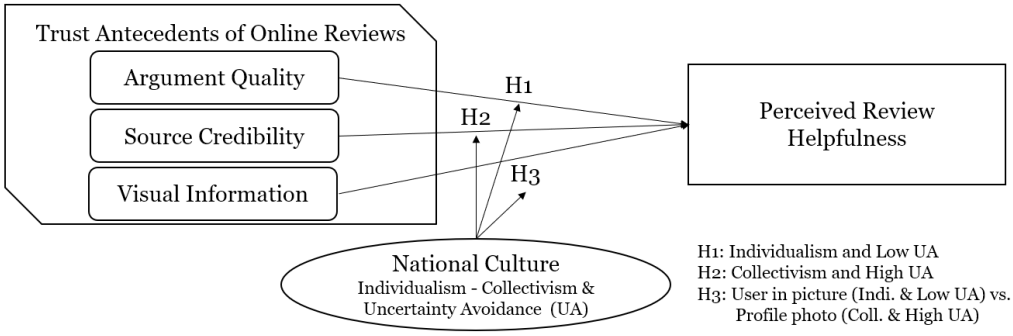


Figure 3.1. Research Design

3.3.1 Cultural Effects with Trust Antecedents

According to ELM, the effects of the two routes are moderated by a recipient's involvement level (Petty and Cacioppo 1986). People would be inclined to adopt the central route to carefully process an argument if they are highly involved in the information processing, whereas they would rely on the credibility of information (i.e., peripheral route) otherwise (Yang et al. 2023).

However, we cannot differentiate prospective consumers' involvement in our focal context (i.e., online reviews) because they all would want to avoid purchasing wrong or bad products, so their effort in processing review information is not likely to be discerned. Rather, we suggest that a recipient's national culture would moderate the effects of the central and peripheral routes on review helpfulness because the trust antecedents may be perceived differently across national cultures (Barbro et al. 2020, Kim et al. 2018, Qin and De-Juan-Vigaray 2021, Tang 2017).

According to Hofstede (1983), different cognitions and behaviors of people across nations can be explained by their national cultures, and the effect of national culture on trust-building has long been maintained (Doney et al. 1998, Fukuyama 1995, Gefen and Heart 2006, Qin and De-Juan-Vigaray 2021). For example, Gefen and Heart (2006) suggest that trust should be understood in the context of national culture, because cultural norms and values may be impotent and misleading in one setting even though they are conducive to effective trust production in another. In a similar vein, Tang (2017) finds significant roles of national culture (e.g., individualism, uncertainty avoidance) on the relationship between online reviews and market performance of products. As for review helpfulness, it has been found that consumers perceive helpfulness from online reviews differently depending on the national culture because the culture influences patterns of behaviors of people and thus consumers' perceptions and evaluations of reviews (Biswas et al. 2022).

National culture is often described in terms of four dimensions (Biswas et al. 2022, Hofstede 1983, Srite and Karahanna 2006): individualism-collectivism, uncertainty avoidance, power distance, and masculinity. Among the four, individualism-collectivism and uncertainty avoidance have been found to play a key role in influencing trust-building online (e.g., Biswas et al. (2022), Pornpitakpan and Francis (2000), Tang (2017), Vishwanath (2003)). Accordingly, we examine

the effects of these two cultural dimensions in depth in this research. Individualism represents the degree to which individuals highlight their own needs as opposed to their group's needs. This dimension deals with a preferred way of a society to handle complexity and its related value system of willingness to rely on strangers (Barbro et al. 2020, Gefen and Heart 2006, Tang 2017). Uncertainty avoidance can be defined as the lack of tolerance for uncertainty (i.e., the degree of risk aversion) of people within the culture (Doney et al. 1998, Gefen and Heart 2006, Srite and Karahanna 2006, Vishwanath 2003). When a society is likely to feel threatened by ambiguous situations, it scores high in uncertainty avoidance.

3.3.2 Individualism and Low Uncertainty Avoidance with Argument Quality

When a national culture is centered on high individualism, people tend to rely on and trust someone unless evidence suggesting not to trust the trustee is manifested (Fukuyama 1995, Tang 2017), and they are found to be independent, self-reliant, trusting others, and utilitarian (Doney et al. 1998, Kim et al. 2018, Srite and Karahanna 2006, Triandis 1995). That is, as Pornpitakpan and Francis (2000) suggest, individualists are more motivated to rely on facts and arguments instead of message sources than collectivists are. Accordingly, in the e-commerce market, individualistic participants can reduce uncertainty and predict the results of purchases by browsing e-commerce websites (e.g., strangers' opinions), whereas more collectivistic people might need information from their in-group (Gefen and Heart 2006). In online reviews, individualists engage with reviews written by unknown reviewers more than collectivists do and would evaluate reviews' helpfulness without discrimination of the reviewers (Tang 2017).

When it comes to uncertainty avoidance, people who live in a low uncertainty avoidance culture tend to trust strangers and ordinary citizens as well (Pornpitakpan and Francis 2000). They are less likely to be reluctant to adopt informal and non-institution-based information (Park

et al. 2012) and tend to be affected equally by both high- and low-credibility sources. Shoppers from a low uncertainty avoidance culture would be open-minded when searching for information (Vishwanath 2003), and they are likely to have higher tolerance to uncertain information providers (Biswas et al. 2022) and perceive lower risk from online reviewers without proper credibility (Tang 2017) than their peers in the opposite culture. Therefore, when reading a review, consumers in a low uncertainty avoidance culture rely on the review argument itself regardless of the credibility of the message source more than those in a high uncertainty avoidance culture.

In most cases, online reviewers do not have personal relationships with review readers, and reviews are often written by anonymous people. Putting these together, we posit that readers who espouse an individualistic and low uncertainty avoidance culture will be inclined to utilize review contents rather than ensuring source credibility, and thus argument quality would play a vital role when they read reviews and evaluate the helpfulness of the arguments. On the other hand, for those in a collectivistic and high uncertainty avoidance culture, the effect of message argument will be attenuated as they would rely on other factors (e.g., reviewer information) in reading reviews. Hence,

H1*: The argument quality of an online review is more likely to advance the perception of review helpfulness in a national culture high in individualism and low in uncertainty avoidance than in the opposite culture.*

3.3.3 Collectivism and High Uncertainty Avoidance with Source Credibility

While people from an individualistic culture would trust information from out-group parties and evaluate it similarly to in-group information, individuals in a collectivistic culture may depend more on relationships (Triandis 1995), social/group norms (Srite and Karahanna 2006), and

harmony with others (Kim et al. 2018). Even in the e-commerce market, prospective consumers in a collectivistic culture put more importance on what their friends are telling than those in an individualistic culture do (Biswas et al. 2022). In online reviews, socially relevant contents are perceived as being more helpful (Biswas et al. 2022) and consumers tend to perceive risk from information without credibility in a collectivistic culture (Tang 2017). Collectivists are likely to be persuaded by a highly credible source and unlikely to perceive a message from a low-credibility source as being believable (Ritchie and Phares 1969). Also, collectivists tend to conform to expert opinions more than individualists do (Triandis 1995).

As to uncertainty avoidance, people in a culture high in uncertainty avoidance are more reluctant to adopt opinions from unknown sources than those in the opposite culture (Doney et al. 1998). They are more concerned with security and thus show more reliance on experts (Hofstede 1983). To avoid uncertainty, they tend to depend on others' opinions by searching for social cues (Srite and Karahanna 2006), formal rules, and relationships (Park et al. 2012). Also, information from personal (e.g., from in-group people) (Srite and Karahanna 2006) or formal sources (Hofstede 1983) is found helpful to relieve uncertainty. However, in the absence of formal rules or guards, such as in the e-commerce market, individuals create a feeling of security by relying on experts (Pornpitakpan and Francis 2000) and credible sources (Qahri-Saremi and Montazemi 2019, Zhu et al. 2014). In the context of ELM, this may imply that source credibility indicating source expertise and authenticity is more important to people from a high uncertainty avoidance culture than to those from a low uncertainty avoidance culture.

All in all, online consumers in a culture low in individualism (i.e., high in collectivism) and high in uncertainty avoidance may prefer online reviews coming from reputable and reliable sources. Without a credible source, they may feel uncomfortable and reluctant to accept a

review, leading to the lack of effect of the review on their purchase decision making. Therefore, the source credibility of a review is expected to be more important to them, whereas the effect is attenuated for their peers in the opposite culture, who may rely on the review information even without a reviewer-driven trust antecedent. Hence,

H2*: The source credibility of an online review is more likely to advance the perception of review helpfulness in a national culture high in collectivism and uncertainty avoidance than in the opposite culture.*

3.3.4 Visual Information

Nowadays, images can be uploaded on almost all review platforms, but relatively little is known about visual information in a review as the extant literature largely focuses on review text (Pooja and Upadhyaya 2022, Zinko et al. 2021). Yet, a picture in a review has been found to increase trust in the review as well because visual materials would provide user experience so that consumers can imagine their product usage result and reduce their uncertainty in an online shopping environment (Filieri 2016, Xu 2015, Zinko et al. 2021). According to Zinko et al. (2021), a review with a picture is likely to have a greater impact on trust than a text-alone review because user experience of the product described in the picture can reduce product uncertainty more than a text-based review can. In addition, Filieri (2016) finds that, in a survey, a picture shows evidence of real product usage, helps consumers evaluate the quality of the product, and thus improves perceived trustworthiness of the review.

In the online review environment, there are two types of visual information containing a reviewer: a profile photo and user appearance in a review picture. The visual element can enhance the persuasiveness of the message (Herr et al. 1991), and it helps a consumer make an informed purchase decision as the expected outcome of the decision can be observed in a better

way with a user image (Karimi and Wang 2017). The pictures in a review tend to improve review helpfulness if they provide authentic environment for product usage (Yang et al. 2023) and relevant information, such as bathroom and amenities in hotel reviews (An et al. 2020). Also, a profile photo of the reviewer has been found to improve review helpfulness (Karimi and Wang 2017, Zhou and Guo 2017).

However, to the best of our knowledge, no research has compared and can suggest different roles of the two types of visual information across national cultures. To bridge this gap, we examine how these two types of images are perceived differently by consumers' own culture and identify cues on how visual-formatted user information functions differently based on the information layout across national cultures.

We posit that a profile photo would serve as a part of source credibility, whereas a review picture would be a trust antecedent as the argument quality construct. This is because the presence of a profile photo raises the awareness of the reviewer and reduces source uncertainty (Xu 2014) and it even improves the social presence of the reviewer, which is an antecedent of trust (Filieri 2016). It may not take much effort for consumers, who are shopping online, to go through a reviewer's profile photo as they would, based on the reviewer information, focus more on the trustworthiness of the review. Meanwhile, user information in a visual-formatted review would reduce equivocality (Weathers et al. 2015), provide detailed information (Lin et al. 2012), and illustrate various aspects about the product (e.g., size, real color, fit, breathability, and stretch) (Kim et al. 2019), which may require cognitive work. Accordingly, the two types of visual information may foster trust in a review and improve review helpfulness, but their importance would be dependent on consumers' national culture. Hence,

H3*: Visual appearance of a product user in an online review is more likely to advance the*

perception of review helpfulness in a national culture high in individualism and low in uncertainty avoidance, whereas the profile photo of a reviewer is more likely in the opposite culture.

