Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

When Images Backfire: The Effect of Customer-Generated Images on Product Rating Dynamics

Yue Guan, Yong Tan, Qiang Wei, Guoqing Chen

To cite this article:
Yue Guan, Yong Tan, Qiang Wei, Guoqing Chen (2023) When Images Backfire: The Effect of Customer-Generated Images on Product Rating Dynamics. Information Systems Research

Published online in Articles in Advance 28 Feb 2023

. https://doi.org/10.1287/isre.2023.1201

informs.
https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Articles in Advance, pp. 1–23
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

When Images Backfire: The Effect of Customer-Generated Images on Product Rating Dynamics

Yue Guan,[a] Yong Tan,[b] Qiang Wei,[c,*] Guoqing Chen[c]

[a] School of Economics and Management, Communication University of China, Beijing 100024, China; [b] Michael G. Foster School of Business, University of Washington, Seattle, Washington 98195; [c] School of Economics and Management, Tsinghua University, Beijing 100084, China

*Corresponding author

Contact: yueguan@cuc.edu.cn, https://orcid.org/0000-0002-9979-5901 (YG); ytan@uw.edu, https://orcid.org/0000-0001-8087-3423 (YT); qiangwei@mail.tsinghua.edu.cn, https://orcid.org/0000-0002-8397-7129 (QW); chengq@sem.tsinghua.edu.cn, https://orcid.org/0000-0001-7386-1133 (GC)

Received: July 2, 2020
Revised: October 10, 2021; September 23, 2022
Accepted: December 3, 2022
Published Online in Articles in Advance: February 28, 2023

https://doi.org/10.1287/isre.2023.1201

Abstract. Customer-generated images (CGIs) on e-commerce platforms have been widely adopted as a promotional tool to persuade customers into purchases. Despite their prevalent applications, the effect of CGIs on customer postpurchase satisfaction has not been extensively examined. This study postulates that CGIs could cause expectation disconfirmation and reduce product uncertainty for customers, therefore making their effect on subsequent product ratings complex. We leverage multiple methods and data sets to gain a better understanding of this problem and underlying mechanisms. We employ a difference-in-differences model to empirically test our hypotheses and find that CGIs lead to a decline in subsequent ratings compared with product ratings not exposed to CGIs. Further heterogeneity analyses demonstrate that high CGI review rating and high aesthetic quality exacerbate the negative effect, whereas reviewer face disclosure could alleviate the negative effect. Through cross-product analyses, we find that the negative effect is more prominent for experience goods (e.g., women's dresses) than for search goods (e.g., lightning cables). Finally, the underlying mechanism is further validated through a laboratory experiment that shows participants experience significantly higher expectation and more negative disconfirmation in the CGI group with high review ratings, whereas uncertainty reduction effect is insignificant, which collectively explains the decline of subsequent product ratings. These findings suggest that platforms and retailers should be aware of the potential negative effect of CGIs on the rating dynamics and take appropriate measures to circumvent it.

History: Ravi Bapna, Senior Editor; Gordon Burtch, Associate Editor.
Funding: This work was supported in part by National Science Foundation of China [Grants 72202220, 72172070, and 71729001].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2023.1201.

Keywords: customer-generated images • rating dynamics • uncertainty reduction • expectation disconfirmation • reviewer subjectivity

1. Introduction

Images generated by users, as a special form of user-generated content (UGC), have become a norm on social sharing and e-commerce platforms. According to Statista, 41% of surveyed users post images or videos on social media sites once a month.[1] The vast amount of multimedia content implies unique business opportunities for companies that could be leveraged as an important marketing tool to attract new customers and increase revenue. For instance, in order to promote its $6 sheet mask product, Sephora launched a campaign named #HolySheetMask Challenge to encourage users to post images with sheet masks on their faces, and it received great success. As is commonly seen, in addition to social media sites, many e-commerce platforms now encourage customers to post images in addition to textual reviews. These reviews with images are considered high-quality reviews by the platforms and have a higher probability to be placed at the top positions of review pages, aiming to quickly grasp potential buyers' attention and further increase product persuasiveness (Hong et al. 2020).

Online customers also have a great demand for visual information to help address the problem of product uncertainty, especially for experience goods (Hong and Pavlou 2014). U.S. digital customers expect an average of about six images and three videos when looking at a product on Amazon or another platform,[2] and between 65% and 85% of people describe themselves as "visual learners" (trend reports).[3] Product images generated by marketers cannot fully address the uncertainty problem because of the fixed presentation style and uniform visual aesthetics as well as customers' "immunity" and discounted trustworthiness on marketer-generated information (Goh et al. 2013). Images generated by customers are different from product images in many aspects. Despite their growing popularity in marketing campaigns and increasing interest paid by customers, the roles of customer-generated images (CGIs) in resolving product uncertainty and affecting satisfaction have not been thoroughly investigated.

In this paper, we are interested in images posted by customers on e-commerce platforms, which are termed CGIs. They usually appear in the review sections together with review text and ratings posted by customers who have previously purchased the products. CGIs are highly heterogeneous with respect to aesthetic quality and content as the reviewers who post CGIs are not professional photographers and have different preferences regarding when and what kind of content to show. In academia, extensive studies investigate the antecedents and consequences of product reviews (Huang et al. 2017), and reviews with CGIs have a positive impact on purchase intention and are perceived to be more useful (Hong et al. 2020, Zinko et al. 2020, Li et al. 2022). However, most of the review-related literature focuses on the textual content or numerical ratings, and the few studies on CGIs mainly discuss their implications on prepurchase outcomes (Liu and Du 2020, Zinko et al. 2020), neglecting their impact on subsequent customer postpurchase satisfaction.

CGIs could play an influential role in affecting customer decision making and postpurchase satisfaction with their unique characteristics. It is well-established that images have perceptual and persuasive advantages over text (Peracchio and Meyers-Levy 2005). They are generally more expressive and possess high attention-grabbing qualities (Childers and Houston 1984). CGIs integrate the properties of both images and user-generated content, making themselves more impressive than textual content generated by customers (Wang et al. 2016) and more trustworthy than seller-provided information (Goh et al. 2013). On one hand, CGIs, as an alternative information source, provide additional information beyond text, which reduce product uncertainty perceived by customers, and we name it the information effect. On the other hand, because people tend to give more trust to reviews with CGIs (Xu 2014, Zinko et al. 2020), reviews with CGIs have larger priority over textual reviews in affecting customers' prepurchase expectation. Furthermore, CGIs are posted by reviewers with subjective preferences (Shen et al. 2015). For instance, when reviewers are satisfied with themselves wearing a new dress, they may choose to post a CGI (Sung et al. 2016). These subjective judgements further influence the expectation of subsequent customers through CGIs. The high or low expectation brought by CGIs could further lead to negative or positive expectation disconfirmation[4] given the product quality is unchanged, and we name it the disconfirmation effect.

Considering the academic research gap and CGIs' dual effects on uncertainty reduction and expectation disconfirmation, we propose the following research questions. (1) What is the effect of CGIs on customers' postpurchase satisfaction as represented by subsequent product ratings? (2) How does CGI heterogeneity affect the ratings differently? And what are the potential mechanisms underlying these effects? In a quasi-experiment setting based on one of the largest e-commerce platforms, we empirically answer these research questions with a combination of econometric models and laboratory experiments. Estimation results indicate that CGIs have a negative effect and lower subsequent product ratings by 1.5% (0.06 stars) on average. This provides initial evidence that CGIs' disconfirmation effect outweighs their information effect and, therefore, decreases customers' postpurchase satisfaction. The heterogeneous effects of CGI aesthetic quality, reviewer information disclosure, CGI review rating are further examined. CGIs with high aesthetics have a negative effect on product ratings, and face disclosure attenuates the negative effect of CGIs by providing more information. Moreover, the negative effect of CGIs is 2.5 times larger in the high-rating group, whereas not significant in the low-rating group, validating that CGI reviewer subjectivity is a major factor in explaining the downside effects on rating dynamics. To further validate the underlying mechanisms, especially CGIs' role in affecting disconfirmation and product uncertainty, a laboratory experiment was conducted to dive into the attitudes of customers toward CGIs. We found evidence of expectation disconfirmation when the review has a high rating, whereas CGIs' effect on uncertainty reduction is significant only when the review has a low rating.

This study makes several contributions. First, CGIs as an important type of UGC, have received much attention by platforms and retailers. On some online platforms, CGIs are placed in salient positions to attract new customers, and new platforms featuring customer-generated multimedia content are also growing rapidly. Despite some studies discussing their positive impacts on purchase intention and trust (Hong et al. 2020, Zinko et al. 2020), most of them focus on prepurchase outcomes and new customer attraction, whereas postpurchase satisfaction and their potential downside effects are rarely discussed. With rigorous econometric analyses to exclude alternative explanations, CGIs' overall negative effect on product ratings is validated, enriching our understanding of the business and social value of CGIs. Second, the underlying mechanism of CGIs' negative effect is established with an additional laboratory experiment, which demonstrates the significance of the disconfirmation effect when the CGI has a high rating, whereas the information effect is limited. Expectation disconfirmation results from the fact that CGIs posted by those satisfied reviewers increase customers' prepurchase expectation and, consequently, cause unsatisfactory purchase experiences when the high expectation is unmet compared with product quality. As investigated by prior literature (Chen and Xie 2008), product ratings significantly impact the sales performance of online retailers. In our context, CGIs lead to a decrease of product rating by 1.5%, which is equivalent to a sales reduction of 0.63% (You et al. 2015). Third, several CGI-related factors are detected that could moderate this negative effect. CGI reviewer subjectivity is a leading factor deciding the upward or downward direction of customers' expectation, and aesthetic quality also plays a role in affecting customers' expectation level. From a content-based perspective, reviewers should be encouraged to post CGIs with less subjectivity and content that truly benefits potential buyers instead of eliciting new purchases. From the platform's perspective, better privacy protection policy could be considered as it could alleviate reviewers' privacy concerns and give reviewers more incentives to disclose personal information (such as faces) to the review system. Moreover, our research also provides discussion on CGIs' impact across multiple product types (clothes, electronics) to provide more enriched theoretical and managerial implications for academia and practice.

2. Theoretical Background

2.1. User-Generated Images on Digital Platforms

Extensive studies in prior literature investigate how the design and composition of images affect customer attitudes. Images are generally more impressive than text, possess high attention-grabbing qualities, and are remembered better (Childers and Houston 1984). Moreover, the presence of human-related information in website design is found to affect social presence and customer trust (Cyr et al. 2009).

The proliferation of deep learning provides tools and models to conduct large-scale image analytics (Howard et al. 2017). Based on e-commerce and social media platforms, some studies examine the correlation between images on the platform and outcome variables, including click behavior, purchase intention, customer conversion, and customer demand (Wang et al. 2016, So and Oh 2018, Zhang et al. 2021). The visual factors of interest in these studies include whether there is a model's face (Wang et al. 2021), image aesthetic quality (Zhang et al. 2021), color properties (So and Oh 2018), image composition and complexity (Wang et al. 2016), and image–text fit (Li and Xie 2020). For instance, So and Oh (2018) extract various image attributes from Facebook images, such as facial expression and stimulus content, and find heterogeneous effects on customer's search and purchase intention. Shin et al. (2020) measure the effect of different visual elements on a post's popularity on Tumblr. They find that proper visual stimuli, such as celebrities and images of high aesthetics, positively affect engagement, whereas complex image content has the opposite effect. Table 1 summarizes recent studies related to user-generated images on platforms, including the image source, visual attributes, and outcomes of interest. There are also algorithms proposed from a methodological perspective to exploit user-generated images. Liu et al. (2020) propose an algorithm to detect brand attributes from Instagram images, and the detected results are mostly consistent with official brand positions and survey results. Other studies include predicting image aesthetics (Talebi and Milanfar 2018), detecting sentiment (Truong and Lauw 2017), and predicting usefulness (Ma et al. 2018) from user-generated images.