3.4 Data

3.4.1 Data Collection

According to Hofstede Insights, the US and Germany show different characteristics in the two cultural dimensions of individualism-collectivism and uncertainty avoidance but are not so distinct in the other two cultural dimensions. Specifically, the US scores high in individualism (91 out of 100), whereas Germany is relatively low in individualism (67), i.e., relatively high in collectivism. As to uncertainty avoidance, Germany is above midrange (65), whereas the US is under midrange (46). Yet, the two countries only have a 5-point difference in power distance (US, 40; Germany, 35) and a 4-point difference in masculinity (US, 62; Germany, 66). Thus, in this research, we chose these two countries to compare national cultures; the US represents an individualistic and low uncertainty avoidance culture, whereas Germany represents a relatively collectivistic and high uncertainty avoidance culture[10]. Although we examined reviews from the two countries in our main analysis, we will strengthen the cultural characteristics by examining more reviews from other countries in a robustness check.

We collected review data from one platform by May 26, 2020, in two countries: US consumer reviews from Amazon.com and German reviews from Amazon.de. By selecting the same platform, we can control for the effects of platform designs on trust-building and consumers' decision making–the issue raised by Dong et al. (2019). The Amazon review

[10] One may point out that Germany is also an individualistic country, but it is relatively collectivistic compared to the US. For example, in the cultural comparison with Netherlands, German people are shown to be more sensitive to expert high-quality opinions (Hornikx and ter Haar 2013). Also, Hornikx and Hoeken (2007) show that even western countries (e.g., France and Netherlands) have a different culture. We also examine the relative effects of national culture of the US versus Germany.

platform has two sections: reviews from the focal country (i.e., the US on Amazon.com and Germany on Amazon.de) and reviews from other countries (e.g., reviews written in Canada on Amazon.com and Amazon.de). We used reviews written in the focal country only (i.e., those written in the US for US reviews and those written in Germany for German reviews), because consumers cannot vote for reviews written in other countries and hence the helpful votes of such reviews may not adequately reflect the helpfulness of the reviews. Also, based on a previous work of Straub et al. (Straub et al. 2002), we assume that consumers are affected by the national culture to which they are exposed rather than the national culture to which their nationality belongs. As a product type, clothing was chosen because, according to the Statista Global Consumer Survey in 2021 (www.statista.com), clothing is the most dominant product category purchased online in both countries. Also, different from products (e.g., electronic devices) that are evaluated by product specifications (e.g., RAM, CPU) where there would be little room for purchasers and potential consumers to communicate via online reviews, clothes may not be perfectly described by product descriptions. Thus, consumers actively search for product reviews when shopping for clothing items, and their decisions are influenced by user-generated information (Harrigan et al. 2021). In this research, we included various types of clothing items, such as outwear, jackets, tracksuits, hoodies, vests, shirts, pants, jeans, skirts, swimwear, socks, underwear, and accessories.

We obtained 61,435 reviews for 401 products on Amazon.com and 60,102 reviews for 406 products on Amazon.de. Then, we removed reviews that were written before 2015 because we found systematic differences between Amazon.com and Amazon.de, i.e., zero picture was uploaded in Germany in 2012 and 2014 and in the US in 2013, and no picture was included in a review before 2012. Also, we excluded reviews written between April 1 and May 6, 2020,

because Amazon temporarily disallowed reviewers to upload pictures in that period[11]. This helps us control for exaggerated image effects. As a result, 53,410 US and 52,812 German reviews remained. Each observation consists of review text, pictures, time stamp, star rating, reviewer information (profile photo, vine badge, verified mark, past review experience, and nickname), and helpful votes.

We mined more variables on the basis of the collected review data. From review pictures, we extracted the number of pictures and the number of pictures presenting a product user. From review text, we extracted review length, topic diversity, and review sentiment. Below, we will present the feature extraction processes–image mining for human detection and text mining for measuring topic diversity and review sentiment.

3.4.2 Visual Information Processing–User Detection

For user detection in a picture, we used an image classification tool for human recognition based on *MobileNet V3,* which is a model using convolutional neural networks. The pre-trained model returns classified objects detected in a picture, and we focused on a human object among many types. As our target product category involves fashion items, such as jackets and jeans, many reviewers showed their whole body, but some reviewers showed specific body parts or even just a product only. The image-mining algorithm can easily find people given a whole body. In the case of body parts, the combination of parts (e.g., a body without a face, the combination of a face and an upper body) also improves detection result. However, when there is a lack of body cues (e.g., only legs wearing socks, a face with sunglasses), the model does not always return the human label. To improve detection accuracy, we also employed a face recognition tool, *Deepface*. As long as the human/face object is recognized by either or both of the algorithms, we

[11] https://www.pocnetwork.net/internet-news/amazon-no-longer-allowing-images-within-consumer-product-reviews/

deem that a user appears in the picture.

For an accuracy check, we randomly picked 600 pictures–265 from the US reviews and 335 from the German reviews. Then, we asked three coders to manually label them as person-included or not. Two coders each labeled 300 pictures. The third coder (one of the authors) independently labeled all 600 pictures. The Cohen's Kappa was 95% and 98% between the first two coders and the third coder, respectively, indicating excellent agreement. The third coder then discussed and debated with the other two coders to finalize the labeling results. As for the criteria in the manual classification, we checked yes if at least a body part is presented in a picture highlighting a product (e.g., a hand wearing a glove or a foot wearing a sock) because it can be a piece of good information for potential consumers. Among the 600 pictures, 307 are positive (i.e., person-included). The user detection algorithms together gave a precision and recall of 89.2% and 73%, respectively, on the 600 pictures.

3.4.3 Natural Language Processing–Topic Modeling and Sentiment Analysis

To measure the topic diversity of a review, we used topic modeling (Blei et al. 2003). We processed reviews separately by country and calculated the entropy of the topic distribution of each review following Gong et al. (2018) and Lee and Zhao (2020). As text pre-processing, we lowered the letter case and removed relatively noisy words, such as numbers, punctuations, white spaces, and stop words (e.g., 'a', 'the' in English and 'ein', 'der' in German), using the *tm* package in R. After this pre-processing, we had 31,065 English words and 65,403 German words left, but these are still too many to conduct topic modeling effectively. Accordingly, we kept only words that appear in more than 1% of the reviews, so that we can exclude unnecessary and exceptional topics possibly caused by brand/product names and spelling errors. As a result, 272 and 182 words remained in English and German, respectively. Then, we applied the term

frequency-inverse document frequency (TFIDF) word weighting scheme. Specifically, a word *w* in a document (i.e., review) *d* is indexed as in Equation (1), and we applied the calculation separately by the languages.

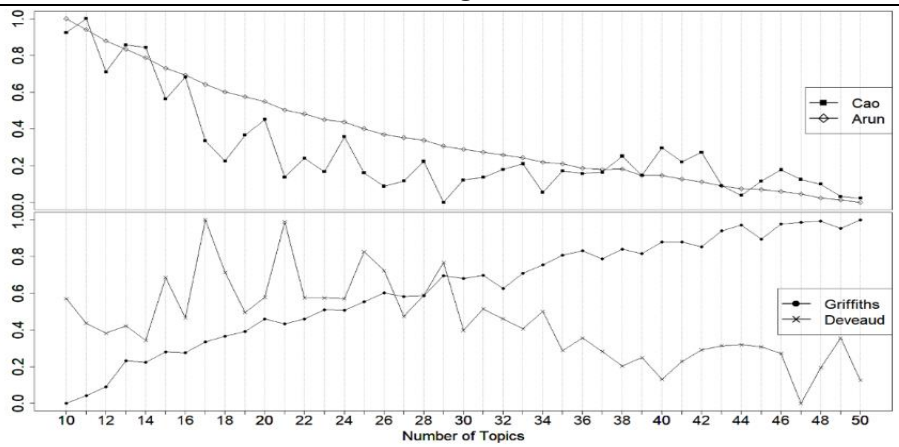
$$Weight_{wd} = frequency_{wd} * log \frac{Number\ of\ Documents}{Number\ of\ Documents\ that\ include\ w} \quad (1)$$

Then, we employed topic modeling to infer semantic features of reviews. As Gong et al. (2018) suggest, the latent Dirichlet allocation (LDA) model is the most widely used approach. LDA, a hierarchical Bayesian model, returns posterior topic probability distribution for each review. Higher entropy of the posterior topic probability distribution can imply a broader range of topics (Gong et al. 2018), and this entropy has been used to measure the amount of information inherent in a review (e.g., Lee and Zhao (2020), Singh et al. (2017)). We used LDA for topic modeling and calculated the entropy using Equation (2), where $\hat{\theta}_{it}$ denotes the posterior probability of topic *t* belonging to review *i* and *T* is the total number of topics.

$$Entropy_i = -\sum_{t=1}^{T} \hat{\theta}_{it}\, log_2(\hat{\theta}_{it}) \quad (2)$$

To determine an appropriate topic number, we compared four metrics across various topic numbers: “Arun”, “Cao”, “Deveaud” and “Griffiths” (Lee and Zhao 2020). Overall, the topic number was chosen based on maximized “Arun” and “Cao” and minimized “Deveaud” and “Griffiths.” Figure 3.2 respectively displays outcomes of the four metrics with the range of topic numbers from 10 to 50 by language. No particular topic number universally dominated others, but we selected 25 considering the results of the four metrics in both English and German languages. Although we used 25 topics for the main analysis, we will examine whether the variable of topic diversity based on different numbers of topics, 15 and 35, will alter the main results in a robustness check.

A. English



B. German

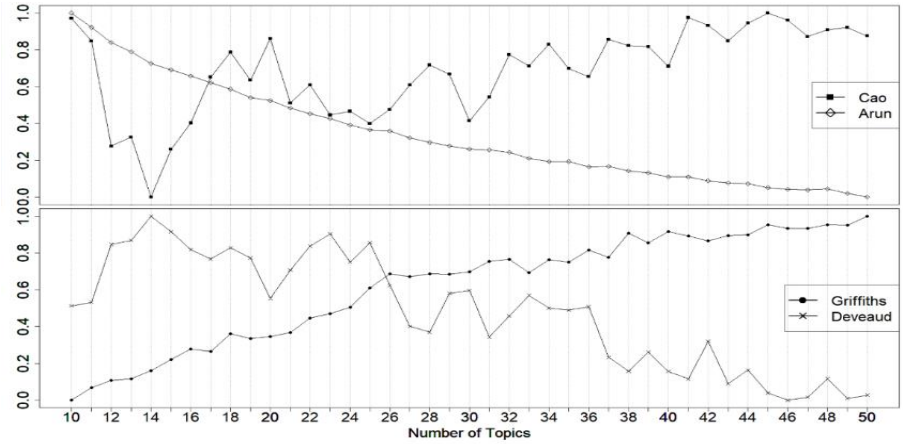


Figure 3.2. Four Metrics for Different Topic Numbers

As we used 25 topics, the maximum and minimum values of entropy are about 4.64 (when all the topics are mentioned with the same probability) [12] and 0 (when a single topic takes 100% probability), respectively. As we handled the two languages (i.e., English and German) separately, this topic modeling is a good way to measure topic diversity in the sense that a language issue may not be raised as long as reviews are treated separately.

We also measured the sentiment of each review. As shown by Salehan and Kim (2016), sentiment expressed in review text is one of the factors that influence review helpfulness. If we use multiple programs to analyze sentiment for different languages (i.e., English and German),

[12] $4.643856 = -25*0.04*\log_2(0.04)$

we may face the issue of measurement difference for quantifying sentiment. Therefore, we chose a service provider for sentiment analysis based on multiple languages–IBM Watson[13], which has been used in previous research (e.g., Song et al. (2022)). The IBM Watson Natural Language Understanding offers built-in deep learning models pretrained on unstructured texts, including social media texts, to support user-generated texts in many languages[14] (English, German, Portuguese, Russian, Spanish, etc.). The models return a score ranging from -1 (negative) to 1 (positive) for the overall sentiment of a review. Using this tool, we avoid the need for language translation, which may lead to meaning distortion and information loss.

3.4.4 Variables

Our focal variables for argument quality include *Length*, *TopicDiversity*, *Pictures,* and *UserPictures*. *Length* is measured by the word count of a review to represent the amount and depth of information. *TopicDiversity* is calculated by topic entropy to reflect how many types of information about a product are conveyed in a review. Both *Pictures* and *UserPictures* are counted by the number of pictures in a review, but *UserPictures* takes into consideration of a person who uses or wears the reviewed product while *Pictures* is the number of all pictures. For source credibility, we used *Verified*, *Vine*, *PastReview*, *Nickname,* and *ProfilePhoto*. All these variables are measured as binary: whether the reviewer is a verified and real purchaser (*Verified*); whether the reviewer has a vine badge, which is awarded to highly-ranked reviewers by Amazon (*Vine*); whether the reviewer has ever written at least one review on the product category (*PastReview*); and whether the reviewer uses his/her own nickname rather than a default reviewer name, "Amazon Consumer"/"Amazon Kunde " (*Nickname*); whether the reviewer has his/her own profile photo (*ProfilePhoto*). Lastly, we marked a review written by a US (German)

[13] https://www.ibm.com/cloud/watson-natural-language-understanding
[14] https://cloud.ibm.com/docs/natural-language-understanding?topic=natural-language-understanding-language-support

reviewer with 1 (0) (*Nationality*).

When reading a review, consumers are willing to vote for it if they think it is helpful for their purchase decision making and they adopt the review information. In many studies, the number of helpful votes of a review has been deemed as a peer-generated evaluation of the review indicating how much the review is helpful for consumers' decision making. Accordingly, we used the total number of helpful votes of a review, which has been widely used in previous studies (e.g., Cao et al. (2011), Liu and Park (2015), Wu (2017), Zhu et al. (2014)), as the dependent variable (*Helpfulness*).

Additionally, in IS research, factors that influence review helpfulness have been widely examined, and we controlled for such basic features. Pan and Zhang (2011) find that the numerical review rating (*Rating*) (on a scale of 1 to 5) of a review is positively associated with review helpfulness. Mudambi and Schuff (2010) find that review extremity (*ReviewExtremity*), measured by the squared deviation of the review star rating from the average product rating, positively affects review helpfulness. The square is to see how far the reviewer evaluation on the product is from the average product rating, and extreme ratings are found to be more influential to consumers than moderate ratings (Forman et al. 2008). When it comes to review age (*DaysElapsed*), measured by the number of days since a review post, Zhu et al. (2014) support that older reviews are more likely to get helpful votes because aged reviews tend to be more revealed to consumers than recently posted reviews. Nonetheless, this is debatable as Cao et al. (2011) conclude that consumers prefer recent reviews because of consequential higher accuracy of information from the integration of previous and recent information. Moreover, Zhou and Guo (2017) find that consumers deem earlier ordered reviews to be more useful (*Order*) because the location of a review can affect its visibility to consumers, and Salehan and Kim (2016) argue that

sentiment (*Sentiment*) expressed in a review is shown to have a significant relationship with review helpfulness as it may reveal defect features of the product, which are not usually found in product description.

Additionally, we included three more picture-driven features as control variables because they can affect the information quality of review pictures and arouse positive feelings of readers, leading to higher helpful votes. According to Zhang et al. (2021), image saturation means the richness and vividness of color, which has a positive relationship with happiness (*Picture–Saturation*); image brightness refers to the overall level of illumination where sufficient illumination helps the image deliver information by making the image content clear to viewers (*Picture–Brightness*); and warm hue of a picture shows dominance of warm colors (e.g., red) over cold colors (e.g., blue), and it is linked with arousal of emotion and thus higher tendency of consumption (*Picture–WarmHue*).

Table 3.1 shows variable statistics, Variance Inflation Factor (VIF) index, and correlations of the variables. One review has 0.05 pictures on average, so we can expect that attached pictures can be meaningful for consumers in reading a review. Because we used two languages, there could be concerns on systematic differences in the variables *Length* and *TopicDiversity*, which can result in biased estimation. Yet, the average values of *Length* and *TopicDiversity* of German versus English are 2.79 versus 2.63 and 4.31 versus 4.28, respectively. Accordingly, we do not expect that the language issue can cause notable systematic differences. In addition, although *Rating* and *ReviewExtremity* show the highest correlation among the variables, there is little concern for multicollinearity based on the VIF–research can obtain multicollinearity-free results if VIF is less than 10 (Kutner et al. 2004).

Table 3.1. Variable Statistics

						VIF	Correlation													
Variable	Min	Med	Mean	Max	St.dev		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Helpfulness	0	0	0.83	414	5.90		-0.05	0.07	0.05	-0.06	-0.06	0.16	-0.09	0.19	0.13	0.00	0.00	-0.01	0.00	0.06	0.01	0.02	0.04	-0.01
2. Rating	1	5	4.38	5	1.13	4.58		-0.87	-0.01	0.01	0.50	-0.17	0.11	-0.04	-0.01	0.01	0.01	0.01	0.02	0.02	-0.01	0.01	0.01	-0.01
3. ReviewExtremity	0	0.25	1.21	14.44	2.61	4.15			-0.02	0.00	-0.41	0.11	-0.06	0.05	0.02	-0.02	0.00	-0.01	-0.02	-0.02	0.01	-0.01	-0.01	0.00
4. DaysElapsed	0	413	572	1971	490	1.12				-0.13	-0.06	0.13	-0.12	-0.03	-0.02	0.01	0.00	0.00	-0.01	-0.01	-0.03	0.01	-0.02	0.23
5. Order	1	123	185	969	185	1.05					0.00	-0.08	0.03	-0.03	-0.02	-0.01	0.00	0.00	0.04	-0.05	-0.03	0.00	-0.01	0.11
6. Sentiment	-1	0.36	0.36	1	0.48	1.41						-0.23	0.19	-0.03	-0.02	0.01	0.00	0.00	0.02	-0.01	-0.02	-0.01	-0.01	-0.13
7. Length	0.69	2.77	2.71	6.86	1.00	2.88							0.77	0.14	0.08	0.00	0.00	0.00	-0.13	0.16	0.01	0.07	0.08	-0.08
8. TopicDiversity	2.03	4.30	4.30	4.64	0.18	2.68								-0.10	-0.05	0.00	0.00	0.00	0.09	-0.10	0.00	-0.06	-0.06	-0.10
9. Pictures	0	0	0.05	15	0.36	1.55									0.57	-0.01	-0.01	-0.01	-0.06	0.15	0.02	0.02	0.09	-0.02
10. UserPictures	0	0	0.01	7	0.14	1.50										0.02	0.03	0.04	-0.03	0.05	0.01	0.01	0.05	0.01
11. Picture–Saturation	0	0	0.01	0.73	0.04	1.05											0.18	0.14	0.01	-0.01	-0.01	0.00	0.00	0.01
12. Picture–Brightness	0	0	0.02	0.97	0.10	1.04												0.03	0.00	-0.01	-0.01	-0.01	-0.01	0.00
13. Picture–WarmHue	0	0	0.02	1	0.12	1.02													0.00	0.00	0.00	0.00	-0.01	0.00
14. Verified	0	1	0.96	1	0.19	1.11														-0.29	0.01	-0.02	-0.08	0.00
15. Vine	0	0	0.01	1	0.10	1.16															0.00	0.03	0.18	-0.02
16. PastReview	0	0	0.002	1	0.05	1.00																0.00	0.00	-0.04
17. Nickname	0	1	0.89	1	0.31	1.02																	0.11	0.04
18. ProfilePhoto	0	0	0.12	1	0.32	1.05																		0.02
19. Nationality	0	1	0.50	1	0.50	1.19																		

Following the literature (e.g., Mudambi and Schuff 2010), we took the natural logarithm for Length.

3.5 Main Analysis

3.5.1 Research Model

We used the negative binomial regression model because the dependent variable (*Helpfulness*), measured by the cumulative number of helpful votes of a review, is an over-dispersed, non-negative count variable. In the research model, our focal variables are presented in vectors *X* and *Y*. *X* represents variables for review argument quality (i.e., *Length*, *TopicDiversity*, *Pictures,* and *UserPictures*), and *Y* includes source credibility variables (i.e., *Verified*, *Vine*, *PastReview*, *Nickname,* and *ProfilePhoto*). Vector Z has control variables–*Rating*, *ReviewExtremity, DaysElapsed*, *Order*, *Sentiment*, *Picture–Saturation*, *Picture–Brightness*, and *Picture–WarmHue*. Lastly, *Nationality* is denoted as *N*, and it will interact with the two sets of focal variables.

$$ln\,(Helpfulness_{ij}) = \beta^X X_{ij} + \delta^Y Y_{ij} + \gamma^Z Z_{ij} + \lambda N_{ij} + \rho^X N_{ij} X_{ij} + \omega^Y N_{ij} Y_{ij} + Yr_{time} + M_{time} + P_j + \varepsilon_{ij} \quad (3)$$

As shown in Equation (3) regarding review *i* for product *j*, product (P_j) and time (Yr_{time}, M_{time}) fixed effects are applied to rule out, on getting helpful votes, the effects of an individual product and time (i.e., year and month of a review timestamp) of a review. We examined how the proposed trust antecedents of a review influence consumers' perception of review helpfulness differently across national cultures.

3.5.2 Results

Across all models in Table 3.2, all estimates of the argument quality factors are statistically significant. This indicates that the amount of information (*Length*), diversity of topics (*TopicDiversity*), pictured information (*Pictures*), and the user presence in a picture (*UserPictures*) are perceived helpful by consumers. Meanwhile, not all the source credibility factors are perceived as being helpful by review readers. A verified purchaser mark (*Verified)*, a

vine badge (*Vine*), and a profile photo of a reviewer (*ProfilePhoto*) have statistically significant positive associations with review helpfulness, but a reviewer's past review (*PastReview*) does not. A reviewer's nickname (*Nickname*) is not consistently significant across the models; the statistical significance disappears if we include the interaction of *Nickname* with *Nationality*.

Table 3.2. Main Results

	(1) Basic	(2) Quality	(3) Source	(4) All
Rating	0.3269***	0.3242***	0.3283***	0.3261***
	(0.0278)	(0.0282)	(0.0269)	(0.0275)
ReviewExtremity	0.2007***	0.2008***	0.2009***	0.2010***
	(0.0102)	(0.0103)	(0.0099)	(0.0101)
DaysElapsed	0.0004	0.0007	0.0003	0.0006
	(0.0018)	(0.0018)	(0.0017)	(0.0017)
Order	-0.0015***	-0.0015***	-0.0015***	-0.0015***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Sentiment	-0.2088***	-0.2080***	-0.2159***	-0.2184***
	(0.0372)	(0.0377)	(0.0361)	(0.0368)
Length	1.2616***	1.1595***	1.2604***	1.1420***
	(0.0225)	(0.0348)	(0.0218)	(0.0342)
TopicDiversity	1.1874***	0.6207***	1.1947***	0.6081***
	(0.1135)	(0.1577)	(0.1100)	(0.1537)
Pictures	0.1928***	0.1794***	0.1927***	0.1685***
	(0.0169)	(0.0217)	(0.0165)	(0.0213)
UserPictures	0.6095***	0.4094***	0.6080***	0.4142***
	(0.0410)	(0.0661)	(0.0398)	(0.0647)
Picture–Saturation	-0.1459	-0.4047	-0.1551	-0.4112
	(0.4334)	(0.4420)	(0.4199)	(0.4301)
Picture–Brightness	0.0719	0.2269	0.0351	0.2027
	(0.4842)	(0.4871)	(0.4698)	(0.4743)
Picture–WarmHue	-0.5507*	-0.4995*	-0.4838*	-0.4159+
	(0.2480)	(0.2488)	(0.2409)	(0.2427)
Verified	1.2120***	1.2369***	1.5362***	1.4945***
	(0.0907)	(0.0924)	(0.1344)	(0.1373)
Vine	0.3338***	0.3656***	0.5340***	0.6117***
	(0.0892)	(0.0907)	(0.0996)	(0.1024)
PastReview	0.9459	0.9375	1.2884	1.0953
	(0.9359)	(0.9482)	(2.0812)	(2.1225)
Nickname	0.1204*	0.1225*	0.0985	0.1135
	(0.0582)	(0.0590)	(0.0779)	(0.0793)
ProfilePhoto	0.3021***	0.3020***	0.3933***	0.4199***
	(0.0411)	(0.0417)	(0.0557)	(0.0567)
Nationality	2.4023***	-2.4194*	3.1143***	-1.9748+
	(0.2222)	(1.0556)	(0.3076)	(1.0574)
Length × Nationality		0.1635***		0.1897***
		(0.0447)		(0.0443)
TopicDiversity × Nationality		0.9661***		0.9811***
		(0.2237)		(0.2179)
Pictures × Nationality		0.0754*		0.1175***
		(0.0351)		(0.0356)