Studies on images in product reviews find that reviews with images are perceived to be more useful (Li et al. 2022), and customers have higher purchase intention and trust when facing reviews with images (Xu 2014, Hong et al. 2020, Liu and Du 2020, Zinko et al. 2020). To the best of our knowledge, prior research efforts on image reviews mainly focus on prepurchase variables and rarely discuss the effect on postpurchase outcomes. Hence, this study aims to fill this gap and

Table 1. Related Literature on User-Generated Images

Literature	Image source	Visual attributes	Outcome variables
So and Oh (2018)	Images on Facebook	Facial expressions, objects, sensitive content, color properties	Click and conversion rate
Shin et al. (2020)	Posts on Tumblr	Celebrities, image–text similarity, sensitive content, complexity	Likes and reblogs
Li and Xie (2020)	Posts on Twitter and Instagram	Aesthetic quality, human face, image–text fit	Likes and retweets
Zhang et al. (2021)	Property images on Airbnb	Twelve visual properties including image composition, color, and figure–ground relationship	Property occupancy rate
Li et al. (2022)	Images in product reviews	Image sentiment	Review helpfulness
Zinko et al. (2020)	Images in product reviews	—	Trust and purchase intention

examine the effect of CGIs on subsequent product ratings on e-commerce platforms.

2.2. Factors in Rating Dynamics

Online product reviews are defined as "peer-generated product evaluations posted on company or third-party websites" (Mudambi and Schuff 2010, p. 186). Existing literature extensively discusses various factors that could affect product reviews and product rating dynamics.

Product rating as a reflection of customer satisfaction, is affected by retailer-provided cues, such as price, store name, country of origin, and product claim (Chen-Yu and Kincade 2001). Social influence is also considered to be an important element in rating dynamics (Muchnik et al. 2013). Herding and differentiation behavior are observed in different situations depending on the rating environment and reviewer characteristics (Lee et al. 2015, Sunder et al. 2019). For instance, prior research finds that initial positive ratings encourage subsequent negative ratings (Lin and Heng 2015). In a social network, reviewers are more likely to show bandwagon behavior and increase the positive emotions in reviews (Lee et al. 2015, Huang et al. 2017, Wang et al. 2018). Structure of the social ties is equally important as reciprocal ties have greater effect than follower and followee ties (Rishika and Ramaprasad 2019). Furthermore, review-posting behaviors are influenced by platform policy, such as free product sampling (Lin et al. 2019), incentivized reviews (Qiao et al. 2020), editorial reviews (Deng et al. 2021), and customers' trust toward the platform (Ho et al. 2017).

Most of the product ratings are observed to follow a decreasing trend (Godes and Silva 2012). In addition, there may be self-selection bias (Hu et al. 2017) as some reviewers may strategically choose specific rating strategies to compete for attention (Shen et al. 2015). Whereas extensive academic progress has been made regarding factors that could impact product ratings, no research has examined the impact of CGIs, that is, an increasingly important and prominent component on the platform, in this dynamic process.

2.3. Information and Disconfirmation Effects of CGIs

Before posting a rating for a specific product, a customer normally goes through two stages. First, the customer forms an expectation of the product and makes a purchase decision. The purchase decision indicates the customer's favorable attitude toward the product through information assimilation and processing before experiencing the product. Second, after receiving the product, the customer knows the actual product utility and chooses whether to post a product review indicating the customer's satisfaction level. Based on theoretical background in image-related research and attitude formation disciplines, we argue that CGIs have two opposing effects, termed the disconfirmation and information effects, on postpurchase satisfaction.

Expectation confirmation theory (ECT) (Oliver 1977) is a classic theory used to explain postpurchase satisfaction. ECT argues that customer postpurchase satisfaction depends on the actual product performance and whether the customer's prepurchase expectation is confirmed. Given the actual product performance, a disconfirmation of expectation leads to lower postpurchase satisfaction (Anderson 1973, Oliver 1980). After controlling factors, including time trend, social influence, and reviewer characteristics (Godes and Silva 2012, Hu et al. 2017, Sunder et al. 2019), each product rating reflects the customer's satisfaction with the purchase with a high rating indicating a high satisfaction level and a low rating indicating a low satisfaction level. The actual product performance is determined by the customer's experience with the product itself, whereas prepurchase expectation can be largely influenced by CGIs (Zinko et al. 2020).

2.3.1. Disconfirmation Effect. The specific characteristics of CGIs significantly affect customers' expectation formation and decision-making process. First and foremost, CGIs contain subjective evaluations of the reviewers, and these can be biased as a result of different personal preferences (Shen et al. 2015). In other words, CGIs do not appear randomly, and some reviewers may be more prone to post CGIs when they feel satisfied with the product, especially posting CGIs with their personal information (Sung et al. 2016). From the perspective of customers, they have more trust for reviews with CGIs and, therefore, give more weight to them in their purchase decision than to reviews without CGIs (Goh et al. 2013, Xu 2014, Zinko et al. 2020). However, as mentioned, the reviews with CGIs don't represent the average product quality because of the subjectivity of CGI reviewers and their selective CGI posting behavior. These biased judgments may further transmit inaccurate signals to potential buyers, decreasing the objectiveness of the information and affecting customers' decision quality. Second, CGIs are heterogeneous in aesthetic quality, and aesthetics could affect customers' perception toward the product (Hagtvedt and Patrick 2008). Generally, CGIs with high aesthetics could transmit information more effectively, but they also generate an appeal in potential customers, and furthermore, the appeal of images transforms into a high expectation toward the product (Zhang et al. 2021), which leads to deviation from a rational expectation. On the contrary, for CGIs with low aesthetics, customers' expectation toward the product could be much lower because of the unpleasant visual experience. To summarize, subjective evaluation of the CGI reviewers and aesthetic quality heterogeneity are the main CGI characteristics that affect customers' prepurchase expectation. If customers hold an unrealistic and high expectation before purchase, they are more likely to feel disappointed after experiencing the product and, therefore, give a low rating to the product.