UserPictures × Nationality		0.2974*** (0.0866)		0.2628** (0.0849)
Verified × Nationality			-0.6478*** (0.1758)	-0.5265** (0.1808)
Vine × Nationality			-0.7840*** (0.1941)	-1.0014*** (0.2066)
PastReview × Nationality			-0.4752 (2.3115)	-0.2782 (2.3577)
Nickname × Nationality			0.0458 (0.1124)	0.0202 (0.1149)
ProfilePhoto × Nationality			-0.1808* (0.0793)	-0.2356** (0.0813)
Review Time Effects	Yes	Yes	Yes	Yes
Product Fixed Effects	Yes	Yes	Yes	Yes
N	106,222	106,222	106,222	106,222
AIC	261,970	261,059	261,469	260,392
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.				

The effects of national culture are present in Models (2), (3), and (4) of Table 3.2. First, the results of Model (2) show that all the interaction terms of the argument quality variables are significantly positive, indicating that, with all other variables held constant, as a review is longer (*Length*), has more topics (*TopicDiversity*), includes more pictures (*Pictures*), or shows more user experience pictured (*UserPictures*), it tends to be perceived as being more helpful for US consumers. For example, one unit increase in topic diversity results in about 2.63 higher helpful votes of the review ($\rho^{TopicDiversity}$*: 0.9661, p <0.001*), and one more picture enhances about 1.08 votes ($\rho^{Pictures}$*: 0.0754, p <0.05*) on Amazon.com. This is because online consumers in an individualistic and low uncertainty avoidance culture are more likely to trust a reviewer and adopt the information of a review than those in the opposite culture. Thus, H1 is supported.

Interestingly, estimates of the interaction terms in Model (3) reveal opposite effects. The interaction coefficients of *Verified*, *Vine,* and *ProfilePhoto* ($\omega^{Verified}$*: -0.6478, p <0.001;* ω^{Vine}*: -0.7840, p <0.001;* $\omega^{ProfilePhoto}$*: -0.1808, p <0.05*) indicate that the three source credibility factors are more valuable for German consumers. This means that the three factors representing reviewer authenticity, reputation, and personal identity are more influential to consumers living in a culture high in collectivism and uncertainty avoidance, because the consumers tend to

mitigate source uncertainty when reading reviews more than consumers in the opposite culture do. Yet, the interaction effects of *PastReview* and *Nickname* are not statistically significant. Therefore, H2 is partially supported.

Remarkably, the two types of visual information showing a person–*UserPictures* and *ProfilePhoto*–are found to be differently helpful to consumers depending on their national cultures. User appearance in a picture is more valuable in an individualistic and low uncertainty avoidance culture ($\rho^{UserPictures}$*: 0.2974, p <0.001*), whereas a profile photo is preferred in a collectivistic and high uncertainty avoidance culture ($\omega^{ProfilePhoto}$*: -0.1808, p <0.05*). These findings are attributed to the fact that the user presence in a review picture provides product information and thus reduces product uncertainty, whereas a profile photo intensifying personal identity of the reviewer alleviates reviewer uncertainty. Therefore, H3 is supported. In Model (4), we included all variables, and all the results consistently hold.

3.6 Robustness Checks

3.6.1 Alternative Measurement and Topic Numbers for *TopicDiversity*

A practical question regarding our main results is whether they will consistently hold under different measurements and regression estimations. We performed several robustness checks to address this question. In the main analysis, we applied the TFIDF word weighting scheme and used 25 topics for *TopicDiversity*. In a robustness check, we re-measured *TopicDiversity* using 15 and 35 topics, respectively, and applying TFIDF and binary word weighting schemes (i.e., word occurrence), respectively. As shown in Table 3.3, the main results consistently hold across the different numbers of topics and word weighting schemes.

Table 3.3. Robustness Check 1–Alternative Measurement and Topic Numbers for *TopicDiversity*

	(1) TFIDF & 15 Topics	(2) TFIDF & 35 Topics	(3) Binary & 15 Topics	(4) Binary & 25 Topics	(5) Binary & 35 Topics
Length × Nationality	0.1715*** (0.0422)	0.2413*** (0.0454)	0.2204*** (0.0555)	0.2276*** (0.0595)	0.2760*** (0.0631)
TopicDiversity × Nationality	1.0272*** (0.2699)	1.1406*** (0.1809)	2.0788* (0.9126)	1.4656* (0.6897)	1.6345** (0.5865)
Pictures × Nationality	0.1276*** (0.0359)	0.1243*** (0.0345)	0.1414*** (0.0375)	0.1390*** (0.0373)	0.1423*** (0.0374)
UserPictures × Nationality	0.2621** (0.0859)	0.2544** (0.0826)	0.2623** (0.0896)	0.2640** (0.0889)	0.2614** (0.0892)
Verified × Nationality	-0.5242** (0.1828)	-0.5104** (0.1763)	-0.5550** (0.1904)	-0.5640** (0.1892)	-0.5709** (0.1898)
Vine × Nationality	-0.9943*** (0.2082)	-0.9721*** (0.2011)	-0.9914*** (0.2177)	-1.0043*** (0.2168)	-0.9831*** (0.2179)
PastReview × Nationality	-0.3823 (2.3878)	-0.2101 (2.2924)	-0.1138 (2.4929)	-0.0858 (2.4770)	-0.1126 (2.4861)
Nickname × Nationality	0.0370 (0.1162)	0.0201 (0.1117)	0.0421 (0.1214)	0.0388 (0.1206)	0.0375 (0.1210)
ProfilePhoto × Nationality	-0.2326** (0.0821)	-0.2330** (0.0792)	-0.2301** (0.0857)	-0.2298** (0.0851)	-0.2302** (0.0854)
N	106,222	106,222	106,222	106,222	106,222
AIC	261,087	259,197	262,159	262,132	261,928

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Product and time fixed effects are applied.
Detailed results are omitted due to the space limit but available from the authors upon request.

3.6.2 Alternative Estimations and More Distinguished Cultural Dimensions

First, we tested whether alternative estimations consistently support the main results estimated with negative binomial regression. To supplement evidence for our results, we employed two alternative estimations: Poisson and zero-inflated negative binomial (ZINB) regressions. Although statistics show that our dependent variable is over-dispersed, we also verified the main results using Poisson regression, which is useful for under-dispersed non-negative count variables. Also, we noticed that our dependent variable has excessive zero values (81% of the reviews have zero vote). Hence, we tested ZINB estimation as well where the same model is used for the zero-inflation model. Model (1) and Model (2) in Table 3.4 show outcomes of Poisson and ZINB estimations, respectively. The results of Poisson regression estimation are consistent with the main results. However, if we consider the excessive zero-voted reviews, the

interaction coefficients of *Length, Pictures, and Verified* become statistically insignificant. That is, there is no significant evidence that they are differently recognized across the national cultures. However, the interaction terms of *UserPictures* and *ProfilePhoto* are still statistically significant. Therefore, H3 is still supported while H1 and H2 are partially supported.

Table 3.4. Robustness Check 2–Alternative Estimations and Cultural Dimensions Back-up

	(1) Poisson	(2) ZINB	(3) New Votes	(4) New Votes + ZINB	(5) More Countries	(6) More Countries + ZINB
ExistingHelpfulVotes			0.0377***	0.0514***		
			(0.0013)	(0.0023)		
Length × Nationality	0.1885***	0.0153	-0.0685	0.0400	0.1715***	0.0635***
	(0.0097)	(0.0219)	(0.0718)	(0.0554)	(0.0218)	(0.0149)
TopicDiversity × Nationality	0.9202***	0.2597*	0.9597**	1.1234***	0.3460***	0.1591*
	(0.0478)	(0.1114)	(0.3353)	(0.2809)	(0.0874)	(0.0645)
Pictures × Nationality	0.0872***	0.0056	0.0006	-0.0518	0.1921***	0.0240*
	(0.0071)	(0.0174)	(0.0749)	(0.0585)	(0.0143)	(0.0110)
UserPictures × Nationality	0.3223***	0.2227***	0.5394***	0.9247***	0.3054***	0.2086***
	(0.0184)	(0.0525)	(0.1255)	(0.1181)	(0.0493)	(0.0423)
Verified × Nationality	-0.5307***	0.0101	-0.8992*	-1.0542***	-0.4439***	-0.3312***
	(0.0416)	(0.0801)	(0.3804)	(0.2329)	(0.1047)	(0.0600)
Vine × Nationality	-0.8999***	-0.2912**	-0.7383*	-0.2718	-0.6028***	-0.3396***
	(0.0471)	(0.1086)	(0.3563)	(0.2293)	(0.0806)	(0.0648)
PastReview × Nationality	-0.1449	0.6962	2.5670+	1.1073	-0.6237***	-0.1736**
	(0.5653)	(0.6110)	(1.3350)	(0.8486)	(0.0850)	(0.0550)
Nickname × Nationality	0.0115	0.0296	-0.1074	-0.0216	-0.1407**	-0.0472
	(0.0272)	(0.0488)	(0.1324)	(0.1095)	(0.0522)	(0.0331)
ProfilePhoto × Nationality	-0.3325***	-0.1149**	-0.5455***	-0.6128***	-0.1330***	-0.0864**
	(0.0181)	(0.0391)	(0.1440)	(0.1019)	(0.0375)	(0.0263)
N	106,222	106,222	90,135[#]	90,135[#]	314,814	314,814
AIC	293,668	154,505	105,375	52,735	772,821	422,344

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Product and time fixed effects are applied.
Detailed results are omitted due to the space limit but available from the authors upon request.
[#] During the one year May 26, 2020—May 28, 2021, the number of products in the data decreases from 783 to 593 and the number of reviews from 106,222 to 90,135 probably because of production and logistics problems due to COVID-19.

Second, although we controlled for the time effect, by including *DaysElapsed* in the model, considering that older reviews may have more opportunities to receive helpful votes, we further tested whether the national cultures are still effective for new helpful votes as well. We re-collected helpful votes for the reviews on May 28, 2021, so new votes received in one year (i.e.,

May 26, 2020—May 28, 2021) can be obtained. We also controlled for existing helpful votes (i.e., we used the dependent variable of the main analysis as an independent variable) in addition to the existing factors. As shown in Model (3) in Table 3.4, H1 and H2 are partially supported (*Length and Pictures* are statistically insignificant), and H3 is still supported. Also, the number of existing helpful votes is found to influence the new votes in a positive way (i.e., the rich get richer). In other words, even if we took into consideration the existing helpful votes, the new votes are more likely caused by pictured user experience in US reviews whereas by a profile photo in German reviews. Furthermore, we applied the alternative estimation ZINB for this analysis as well and got the same result (Model (4) in Table 3.4).