2.3.2. Information Effect. In addition to the effect on expectation, a CGI is another form of information that complements verbal information. Verbal language excels at conveying the meaning of conditional events or causes, whereas visual language is more effective at transmitting emotions or attitudes (Lim et al. 2000). Therefore, CGIs provide customers with enriched information in addition to textual reviews. Moreover, different from marketer-generated images (MGIs), CGIs are taken from the perspective of customers and may contain some reviewer-related personal information. The additional information provided by CGIs helps potential buyers evaluate the product more thoroughly and reduce product uncertainty (Dimoka et al. 2012) faced in the online shopping environment, and the reduced uncertainty may further help customers make a more satisfying purchase decision than comparable products without CGIs (Luo et al. 2012, Hong and Pavlou 2014, Chen et al. 2021).

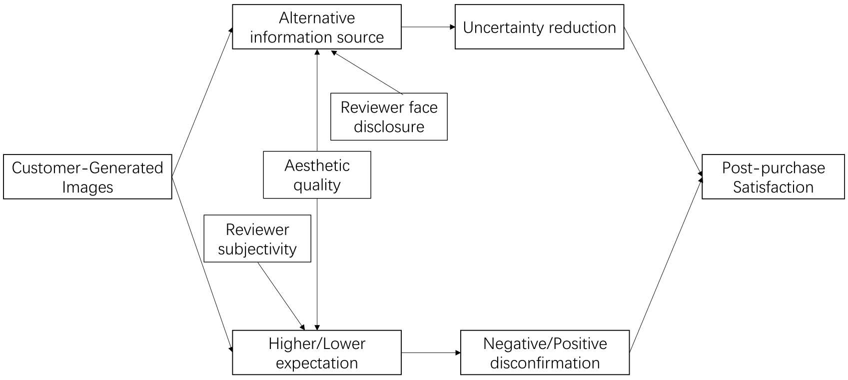
To summarize, as shown in Figure 1, on one hand, CGIs cause bias to customers' expectation-formation process and, therefore, lead to a higher or lower prepurchase expectation depending on the CGIs' characteristics. With a higher expectation, customer postpurchase satisfaction is negatively affected through negative disconfirmation. On the other hand, CGIs provide more information and reduce information uncertainty and, therefore, help customers make more satisfied purchase decisions (Hong and Pavlou 2014). The net effect on postpurchase satisfaction or subsequent product ratings depends on the relative strength of these two effects.

There are several factors in the discussion that can greatly influence the strength of these two effects, including aesthetic quality, reviewer-related information disclosure, and CGI reviewer subjectivity. To further uncover the mechanism of how CGIs affect customer postpurchase satisfaction, we fully exploit the heterogeneity of CGIs on these factors and discuss how these factors affect subsequent product ratings differently through the information and disconfirmation effects.

2.3.3. CGI Aesthetic Quality. Aesthetics deals with the nature of beauty and taste as well as the philosophy of art. Photography literature extensively discusses criteria of image aesthetics (Freeman 2007), such as composition, color, and figure–ground relationship. In recent years, image aesthetics assessment has been widely investigated in the research field of computer vision, which could automatically categorize images into different aesthetic levels (Deng et al. 2017). Especially, models based on convolutional neural networks (CNNs) (Talebi and Milanfar 2018) have made significant progress on image processing, gradually replacing feature engineering methods.

Despite CGIs' trustworthiness, they generally contain more imperfections compared with MGIs, which is typical of user-generated content. The aesthetic quality of a CGI affects the degree of information and disconfirmation effect. On one hand, images with high aesthetics generate more arousal and appeal in customers (Childers and Houston 1984, Jiang et al. 2016), which create an enlightening mood. The spillover effect from the image aesthetics to the product itself could lead to a high expectation toward the product (Hagtvedt and Patrick 2008, Zhang et al. 2021), despite that aesthetics may be independent of product quality, and therefore, is more likely to cause unsatisfactory purchase experiences. On the other hand, from an information-provision perspective, an image with high aesthetic quality implies more clarity, clear separation of foreground and background, appropriate brightness and contrast, and comfortable colors (Freeman 2007, Zhang et al. 2021). Because of these aspects, more information can be clearly transmitted to customers and further help customers make a more rational purchase decision (Kisielius and Sternthal 1984). In other words, the roles of CGI aesthetic quality

Figure 1. Conceptual Model of CGIs' Effect on Postpurchase Satisfaction



are twofold in that a CGI with high aesthetics generates a high expectation for potential customers as well as transmits information more effectively. Thus, the effect of CGI aesthetics on ratings could be positive or negative.

2.3.4. CGI Reviewer Information Disclosure. Some reviewers choose to reveal their personal information when posting CGIs. Reviews with reviewer personal information are less likely to be perceived as fake reviews, thus increasing the credibility and trustworthiness of the CGIs from the perspective of potential buyers (Cyr et al. 2009). For fashion products specifically, some reviewers disclose their faces in CGIs, such as a CGI with the reviewer wearing a dress at a party. People have a positive reaction to reviews that disclose identity-descriptive information (Forman et al. 2008). The disclosure of reviewer personal information could be an information source as reviewers are more comparable entities to potential customers than the models displayed in MGIs, which helps buyers form a more accurate judgment about potential fitness through mental imagination and reduce product fit uncertainty (Wheeler et al. 1997, Aydinoğlu and Cian 2014, Hong and Pavlou 2014). Therefore, CGIs with reviewer personal information disclosure have a more positive effect on product ratings than CGIs without.