Third, we included more reviews from additional countries to better distinguish the cultural dimensions–reviews from the UK and Canada in line with US reviews, and reviews from Spain, Brazil and Poland in line with German reviews[15]. Interestingly, all the hypotheses are supported. As shown in Model (5), all the interaction coefficients for argument quality are significantly positive at the 0.1% level whereas those for source credibility are significantly negative. We again applied the ZINB estimation and obtained similar results in Model (6) except for the interaction coefficient of *Nickname*.

3.6.3 Mitigation of Endogeneity

Although we did robustness checks in various ways, our main analysis may not be free from an endogeneity concern because of (1) an unobserved factor that may influence both focal variables and helpful votes and (2) measurement error of image-driven features. To address the first issue (e.g., existing review posts of the product may influence a subsequent reviewer's behavior), we applied propensity score matching (PSM). PSM has been widely used to mitigate the systematic

[15] For these new reviews, we conducted image- and text-mining processes in the same way; we analyzed sentiment in their own languages and identified a human feature using the image-mining tools.

difference between treatment group (i.e., reviews with at least one user-appeared picture) and control group (i.e., reviews without a user-appeared picture). The propensity score is the probability assigning a treatment to an individual unit (here, a review) given observed covariates (Rosenbaum and Rubin 1983). This can be used to balance the two group samples based on the covariates and thus help us make an experiment-like setting. To calculate the score, we did not use the features used for our main analysis due to the violation of matching rule (Ho et al. 2007)– treatment (i.e., whether a user appears in a review picture) may not be the result of current review features such as star rating and review length. Instead, we generated features of explanatory variables based on previous reviews aggregated at the product level, e.g., average star rating of previous reviews of review *i*, to match the groups[16].

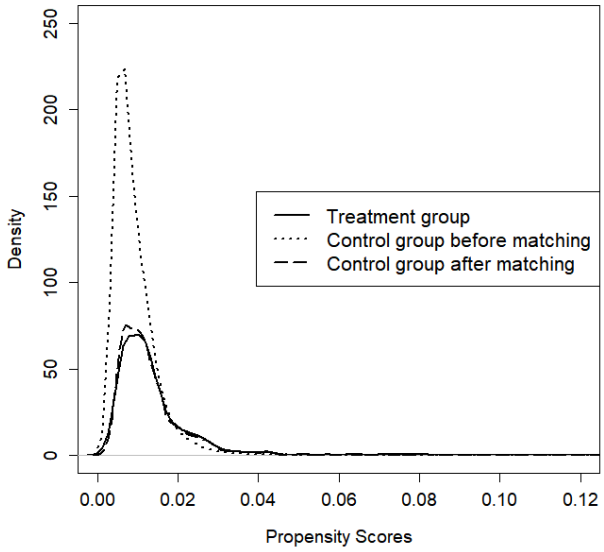


Figure 3.3. Matching Results of PSM

We ran a logit model to calculate the propensity score for each review and then matched two groups using nearest neighbor. Figure 3.3 shows the matching result (a covariate balance table is available from the authors upon request). We used a 1:10 ratio so that one review with a

[16] By using previous reviews, we can create an experiment-like setting reflecting (1) a reviewer's tendency to post a user-appeared picture and (2) existing characteristics of reviews of the product.

user-appeared picture is paired with 10 most similar reviews without a user-appeared picture. In our data, 1,020 reviews have at least one user-appeared picture, so our matched data sample has 11,220 observations in total.

$$Weight_i = \frac{TG}{PS_i} + \frac{1 - TG}{1 - PS_i} \quad (4)$$

Table 3.5. Robustness Check 3–Mitigation of Endogeneity

	(1) PSM	(2) PSW	(3) Measurement Error
Length × Nationality	0.0741** (0.0278)	0.0791*** (0.0120)	0.2129*** (0.0625)
TopicDiversity × Nationality	0.6789*** (0.1470)	0.3007*** (0.0464)	1.1830*** (0.3252)
Pictures × Nationality	0.0159 (0.0351)	-0.0171 (0.0264)	-0.2563*** (0.0594)
UserPictures × Nationality	0.1829* (0.0785)	0.1612* (0.0715)	1.4140*** (0.2084)
Verified × Nationality	-0.3182*** (0.0827)	-0.0141 (0.0283)	-0.4734+ (0.2619)
Vine × Nationality	-0.5961*** (0.1737)	-0.2527* (0.1149)	-0.5522+ (0.2901)
PastReview × Nationality	0.4262 (0.2954)	0.3688** (0.1251)	0.3164 (1.5740)
Nickname × Nationality	0.0475 (0.0378)	0.0005 (0.0089)	-0.1050 (0.1741)
ProfilePhoto × Nationality	-0.5789*** (0.0535)	-0.0285* (0.0115)	-0.6087** (0.1976)
N	11,220	106,222	106,222
AIC	23,174	308,462	N/A

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Product and time fixed effects are applied.
Detailed results are omitted but available from the authors upon request.

Nevertheless, PSM has a weakness, that is, reduced observations may not use the information contained in excluded data. To address this drawback (i.e., not to exclude any observations), we employed propensity score weighting (PSW) as well where the propensity score is used to compute an observation weight. Equation (4) introduces the inverse probability of treatment weighting (IPTW) method specifying how to derive the weight of review *i* where PS_i is the estimated propensity score of review *i* and TG is a dummy variable taking 1 if review *i*

is in the treatment group. Both PSM and PSW have an assumption of linearity (Ho et al. 2007, Rosenbaum and Rubin 1983), so we took a logarithm of votes and used linear regression. The outcomes of propensity score-based estimations are presented in Models (1) and (2) in Table 3.5, respectively, and robustness standard errors are reported. The estimates consistently indicate the same results; H3 is supported, whereas H1 and H2 are partially supported.

Lastly, we conducted measurement error correction using simulation extrapolation (SIMEX) proposed by Cook and Stefanski (1994) for the two image-oriented variables, *Profile* and *UserPictures*. We randomly selected 250 reviews, which have 200 profile photos and 523 review pictures in total. We asked three coders to label the true profile photos and user appearance in review pictures. One coder labeled 67 profile photos and 219 review pictures. The second coder labeled the remaining 133 profile photos and 304 review pictures. The third coder (one of the authors) independently labeled all the profile photos and review pictures. The Cohen's Kappa between the first two coders and the third coder was 94% and 92% for the profile photos and 96% and 98% for the review pictures, respectively, indicating excellent agreement. The third coder then discussed and debated with the other two coders to finalize the labeling results. As shown in Model (3) in Table 3.5, H3 is supported and H2 is partially supported. However, as the interaction term of *Pictures* has a negative coefficient (i.e., the attached pictures are deemed more helpful for German consumers), H1 is no longer supported.

3.7 Discussions

Our findings suggest that the importance of trust antecedents in online reviews (i.e., argument quality and source credibility) tends to be differently perceived by online consumers across national cultures. Review content that enriches argument quality tends to be perceived to be helpful by US consumers, who have a culture high in individualism and low in uncertainty

avoidance. On the other hand, when consumers are more likely to avoid uncertainty and be collectivistic (e.g., German consumers), source credibility factors are more helpful. The former consumers tend to rely on information quality rather than checking a message source because they are inclined to trust even strangers and rather focus on argument contents for their purchase decisions. Yet, the latter consumers would pay attention to a message source before adopting information as they do not tend to believe in strangers' opinions.

Among the findings, one is particularly noteworthy: a user appearing visually in reviews would be recognized differently across national cultures. Both *UserPictures* and *ProfilePhoto* capture a human object in an image, but the presence of a user in a review picture is more helpful to US consumers whereas a profile photo is more helpful to German consumers. We can explain this phenomenon using the uncertainty reduction concept. On the one hand, user appearance in a review picture can mitigate uncertainty about product quality. For example, a flexibility feature can be disclosed by movement of a user, and the real color and fit of a product can be visually clarified–they are likely to improve the information quality of the review. On the other hand, this user appearance does not always reduce the uncertainty of a message source–only small parts of the human body highlighting a product (e.g., only feet wearing socks) can be captured. In contrast, a profile photo tends to center a reviewer rather than a product and hence show personal identity. This characteristic strengthens the source credibility of a review and reduces source uncertainty.

Also, by testing the effects of review pictures and profile photos on new votes collected during early COVID-19 (May 2020 ~ May 2021), we found increased importance of a user appearing visually in reviews. As shown in the main analysis, before COVID-19, one more review picture with user appearance and one attached profile photo turned out to improve review

helpfulness by 1.29 and 1.27 votes, respectively, in the US and Germany. The effects are enlarged in the pandemic era–the picture and profile photo improve helpfulness by 2.36 and 1.38, respectively, as shown in Model (3) in Table 3.4. This is possibly because of (1) higher reliance on visual information and (2) restricted travel between countries. Nowadays, online consumers get used to optical information processing from videos and images and thus may have increased tendency to view pictures and reduced tendency to read reviews. To comply with this trend, Amazon launched a live-streaming service (Amazon Live) in early 2019 as well as providing consumers a link to see only consumer-uploaded images on the review platform. Moreover, during the pandemic, this phenomenon is combined with cultural effects. As people more often stay in one country because of travel restriction and social distancing, they are more likely to get influenced by the culture where they live. Therefore, US online consumers rely more on user appearance in review pictures to get product information whereas less on text-based information. In a similar vein, German online consumers also show higher reliance on profile photos.

3.8 Contributions and Limitations

Our research contributes to the growing literature on online reviews by combining cultural aspects and trust through text and image analytics. In quantitative empirical research, online reviews have rarely been analyzed across countries and/or cultures due to the language issue. To overcome such a problem, we employed text analytics such as sentiment analysis and topic modeling. Without translating languages, which may result in information loss and distortion of the original meaning, we extracted text-oriented features and took the first step in consumer review analysis across national cultures. Moreover, we mined user appearance from review pictures and included picture-based control variables. By employing image analytics, this

research can inspire future studies to investigate the effects of visual information (e.g., color, sentiment) of online reviews. Overall, we expanded the perspectives of review helpfulness research beyond one country and typical data analytics approach.

As for business implications, we recommend that e-commerce businesses consider culture-specific strategies in managing their review systems. For example, even before a review has received enough helpful votes from consumers, review platform managers can place it in an appropriate position considering the national culture where they sell the product (i.e., reviews conveying a great amount of information in the US and reviews ensuring reviewer credibility in Germany can be displayed in more prominent positions). Also, when sellers target purchasers (or provide their products for free or at discounted prices) to encourage review writing, they may adopt different strategies based on national cultures: in Germany, sellers may target those who have higher credibility on the review platform, whereas in the US, they may target those who tend to write meticulous and informative reviews.

Our work has several limitations, which may be addressed in future research. First, while we examined two western countries (i.e., the US and Germany) in the main analysis and a few more countries in a robustness check, future research may extend the investigation to more differing countries to validate the generalizability of our findings. Second, we used the number of helpful votes a review has received as the measure for review helpfulness, which is different from the original measure (i.e., the percentage of people who found the review helpful) provided by Mudambi and Schuff (2010), because Amazon only allows helpful votes now and does not allow unhelpful votes anymore. With data from platforms that also allow unhelpful votes, future research may apply a percentage measure (Forman et al. 2008, Mudambi and Schuff 2010) to test the robustness of our findings with respect to the helpfulness measurement. Third, there

could be more text-mined features, such as readability and spelling errors, which we did not employ because there is currently no tool for them that is compatible across all the languages we examined. Future research may develop multilingual tools to measure such text-based features reliably and consistently across languages. Fourth, in-depth image analysis, for example, more objects extracted from a review picture (besides humans) and picture-based sentiment (as a counterpart of text-based sentiment), can further extend this research. Fifth, it is infeasible to add an additional variable measuring trust in this research to fully examine how different antecedents first influence trust and then influence helpfulness, because Amazon does not provide data about review readers. Addressing this limitation would require another study with experiments or surveys. It will be interesting for future research to further explore how the role of consumer trust in review helpfulness differs and how innate tendency of review readers to trust a review influences review helpfulness across national cultures. Sixth, we targeted clothing products (i.e., experience goods) in this study because reviewers are more likely to upload human-included pictures along with their review texts than reviewers for search goods (e.g., laptop, smartphone), allowing us to extract the user appearance feature better. Future research may extend the examination by comparing the trust antecedent effects in search goods versus experience goods.