2.3.5. CGI Reviewer Subjectivity. Each CGI is accompanied with a review rating indicating the satisfaction of the CGI reviewer with the product. Reviewers choose to post CGIs in specific circumstances and with specific preferences (Shen et al. 2015, Hu et al. 2017). A high CGI rating acts as a positive cue, whereas a low CGI rating acts as a negative cue. The positive cues convey a signal of high quality to customers who are not carefully considering the true merits of a product and, thus, are more likely to generate a high expectation in potential buyers (Lin and Heng 2015). However, as this favorable attitude is based on a specific review, it is temporary and less persistent, which, according to ECT (Oliver 1977), leads to dissatisfaction when the high-rating cue is inconsistent with the true quality of the product. With respect to the negative cue implied by a low CGI rating, it does not illicit a favorable attitude in the very beginning, and the expectation toward the product is not high. Therefore, CGIs with a high rating have a more negative effect on product ratings than CGIs with a low rating.

3. Research Context

3.1. Data Description

To empirically answer our research questions, we chose the Amazon platform as our main context. In 2021, this platform had net sales of 469.82 billion dollars and net income of 33.36 billion, of which the biggest revenue segment is its online stores. A tremendous number of reviews are generated for various products on this platform every day, providing a perfect empirical context to conduct our analyses. Specifically, we collected data generated on the Amazon platform until April 1, 2017, which included product- and review-related data of all the products in the category of women's casual dresses.[5] In Section 5.5, we further enrich the data set with other product categories. As typical experience goods, fashion goods, such as dresses, involve high uncertainty before purchase; therefore, customers are more likely to rely on user-generated content, especially visual information, to reduce the uncertainty, and platform managers also leverage this kind of content to conduct marketing campaigns and attract new customers (Somerfield et al. 2018). That is the key reason for us to choose this specific product category as our main research object. Product review information includes reviewer ID, review time, review title, review text, and CGIs if any. Figure 2 is a

Figure 2. (Color online) Screenshot of the First Review Page

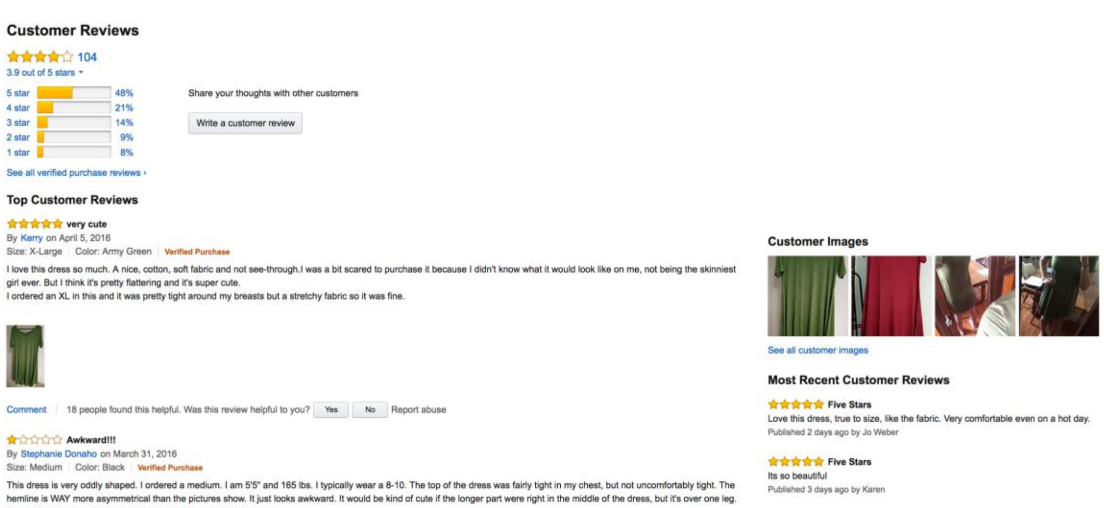


Table 2. Variable Description

Variables	Description
Product level	
NUM_REVIEWS	Total number of reviews of a product
PRICE	Product price, in dollars
QUALITY	Product quality, ranging from one to five
FIRST_AVAIL_MONTH	The month when a product appears on the platform. September 2009 is coded as one.
FIRST_REV_MONTH	The month when the first review of a product appears. September 2009 is coded as one.
TREAT_DUMMY	Whether a product belongs to the treatment group
Review level	
RATING	Ranging from one to five stars
TEXT_LEN	Number of words in the review text
TITLE_LEN	Number of words in the review title
MONTH	The month when a review appears. September 2009 is coded as one.
REVIEWER_EXPE	Whether the reviewer has posted a review before in this product category
CGI_DUMMY	Whether there is a CGI in a review
AESTHETIC QUALITY	Aesthetic quality score of a CGI
FACE DISCLOSURE	Whether there's a face in the CGI

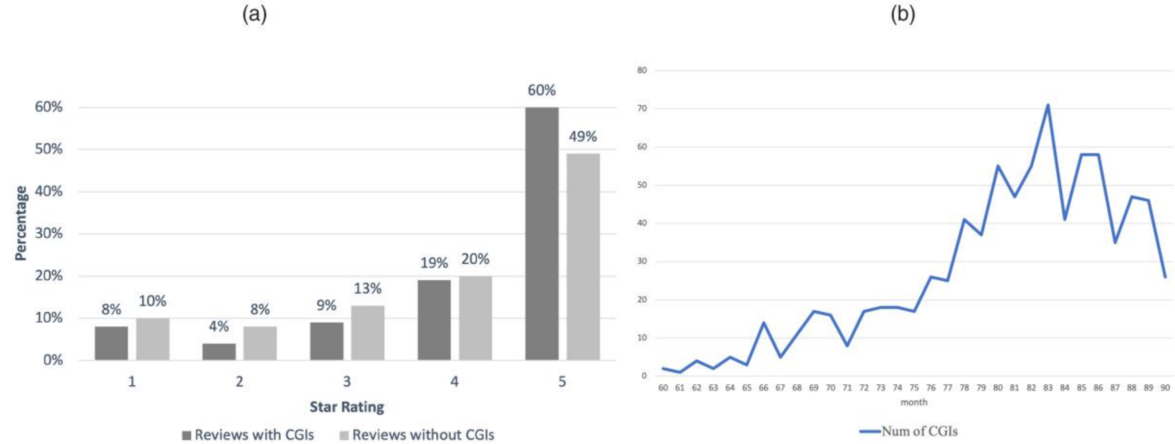
screenshot of the product reviews displayed on the product information page, which is also the first review page. Apart from being embedded in the reviews, CGIs are also placed on the right side as a separate section, further highlighting the critical attention paid to CGIs by platform managers.