REFERENCES

Adaval, R., Isbell, L. M., and Wyer, R. S. 2007. "The Impact of Pictures on Narrative-and List-Based Impression Formation: A Process Interference Model," *Journal of Experimental Social Psychology* (43:3), pp. 352-364.

Adaval, R., and Wyer, R. S. 1998. "The Role of Narratives in Consumer Information Processing," *Journal of Consumer Psychology* (7:3), pp. 207-245.

Aggarwal, R., Midha, V., and Sullivan, N. 2021. "Superlatives and Scope of Improvement in Online Recommendations: Breath of Life or a Kiss of Death?," MIS Quarterly (45:3), pp. 1411-1432.

An, Q., Ma, Y., Du, Q., Xiang, Z., and Fan, W. 2020. "Role of user-generated photos in online hotel reviews: An analytical approach," *Journal of Hospitality and Tourism Management* (45), pp. 633-640.

Archak, N., Ghose, A., and Ipeirotis, P. G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management Science* (57:8), pp. 1485-1509.

Awad, N. F., and Ragowsky, A. 2008. "Establishing Trust in Electronic Commerce Through Online Word of Mouth: An Examination Across Genders," Journal of Management Information Systems (24:4), pp. 101-121.

Ba, S., and Pavlou, P.A. 2002. "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior," MIS Quarterly (26:3), pp. 243-268.

Barbro, P. A., Mudambi, S. M., and Schuff, D. 2020. "Do country and culture influence online reviews? An analysis of a multinational retailer's country-specific sites," *Journal of International Consumer Marketing* (32:1), pp. 1-14.

Berger, J., and Fitzsimons, G. 2008. "Dogs on the Street, Pumas on Your Feet: How Cues in the Environment Influence Product Evaluation and Choice," *Journal of Marketing Research* (45:1), pp. 1-14.

Biswas, B., Sengupta, P., and Ganguly, B. 2022. "Your reviews or mine? Exploring the determinants of "perceived helpfulness" of online reviews: a cross-cultural study," *Electronic Markets* (32), pp. 1083-1102.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," Journal of Machine Learning Research, 3(Jan), pp. 993-1022.

Bond, C. 2019. "User-Generated Images Come to Google Shopping Reviews," WordStream (https://www.wordstream.com/blog/ws/2019/10/16/user-generated-images-google-shopping-reviews).

Boring, E. G. 1946. "The Perception of Objects," *American Journal of Physics* (14:2), pp. 99-107.

Brand, B. M., and Reith, R. 2022. "Cultural Differences in the Perception of Credible Online Reviews–The Influence of Presentation Format," Decision Support Systems (154), pp. 113710.

Cameron, A. C., and Trivedi, P. K. 2013. *Regression Analysis of Count Data,* Cambridge, UK: Cambridge University Press.

Cao, Q., Duan, W., and Gan, Q. 2011. "Exploring Determinants of Voting for the "Helpfulness" of Online User Reviews: A Text Mining Approach," *Decision Support Systems* (50:2), pp. 511-521.

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), pp. 752-766.

Chen, P. Y., Hong, Y., and Liu, Y. 2018. "The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments," *Management Science* (64:10), pp. 4629-4647.

Cheung, C. M. Y., Sia, C. L., and Kuan, K. K. 2012. "Is This Review Believable? A Study of Factors Affecting the Credibility of Online Consumer Reviews from an ELM Perspective," Journal of the Association for Information Systems (13:8), pp. 618-635.

Childers, T. L., and Houston, M. J. 1984. "Conditions for a Picture-Superiority Effect on Consumer Memory," *Journal of Consumer Research* (11:2), pp. 643-654.

Chou, Y. C., Chuang, H. H. C., and Liang, T. P. 2022. "Elaboration likelihood model, endogenous quality indicators, and online review helpfulness," *Decision Support Systems* (153), 113683.

Cook, J. R., and Stefanski, L. A. 1994. "Simulation-Extrapolation Estimation in Parametric Measurement Error Models," *Journal of the American Statistical Association* (89:428), pp. 1314-1328.

Ceylan, G., Diehl, K., and Proserpio, D. 2021. "Words Meet Photos: When and Why Visual Content Increases Review Helpfulness," Available at (https://ssrn.com/abstract=3896569).

Daugherty, T., Li, H., and Biocca, F. 2005. "Experiential Ecommerce: A Summary of Research Investigating the Impact of Virtual Experience on Consumer Learning," in *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World,* C. Haugtvedt, K. Machleit and R. Yalch (eds.), Mahwah, NJ: Lawrence Erlbaum Associates, pp. 425-459.

De, P., Hu, Y., and Rahman, M. S. 2013. "Product-Oriented Web Technologies and Product Returns: An Exploratory Study," *Information Systems Research* (24:4), pp. 998-1010.

Deloitte Consumer Review. 2016. "The Deloitte Consumer Review Digital Predictions 2016," Deloitte (https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/consumer-business/deloitte-uk-consumer-review-march-2016.pdf).

Doney, P. M., Cannon, J. P., and Mullen, M. R. 1998. "Understanding the Influence of National Culture on the Development of Trust," The Academy of Management Review (23:3), pp. 601-620.

Dong, B., Li, M., and Sivakumar, K. 2019. "Online Review Characteristics and Trust: A Cross-Country Examination," Decision Sciences (50:3), pp. 537-566.

Duan, W., Gu, B., and Whinston, A. B. 2009. "Informational Cascades and Software Adoption on the Internet: An Empirical Investigation," *MIS Quarterly* (33:1), pp. 23-48.

Epstein, W. 1973. "The Process of 'Taking-into-Account' in Visual Perception," *Perception* (2:3), pp. 267-285.

Filieri, R. 2016. "What Makes an Online Consumer Review Trustworthy?," Annals of Tourism Research (58), pp. 46-64.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp. 291-313.

Fukuyama, F. 1995. Trust: The Social Virtues and the Creation of Prosperity, New York, Free Press.

Garvin, D. A. 1984. "What Does "Product Quality" Really Mean?" *Sloan Management Review* (26:1), pp. 25-43.

Gefen, D., and Heart, T. 2006. "On the Need to Include National Culture as a Central Issue in E-Commerce Trust Beliefs," Journal of Global Information Management (14:4), pp. 1-30.

Ghose, A., Goldfarb, A., and Han, S. P. 2013. "How Is the Mobile Internet Different? Search Costs and Local Activities," *Information Systems Research* (24:3), pp. 613-631.

Ghose, A., and Ipeirotis, P. G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering* (23:10), pp. 1498-1512.

Gong, J., Abhishek, V., and Li, B. 2018. "Examining the Impact of Keyword Ambiguity on Search Advertising Performance: A Topic Model Approach," MIS Quarterly (42:3), pp. 805-830.

Harrigan, M., Feddema, K., Wang, S., Harrigan, P., and Diot, E. 2021. "How Trust Leads to Online Purchase Intention Founded in Perceived Usefulness and Peer Communication," Journal of Consumer Behaviour (20:5), pp. 1297-1312.

Herr, P. M., Kardes, F. R., and Kim, J. 1991. "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective," Journal of Consumer Research (17:4), pp. 454-462.

Higgins, E. T., Rholes, W. S., and Jones, C. R. 1977. "Category Accessibility and Impression Formation," *Journal of Experimental Social Psychology* (13:2), pp. 141-154.

Ho D. E., Imai K., King G., and Stuart E. A. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference," Political Analysis (15:3), pp. 199-236.

Hofstede, G. 1983. "National Cultures in Four Dimensions: A Research-Based Theory of Cultural Differences Among Nations," International Studies of Management & Organization (13:1-2), pp. 46- 74.

Holway, A. H., and Boring, E. G. (1941). "Determinants of Apparent Visual Size with Distance Variant," *American Journal of Psychology* (54:1), pp. 21-37.

Hong, H., Xu, D., Wang, G. A., and Fan, W. 2017. "Understanding the determinants of online review helpfulness: A meta-analytic investigation," *Decision Support Systems* (102), pp. 1-11.

Hornikx, J., and Hoeken, H. 2007. "Cultural Differences in the Persuasiveness of Evidence Types and Evidence Quality," Communication Monographs (74:4), pp. 443-463.

Hornikx, J., and ter Haar, M. 2013. "Evidence Quality and Persuasiveness: Germans Are Not Sensitive to the Quality of Statistical Evidence," Journal of Cognition and Culture (13:5), pp. 483-501.

Jensen, M. L., Averbeck, J. M., Zhang, Z., and Wright, K. B. 2013. "Credibility of anonymous online product reviews: A language expectancy perspective," *Journal of Management Information Systems* (30:1), pp. 293-324.

John, S. 2020. "How to Write a Review on Amazon in 5 Simple Steps," Business Insider (https://www.businessinsider.com/how-to-write-a-review-on-amazon; accessed July 3, 2020).

Karimi, S., and Wang, F. 2017. "Online Review Helpfulness: Impact of Reviewer Profile Image," *Decision Support Systems* (96), pp. 39-48.

Kavanagh, M. 2021. "The Impact of Customer Reviews on Purchase Decisions," Bizrate Insights (https://www.bizrateinsights.com/resources/shopper-survey-report-the-impact-reviews-have-on-consumers-purchase-decisions/).

Kim, J. M., Jun, M., and Kim, C. K. 2018. "The effects of culture on consumers' consumption and generation of online reviews," *Journal of Interactive Marketing* (43:1), pp. 134-150.

Kim, Y., Wang, Y., and Ghoshal A. 2019. "Imagine It or Image It? The Effectiveness of User Generated Reviews in Pictures," In Proceedings of the 2019 Workshop on Information Technologies and Systems, Munich, Germany.

King, J. 2018. “For Online Shoppers, Photos Can Influence a Purchase,” eMarketer (https://www.emarketer.com/content/for-online-shoppers-photos-can-influence-a-purchase).

Kokkodis, M., and Lappas, T. 2016. “The relationship between disclosing purchase information and reputation systems in electronic markets,” In Proceedings of the 37th International Conference on Information Systems, Dublin, Ireland.

Kutner, M. H., Nachtsheim, C., and Neter, J. 2004. Applied Linear Regression Models, McGraw-Hill/Irwin.

Lancaster, K. J. 1971. *Consumer Demand: A New Approach*, New York, Columbia University Press.

Lantzy, S., Hamilton, R. W., Chen, Y. J., and Stewart, K. 2021. “Online reviews of credence service providers: what do consumers evaluate, do other consumers believe the reviews, and are interventions needed?” *Journal of Public Policy & Marketing* (40:1), pp. 27-44.

Lee, J., and Hong, I. B. 2019. “Consumer’s Electronic Word-of-Mouth Adoption: The Trust Transfer Perspective,” International Journal of Electronic Commerce (23:4), pp. 595-627.

Lee, A. Y., and Labroo A. A. 2004. “The Effect of Conceptual and Perceptual Fluency on Brand Evaluation,” *Journal of Marketing Research* (41:2), pp. 151-165.

Lee, E., and Zhao, H. 2020. “Deriving Topic-Related and Interaction Features to Predict Top Attractive Reviews for a Specific Business Entity,” Journal of Business Analytics (3:1), pp. 17-31.

Lei, Z., Yin, D., and Zhang, H. 2021. “Focus Within or On Others: The Impact of Reviewers’ Attentional Focus on Review Helpfulness,” *Information Systems Research* (32:3), pp. 801-819.

Leong, L. Y., Hew, T. S., Ooi, K. B., Chong, A. Y. L., and Lee, V. H. 2021. “Understanding trust in ms-commerce: The roles of reported experience, linguistic style, profile photo, emotional, and cognitive trust,” *Information & Management* (58:2), 103416.

Li, M., Huang, L., Tan, C.-H., and Wei, K.-K. 2013. “Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features,” International Journal of Electronic Commerce (17:4), pp. 101-136.

Li, C., Kwok, L., Xie, K. L., Liu, J., and Ye, Q. 2021. “Let photos speak: The effect of user-generated visual content on hotel review helpfulness,” *Journal of Hospitality & Tourism Research*, 10963480211019113.

Lin, T. M., Lu, K., and Wu, J. 2012. “The Effects of Visual Information in eWOM Communication,” Journal of Research in Interactive Marketing 6(1), pp. 7-26.

Liu, X., Lee, D., and Srinivasan, K. 2019. "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *Journal of Marketing Research* (56:6), pp. 918-943.

Liu, Z., and Park, S. 2015. "What makes a useful online review? Implication for travel product websites," *Tourism Management* (47), pp. 140-151.

Ma, Y., Xiang, Z., Du, Q., and Fan, W. 2018. "Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning," *International Journal of Hospitality Management* (71), pp. 120-131.

Mayer, R. C., Davis, J. H., and Schoorman, F. D. 1995. "An Integrative Model of Organizational Trust," Academy of Management Review (20:3), pp. 709-734.

Mayfield, J. 2022. "New Data Shows FTC Received 2.8 Million Fraud Reports from Consumers in 2021," Available at https://www.ftc.gov/news-events/news/press-releases/2022/02/new-data-shows-ftc-received-28-million-fraud-reports-consumers-2021-0.

Mudambi, S. M., and Schuff, D. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," *MIS Quarterly* (34:1), pp. 185-200.

Pan, Y., and Zhang, J. Q. 2011. "Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews," Journal of Retailing (87:4), pp. 598-612.

Park, J., Gunn, F., and Han, S. L. 2012. "Multidimensional Trust Building in E-retailing: Cross-Cultural Differences in Trust Formation and Implications for Perceived Risk," Journal of Retailing and Consumer Services (19:3), pp. 304-312.

Park, C. W., Sutherland, I., and Lee, S. K. 2021. "Effects of online reviews, trust, and picture-superiority on intention to purchase restaurant services," *Journal of Hospitality and Tourism Management* (47), pp. 228-236.

Peschel, A. O., and Orquin, J. L. 2013. "A Review of the Findings and Theories on Surface Size Effects on Visual Attention," *Frontiers in Psychology* (4), pp. 21-30.

Petro, G. 2021. "Online Retail's Hidden Post-Covid Cost: Returns," Forbes (https://www.forbes.com/sites/gregpetro/2021/03/02/online-retails-hidden-post-covid-cost-returns/?sh=1b9d5a217002).

Petty, R. E., and Cacioppo, J. T. 1986. Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York: Springer-Verlag.

Pooja, K., and Upadhyaya, P. 2022. "What makes an online review credible? A systematic review of the literature and future research directions," *Management Review Quarterly*, pp. 1-33.

Pornpitakpan, C., and Francis, J. N. 2000. "The Effect of Cultural Differences, Source Expertise, and Argument Strength on Persuasion: An Experiment with Canadians and Thais," Journal of International Consumer Marketing (13:1), pp. 77-101.

Potter, M. C., Wyble, B., Hagmann, C. E., and McCourt, E. S. 2014. "Detecting Meaning in RSVP at 13 ms Per Picture," *Attention, Perception, & Psychophysics* (76:2), pp. 270-279.

Qahri-Saremi, H., and Montazemi, A. R. 2019. "Factors Affecting the Adoption of an Electronic Word of Mouth Message: A Meta-Analysis," Journal of Management Information Systems (36:3), pp. 969-1001.

Qin, L., and De-Juan-Vigaray, M. D. 2021. "Social commerce: Is interpersonal trust formation similar between USA and Spain?" *Journal of Retailing and Consumer Services* (62), 102642.

Racherla, P., Mandviwalla, M., and Connolly, D. J. 2012. "Factors Affecting Consumers' Trust in Online Product Reviews," Journal of Consumer Behaviour (11:2), pp. 94-104.

Ritchie, E., and Phares, E. J. 1969, ''Attitude Change as a Function of Internal-External Control and Communicator Status,'' Journal of Personality (37:3), pp. 429-443.

Rosenbaum P. R., and Rubin D. B. 1983. "The central role of the propensity score in observational studies for causal effects," Biometrika (70:1), pp. 41-55.

Sahoo, N., Dellarocas, C., and Srinivasan, S. 2018. "The Impact of Online Product Reviews on Product Returns," *Information Systems Research* (29:3), pp. 723-738.

Salehan, M., and Kim, D. J. 2016. "Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach to Big Data Analytics," Decision Support Systems 81, pp. 30-40.

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., and Lee, K. C. 2020. "Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach," *MIS Quarterly* (44:4), pp. 1459-1492.

Shukla, A., and Mishra, A. 2021. "Effects of visual information and argument concreteness on purchase intention of consumers towards online hotel booking," *Vision*.

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., and Roy, P. K. 2017. "Predicting the "Helpfulness" of Online Consumer Reviews," Journal of Business Research 70, pp. 346-355.

Son, J., Negahban, A., and Chiang, D. 2019. "Topic Diversity of Online Consumer Reviews and Its Effect on Review Helpfulness," in Proceeding of the 25th Americas Conference on Information Systems, Cancun, Mexico.

Song, Y., Li, Z., and Sahoo, N. 2022. "Matching Returning Donors to Projects on Philanthropic Crowdfunding Platforms," Management Science (68:1), pp. 355-375.

Srite, M., and Karahanna, E. 2006. "The Role of Espoused National Cultural Values in Technology Acceptance," MIS Quarterly (30:3), pp. 679-704.

Standing, L., Conezio, J., and Haber, R. N. 1970. "Perception and Memory for Pictures: Single-Trial Learning of 2500 Visual Stimuli," *Psychonomic Science* (19:2), pp. 73-74.

Straub, D., Loch, K., Evaristo, R., Karahanna, E., and Srite, M. 2002. "Toward a theory-based measurement of culture," *Journal of Global Information Management* (10:1), pp. 13-23.

Tang, L. 2017." Mine your customers or mine your business: the moderating role of culture in online word-of-mouth reviews," *Journal of International Marketing* (25:2), pp. 88-110.

Triandis, H. C. 1995. Individualism & Collectivism, Westview Press.

Ulrich, K. T. 2011. *Design: Creation of Artifacts in Society,* Philadelphia, University of Pennsylvania.

Vishwanath, A. 2003. "Comparing online information effects: A cross-cultural comparison of online information and uncertainty avoidance," *Communication Research* (30:6), pp. 579-598.

Walk-Morris, T. 2019. "Consumers' common return reason? Wrong color, size or fit," Industry Dive (https://www.retaildive.com/news/consumers-common-return-reason-wrong-color-size-or-fit/569328/).

Walker, B. A., and Olson, J. C. 1997. "The Activated Self and Consumer Behavior: A Cognitive Structure Perspective," Research in Consumer Behavior (8:2), pp. 135-171.

Wang, Y., Ramachandran, V., and Liu Sheng, O. R. 2021. "Do Fit Opinions Matter? The Impact of Fit Context on Online Product Returns," *Information Systems Research* (32:1), pp. 268-289.

Weathers, D., Swain, S. D., and Grover, V. 2015. "Can Online Product Reviews Be More Helpful? Examining Characteristics of Information Content by Product Type," Decision Support Systems (79), pp. 12-23.

Wu, J. 2017. "Review Popularity and Review Helpfulness: A Model for User Review Effectiveness," Decision Support Systems (97), pp. 92-103.

Xu, Q. 2014. "Should I Trust Him? The Effects of Reviewer Profile Characteristics on eWOM Credibility," Computers in Human Behavior (33), pp. 136-144.

Xu, P., Chen, L., and Santhanam, R. 2015. "Will Video Be the Next Generation of E-Commerce Product Reviews? Presentation Format and the Role of Product Type," Decision Support Systems (73), pp. 85-96.

Yang, M., Adomavicius, G., Burtch, G., and Ren, Y. 2018. "Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated via Data Mining," *Information Systems Research* (29:1), pp. 4-24.

Yang, Y., Wang, Y., and Zhao, J. 2023. “Effect of user-generated image on review helpfulness: Perspectives from object detection,” *Electronic Commerce Research and Applications* (57), 101232.

Yin, D., Bond, S. D., and Zhang, H. 2021. "Anger in Consumer Reviews: Unhelpful but Persuasive?," *MIS Quarterly* (45: 3), pp.1059-1086.

Yin, D., Bond, S. D., and Zhang, H. 2014. “Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews,” *MIS Quarterly* (38:2), pp. 539-560.

Yin, D., Bond, S. D., and Zhang, H. 2017. “Keep Your Cool or Let It Out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews,” *Journal of Marketing Research* (54:3), pp. 447-463.

Yin, D., Mitra, S., and Zhang, H. 2016. “Research Note—When Do Consumers Value Positive vs. Negative Reviews? An Empirical Investigation of Confirmation Bias in Online Word of Mouth,” *Information Systems Research* (27:1), pp. 131-144.

Zhang, S., Lee, D., Singh, P., and Srinivasan, K. 2021. “What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features,” *Management Science* (68:8), pp. 5644-5666.

Zhao, Y., Wang, L., Tang, H., and Zhang, Y. 2020. “Electronic word-of-mouth and consumer purchase intentions in social e-commerce,” *Electronic Commerce Research and Applications* (41), 100980.

Zhou, S., and Guo, B. 2017. “The Order Effect on Online Review Helpfulness: A Social Influence Perspective,” *Decision Support Systems* (93), pp. 77-87.

Zhu, Y., Liu, M., Zeng, X., and Huang, P. 2020. “The effects of prior reviews on perceived review helpfulness: A configuration perspective,” *Journal of Business Research* (110), pp. 484-494.

Zhu, L., Yin, G., and He, W. 2014. “Is This Opinion Leader's Review Useful? Peripheral Cues for Online Review Helpfulness,” Journal of Electronic Commerce Research (15:4), pp. 267-280.

Zinko, R., de Burgh-Woodman, H., Furner, Z. Z., and Kim, S. J. 2021. “Seeing Is Believing: The Effects of Images on Trust and Purchase Intent in eWOM for Hedonic and Utilitarian Products,” Journal of Organizational and End User Computing (33:2), pp. 85-104.

APPENDICES

Appendix A: Lists of Detected Objects in Visual Situational Cues

For the 857 pictures in the raw data, we extracted 1,768 objects and used 127 to construct the pool for the situational cues in the review images. Table A enumerates detailed objects pertaining to situational cues (i.e., nature, plant, sport, and defensive structure) recognized in the machine learning algorithms. Figure A presents several examples of *Visual Situational Cues*.