The number of reviews across different products demonstrates a long-tail distribution. Consistent with prior literature (Li and Hitt 2008) and considering the representativeness of our data sample, we deleted products with more than 1,000 or fewer than 5 reviews with the following reasons. For products with fewer than 5 reviews, CGIs are less likely to appear, and with such a short review sequence, it is hard to claim the detected effect of CGIs on subsequent product ratings is robust and consistent. For products with more than 1,000 reviews, they become the most popular products and are not representative of an average product. For a specific product, there could be more than one CGI appearing in different reviews at different time points. Table 2 provides the variable description. The descriptive statistics of product- and review-level variables before and after propensity score matching are shown in Table 3. On average, each product has 49 reviews with an average price of 28.97 U.S. dollars. On the review level, review text has an average length of 32 words, and review titles have an average length of 4 words. The average rating of all the reviews is 3.93 stars with a J-shaped distribution. Only 12.6% of the reviews are posted by reviewers with prior review experiences: 41.9% of the products have at least one CGI, and reviews with CGIs account for 4% of the total number of reviews.

Table 3. Descriptive Statistics Before and After Matching

	Observations		Mean		Standard deviation		Minimum		Maximum	
Variables	Before	After	Before	After	Before	After	Before	After	Before	After
Product level										
NUM_REVIEWS	2,802	1,650	49.25	48.09	93.693	86.03	5	5	997	997
PRICE	2,802	1,650	28.965	25.807	23.292	18.919	2.48	2.48	306.99	175.51
QUALITY	2,802	1,650	3.816	3.816	0.593	0.572	1.1	1.3	5	5
FIRST_AVAIL_MONTH	2,802	1,650	70.457	70.623	13.324	13.035	1	3	89	89
FIRST_REV_MONTH	2,802	1,650	73.942	73.711	11.377	11.496	11	11	90	90
TREAT_DUMMY	2,802	1,650	0.419	0.5	0.494	0.5	0	0	1	1
Review level										
RATING	190,145	79,396	3.926	3.912	1.341	1.348	1	1	5	5
TEXT_LEN	190,145	79,396	31.828	31.472	37.703	37.379	1	1	1,377	610
TITLE_LEN	190,145	79,396	4.453	4.468	3.913	3.928	1	1	31	31
MONTH	190,145	79,396	78.617	79.089	9.425	8.921	1	11	90	90
REVIEWER_EXPE	190,145	79,396	0.126	0.131	0.332	0.338	0	0	1	1
CGI_DUMMY	190,145	79,396	0.038	0.037	0.191	0.189	0	0	1	1
AESTHETIC QUALITY	8,400	825	3.525	3.487	0.378	0.435	1.561	1.715	4.871	4.728
FACE DISCLOSURE	8,400	825	0.461	0.424	0.498	0.495	0	0	1	1

Figure 3. (Color online) Distributions of Ratings and New Treated Products



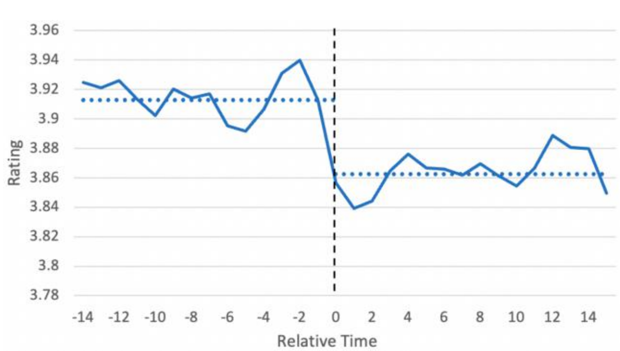
Notes. (a) Rating distribution for reviews with and without CGIs. (b) Number of new treated products in each month. The first CGI in panel (b) appears in month 60.

Figure 3(a) shows the rating distributions of reviews with and without CGIs. Given that the average rating is around four stars (Table 3), high ratings (five stars) account for 60% in the CGI group, and low ratings (three stars or lower) only account for 21%. For the review group without CGIs, high ratings (five stars) account for 49%, much lower than the CGI group, and low ratings account for 31%. In other words, CGIs posted by satisfied reviewers are almost three times as many as CGIs posted by unsatisfied reviewers. This could lead to more negative disconfirmation (expectation higher than product quality) than positive disconfirmation (expectation lower than product quality) for subsequent customers as discussed in Section 2.3. In Section 5.3, we provide more empirical evidence on the disconfirmation effect brought by the high ratings of CGIs. Figure 3(b) shows the number of new treated products in each month with an overall bell shape. (New treated products are defined as products that have a CGI generated for the first time).

3.2. Trend Analysis

Figure 4 displays the average rating trend of the treatment group (products with at least one CGI). The x-axis represents the relative order of a review for each treated product, and the moment a product receives a CGI is denoted as time 0. The y-axis represents average ratings across all the treated products at each time point.[6] The horizontal trendlines before and after time 0 are also presented. It can be observed that there is a significant drop ($mean_{treat\_before} = 3.913$, $mean_{treat\_after} = 3.863$, diff = 0.050, p-value < 0.001) in product ratings for the treated products before and after treatment. This visualization provides us with some initial evidence regarding the effect of CGIs on subsequent product ratings.