Table A. Image-Mined Object

Classes	Detected Object
Nature	alp, beach, campsite, canyon, cave, cavern, cliff, crag, creek, crest, crevasse, deep ocean, desert, downhill, forest, frozen waterbody, glacier, gorge, highland, hill, hillock, hillside, ice, ice floe, icecap, ice mass, lake, lakefront, lakeside, landfill, ledge, meander, mound of stones, mountain, mountain path, mountain range, mountain ridgeline, mountainside, natural elevation, ocean, ocean floor, park, plateau, pond, precipice, rainforest, ravine, ridge, river, riverbank, riverbed, rock, rock arch, rock outcrop, sand dune, sandbank, sands, seashore, shoreline, ski slope, slope, snow scene, snowdrift, snowfield, snowy mountains, steep, stream, swamp, uphill, valley, waterfalls, waterside.
Plant	flower, forest, grass, tree (e.g., oak, pine tree, bamboo, eucalyptus, logwood, snag, baobab, alpine ash, herb), vine.
Sports	athletic game, ball game, belay, bicycling, biking, boat racing, climbing, court game, cross-country skiing, downhill skiing, dune cycling, fishing, fly-fishing, kayaking, motorbike racing, motorcycling, outdoor sport, racing, riding, rowing, skateboarding, skating, ski jumping, skiing, sledding, snorkeling, surfing, swimming, trekking, water sport, water-skiing.
Defensive Structure	escarpment, fortification.

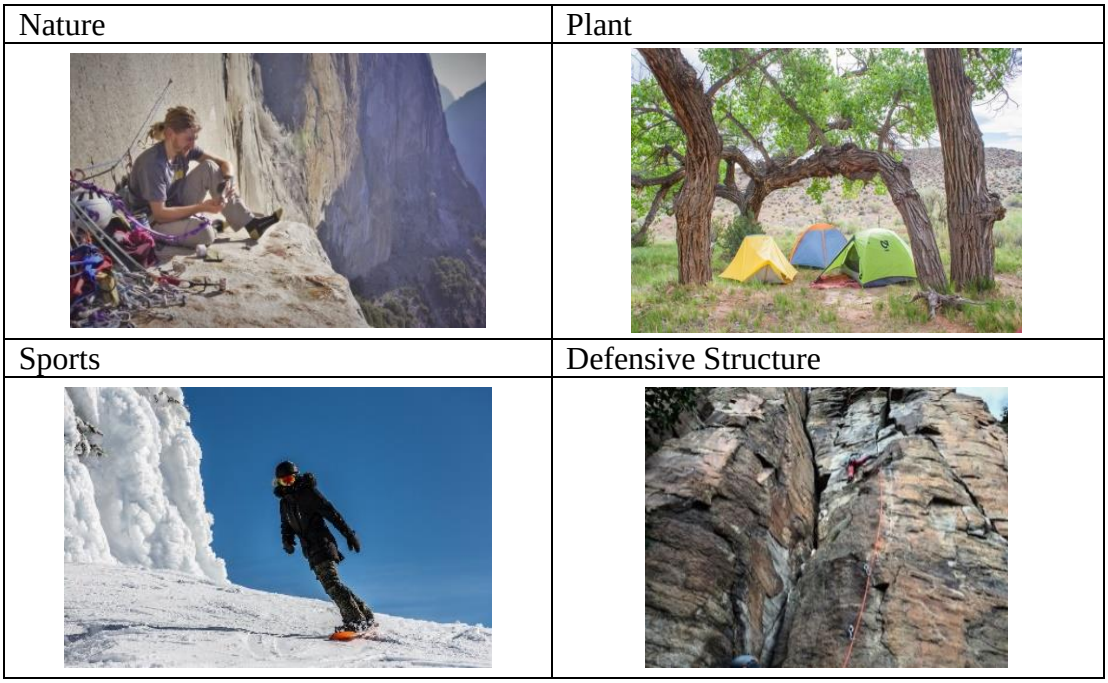


Figure A. Examples of Visual Situational Cues

Appendix B: Keyword Lists for the Text Mining Tasks

Table B1 presents all the expressions and keywords we used to extract the situational cues based on the 4,179 most frequent terms appearing in the review text. Although we included all parts of speech (i.e., noun, verb, adjective) during the text mining procedure, we present the most frequently appeared form in the list. Also, when a word is a gerund, we did not account for the other words that share the same root because they may have different meanings, and therefore, do not represent a context. For example, *mountaineering* means an activity, while a *mountaineer* represents a person who is climbing. As another example, *rappelling* denotes sliding down a cliff using a rope, while *rappel* is a method of moving down a steep.

Table B.1. Text-Mined Keyword List–Situational Cues

Aspects	Keyword List	Example from Actual Reviews
Nature and Plant	alpine, avalanche, backyard, basecamp, beach, boulder, campsite, canyon, cliff, coast, crags, creeks, crevasse, desert, downhill, elevation, environment, forest, garden, glacier, grass, hill, ice, lake, lawn, ledge, mountain, mud, nature, ocean, outdoors, park, reservoir, river, sandstone, slush, steeps, terrain, trailhead, trail, valley, waterfall, waters, weather.	*"I've skied in the wet thawing **slush** of March."* *"This is best suited for a drier **environment**."* *"Used it **elevation** ranging from 6~ 9,700 ft."*
Nature: Seasonality	chill, climate, cold, drizzle, F, freeze, hail, humidity, monsoon, rain, rainstorm, season, sleet, snow, spring, storm, summer, sun, temp, temperature, winter.	*"Perfect to keep warm on a **chilly** day."*
Activity & Defensive Structure	activity, adventure, backpacking, baseball, belaying, biking, boarding, bouldering, camp, canoeing, canyoneering, climb, commute, cragging, cycling, excursion, exercise, expedition, exploring, fishing, football, hike, hooking, kayaking, marathon, mount, mountaineering, outing, paddling, pedaling, race, rafting, rappelling, ride, roping, skate, ski, sledding, snowboard, snowmobiling, snowshoeing, sport, surfing, swim, tennis, tour, travel, traverse, trek, trip, walk, workout, yoga.	*"I wear them for everything; **kayaking**, **rafting**, **fishing**, **canyoneering**, and **hiking**."* *"Great for any type of outdoor **adventure**."* *"I guess it's not all that necessary for day **hikes** or short **outings**, but on longer **trips** it's essential."*

Table B2 details all the keywords we used to indicate the three categorized product features: functionality, aesthetics, and durability. Because some keywords are category-specific (e.g., the word *sensitivity* is mainly used to describe outdoor gears while the word *shrinkage* is used to describe apparel), we separately present the wordlist of outdoor gear and the wordlist of apparel. We considered all nouns, verbs, and adjectives when extracting the keywords. For example, both words – *durable* and *durability* – would be regarded as the keyword for product durability.

Table B.2. Text-Mined Keyword List–Product Features

Categories	Aspects	Keyword List	Example from Actual Reviews
Apparel	Aesthetics	appearance, baggy, beauty, boxy, bulky, fancy, fashionable, fit, floppy, fluffy, logo, looking, loose, oversized, pattern, print, puffy, saggy, sleek, slim, sloppy, snug, style, thick. (color) black, blue, bright, brown, color, dark, green, gray, grey, orange, pink, purple, red, white, yellow.	*"Not too **thick** yet incredibly warm for even the coldest temperatures."* *"it's just a good **sleek look**!"* *"the **color**/**print** is awesome."*
	Functionality	adjustability, breathability, clammy, comfortability, compression, damp, dry, moisture-proof, elasticity, extendable, flexibility, insulation, overheating, packability, roomy, soak, stretch, squeeze, squish, sweat, ventilation, versatility, warmth, washability, waterproof, windproof.	*"This jacket is very **versatile** especially if you are heading into different weather zones."*
	Durability	durability, fade, feather, pilling, quality, reliability, rip, shrinkage, sturdiness, tear.	*"**Durable**, **reliable**, not smell."*
Gear	Aesthetics	appearance, baggy, beefy, beauty, bulky, curvature, fancy, fashionable, fit, floppy, fluffy, logo, looking, loose, oversized, pattern, precision, print, puffy, saggy, sensitivity, sleek, slim, sloppy, snug, style, vintage. (color) black, blue, bright, brown, color, dark, green, gray, grey, neon, orange, pink, purple, red, silver, tint, white, yellow.	*"So, if you want a **sensitive** and **precise** shoe for wide or narrow feet, this is it."* *"He likes the aggressive **curvature** and says they are more comfortable"*
	Functionality	accommodation, adjustability, breathability, capacity, capability, clammy, comfortability, compression, damp, dry, moisture-proof, elasticity, extendable, extendibility, flexibility, friction, inflatability, insulation, overheating, packability, roomy, slippage, soak, stretch, squeeze, squish, sweat, ventilation, versatility, warmth, washability, waterproof, windproof.	*"It opens large enough to **accommodate** large cookware."* *"I gave it three stars only because it does work, albeit in a limited **capacity**."*
	Durability	abrasion, bombproof, bullet, cracking, damage, dent, durability, firm, flimsy, fragility, fray, indestructible, leakage, longevity, quality, rattling, rip, sturdiness, tear, withstand.	*"It's been almost 6 months now and its not **frayed** on the sheathe at all."*

Appendix C: User Appearance in Image and Text, and Product Durability in Picture

To identify the user appearing in the picture, we applied face detection algorithms to detect a person in the image. A drawback of this algorithm is that its performance drops when there are obstacles on the face (e.g., a pair of sunglasses or a mask). To improve the accuracy, we adopted the object classification algorithm as an additional check. This algorithm returns the class person if it detects (even part of) a user's body appearing in the picture. For the text mining task, we collected peer-recognizable user information such as gender, age, and body size. Since age and body are usually expressed in multiple words, we used word associations to handle this issue. For example, if a sentence has both "year/years/yr/yrs" and "old," we regarded it as having the age information.

Table C1 enumerates detailed objects of user appearance recognized in the machine learning algorithms. Table C2 presents all the expressions and keywords we used to extract the reviewer identity disclosure information from the review text. Examples of *User Appearing in Image* are shown in Figure C1.

Table C.1. Image-Mined Object–User Appearance

Object Classes	**Detected Object**
Face	male, female.
Person	athlete, alpinist, backpacker, child, children, climber, contestant, couple, crowd, motorcyclist, runner, skier, skin-diver, social gathering, tourist, traveler, yogi.

Table C.2. Text-Mined Keyword List–User Appearance

Aspects	**Keyword List**	**Example from Actual Reviews**
Age	Combination of year/years/yr/yrs and old in a sentence	*"I am 26 **years old** and have a tiny head so I have to wear kid's helmets."*
Gender [1]	male, female	*"I'm a **male** but buy women's shoes because of my narrow feet."*
Body Size [2]	(height) any numbers and ' in a sentence (weight) lb/lbs/lbs./pound/pounds and ' in a sentence (body size) waist/head/chest/shoe/shoes, any numbers, and size in a sentence	*(height) "I'm **6'4** with long arms and the large is perfect."* *(weight) "I'm 6'2", **190lbs**."* *(body shape) "I'm a **size 34** in **waist** and the size M fits perfectly."*
Notes: [1] We excluded the situations when the words do not refer to the gender of the person but rather represent product categories. For example, expressions such as *4-man tent"* and *"women's shoes"* were not considered gender disclosure. [2] We excluded the situations when weight information is used for product description. For example, "*I was using a 3.7 lb. zero deg bag*," "*Note the 75 - 225lb weight limit*" were not treated as body size information.		

Object Classified as a Person	Face Detection
	

Figure C.1. Examples of User Appearing in Image

We included an additional image measure, *Image: Product Malfunction*, to indicate a particular type of product durability in the image. This feature was manually measured by coders due to the complexity of using algorithms to detect such an issue in images. Among the 772, only 29 pictures in our data show a product malfunction (e.g., a cracked helmet, a broken bag string). Figure C2 shows examples of *Product Malfunction in Image*.

Figure C.2. Examples of Product Malfunction in Image