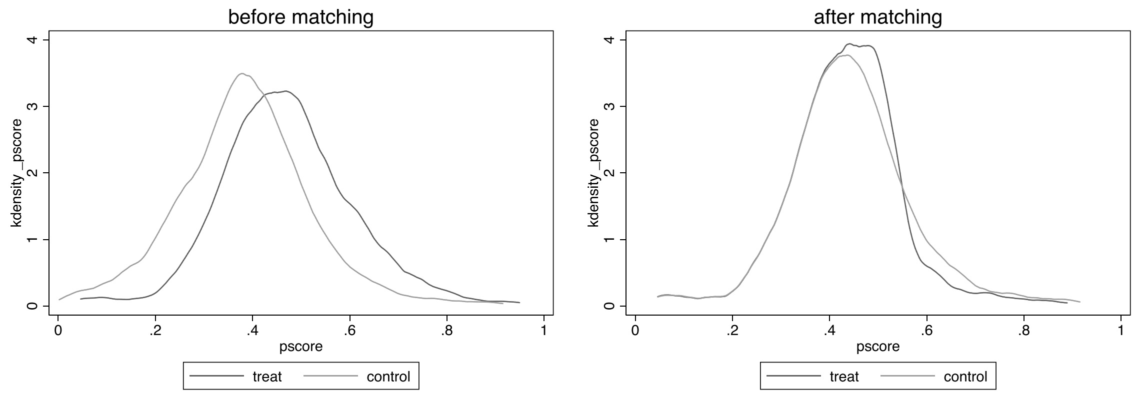
Figure 4. (Color online) Rating Trend of the Treatment Group



4. Model and Results

A difference-in-differences (DID) model is employed to investigate the influence of CGIs on subsequent product ratings. Widely applied in policy evaluations in the field of social science, DID evaluates the effect on outcomes under circumstances with or without policy intervention based on a counterfactual framework. In our context, a product is "treated" if there is a CGI appearing in the product reviews. The treatment group consists of treated products, and the control group consists of untreated products. Our panel data allows us to observe the rating dynamics of the treatment and control groups before and after the treatment. Therefore, by referring to previous literature (Qiao et al. 2020, Deng et al. 2021, Zhang et al. 2021), a DID approach is appropriate in our research context. One thing noteworthy is that DID assumes that the treatment is randomly assigned, which may not apply in our research setting. That is, there may be some variables that affect the appearance of CGIs as well as product ratings, which violates the randomness assumption. To address this concern, a propensity score matching (PSM) (Rosenbaum and Rubin 1983) is conducted before the DID analyses.

Figure 5. Propensity Score Distributions Before and After Matching



4.1. PSM

Propensity score is defined as the probability of a product receiving the treatment. It is usually estimated with a logit model, $pscore_i = 1/(1 + e^{-X_i\boldsymbol{\alpha}})$. $\boldsymbol{X_i}$ represents observable covariates of product i that could potentially affect the treatment assignments. Coefficients $\boldsymbol{\alpha}$ are obtained by maximizing the likelihood of the treatment assignments in the observed data.

PSM aims to ensure that the two groups are comparable or have similar trends before treatment. Specifically, the number of reviews and average rating in the first month are obtained for all the products (denoted as *PRE_REV_COUNT* and *PRE_RATING*), which represents the trend before treatment as the treatment group is not treated at this time (products that are treated in the first month only account for a small percentage and are not included in the data set). The number of reviews and prior ratings are relevant because the probability of posting a CGI may be dependent on product popularity measured by the number of accumulated reviews and product quality measured by product ratings (Lin and Heng 2015, Shen et al. 2015). Additionally, the following covariates for PSM: *PRICE*, *FIRST_AVAIL_MONTH*, *FIRST_REV_MONTH*, are added to the logit model. The descriptions of these covariates are shown in Table 2. It is expected that whether a product receives treatment is related to these product characteristics. For instance, if a product is so expensive that few people could afford it, the probability of a CGI being posted is low. Similarly, product available time and time of the first review are important factors as the longer time a product appears on the platform, the more likely a CGI is to be posted.

The result from the logit regression generates a propensity score for each product, and then, a one-to-one nearest neighbor without replacement matching algorithm is performed to ensure that the products in the control and treatment groups have similar probabilities to be treated. After matching, we have 825 products in the control group and 825 products in the treatment group. Figure 5 shows the distributions of propensity scores before and after matching. It can be observed that the distribution of propensity score in the treatment group is identical to that in the control group after matching.

To further check the validity of PSM, we conducted a balance check on the product covariates. Table 4 shows the differences and significance levels of variables between two groups before and after matching. The comparison results before matching indicate that systematic differences exist between the two groups before treatment. Review rating and the number of reviews are significantly higher in the treatment group than those in the control group, whereas average product price is significantly lower. Moreover, products in the treatment group appear earlier and receive the first review earlier.

Table 4. Balance Check Results Before and After Propensity Score Matching

	Before matching				After matching			
	Treatment	Control	t	p-value	Treatment	Control	t	p-value
PRICE	24.491	32.186	−8.07	0.000	25.679	25.794	−0.12	0.902
PRE_RATING	4.101	4.004	2.30	0.022	4.088	4.092	−0.08	0.935
PRE_REV_COUNT	3.714	2.665	7.77	0.000	2.965	3.160	−1.33	0.185
FIRST_AVAIL_MONTH	69.872	70.854	−1.78	0.076	70.988	70.636	0.56	0.578
FIRST_REV_MONTH	72.726	74.799	−4.41	0.000	74.113	73.619	0.88	0.380

Table 5. Difference-in-Differences Model: The Effect of CGIs

Variables	(1) Control		(2) Full		(3) Matched	
AfterTreat			−0.0566***	(0.0137)	−0.0602***	(0.0207)
Volume	−0.0002	(0.0001)	−0.0000	(0.0001)	−0.0002	(0.0002)
Valence	−0.1266***	(0.0091)	−0.1128***	(0.0062)	−0.1266***	(0.0091)
Variance	0.0524***	(0.0139)	0.0534***	(0.0100)	0.0580***	(0.0140)
TextLen	0.0009	(0.0006)	0.0011***	(0.0003)	0.0009	(0.0006)
TitleLen	0.0096	(0.0062)	0.0092**	(0.0043)	0.0096	(0.0062)
ReviewerExpe	0.1504***	(0.0150)	0.1225***	(0.0098)	0.1505***	(0.0150)
Observations	79,396		190,145		79,396	
Product fixed effects	Y		Y		Y	
Month fixed effects	Y		Y		Y	
Adjusted R^2	0.126		0.120		0.126	

Note. Robust standard errors in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

After PSM, there are no significant differences at the 10% level on all of the matching variables.

4.2. Difference-in-Differences Model

Our DID estimation model is shown in Equation (1), which is performed at the product–reviewer–time level:

$$\begin{aligned} Rating_{ijt} = {} & \beta_1 \cdot AfterTreat_{it} + \beta_2 \cdot Volume_{it} + \beta_3 \cdot Valence_{it} \\ & + \beta_4 \cdot Variance_{it} + \beta_5 \cdot TextLen_{it} \\ & + \beta_6 \cdot TitleLen_{it} + \beta_7 \cdot ReviewerExpe_{jt} \\ & + T_t + \alpha_i + \epsilon_{ijt}. \end{aligned} \tag{1}$$

Let subscript i denote product, subscript j denote reviewer, and subscript t denote time. $Treat_i = 1(0)$ represents that product i is in the treatment (control) group. $After_{it}$ equals one if period t is after the time when product i receives the first CGI. The key DID estimator in the model is $AfterTreat_{it}$, an interaction term of $Treat_i$ and $After_{it}$, which measures the net impact of CGIs on subsequent ratings. The dependent variable, $Rating_{ijt}$, indicates the rating given to product i by reviewer j at time t. Guided by prior literature (Godes and Silva 2012, Muchnik et al. 2013), we add control variables, including *Volume* (the number of previous reviews for product i before period t), *Valence* (average rating of previous reviews), *Variance* (variance of previous reviews), *TextLen* (average length of previous review texts), *TitleLen* (average length of previous review titles), and *ReviewerExpe* (whether the reviewer has previously posted a review).[7] Product fixed effects α_i is included to capture other product-specific features that may affect the outcome variable. Time fixed effects T_t is included to control seasonality patterns in product rating dynamics. As pointed out by Bertrand et al. (2004), the independent variables and outcomes may have serial correlation issue because of the nature of the DID model. Therefore, we allow for an arbitrary variance–covariance error structure within products over time by using clustered standard error in the estimation results.

4.3. The Effects of CGIs

We first estimate the model with all control variables, and the estimation results are shown in Table 5, column (1). The results show that the number of previous reviews has no significant effect on ratings, the reason for which could be that time fixed effects are controlled in the model. Evidence of social influence is further confirmed in that previous review valence has a negative effect on subsequent ratings, which is consistent with prior literature that reviewers are less motivated to post additional positive reviews for already highly rated products and reviewers demonstrate differentiation behaviors when posting reviews (Moe and Trusov 2011, Lin and Heng 2015). Moreover, previous review variance has a positive influence on ratings. One of the explanations is that the divergent reviews prevent customers from being influenced by potential bias in the reviews and assist customers in finding products that match their preferences (Godes and Silva 2012).

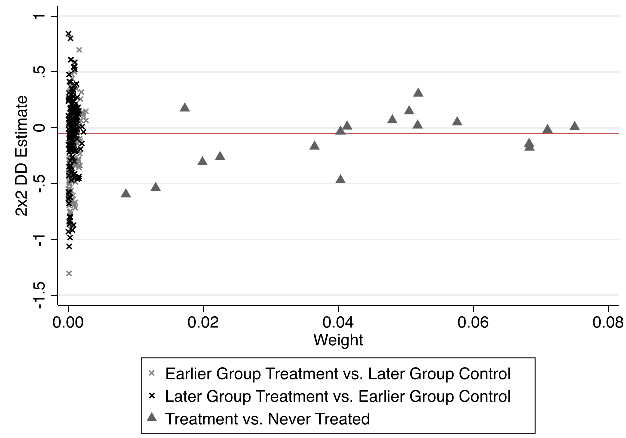
Adding the key DID estimator on top of the control variables, we then estimate a full model and summarize the results in column (2). This result is based on the whole data sample without propensity score matching. The coefficient of *AfterTreat* is significant and negative with a value of −0.0566. Next, we estimate the model with matched data set as the results shown in column (3). The coefficient of the key variable, $AfterTreat_{it}$, remains significant and negative, implying that CGIs lead to a decrease of product ratings by 0.0602 out of 5 stars, which is equivalent to a 1.5% decrease given that the average product rating is 4 stars. Prior research shows that the elasticity of review ratings on sales is 0.417, meaning that a 1% decrease of ratings could decrease sales by 0.417% (You et al. 2015). In our context, a decrease of 1.5% in product ratings means a sales reduction of 0.63%, which is a paramount number for online retailers and should not be underestimated. Referring to our theoretical argument, the negative effect indicates that the CGI disconfirmation effect plays a larger role than its information effect for subsequent customers.

4.4. Additional Analyses

4.4.1. Alternative Tests to DID Fixed Effects Estimation. The DID design in our research context is not the classic DID in that the treatment group receives treatment at different times, which forms a staggered DID design. One concern raised by prior literature (de Chaisemartin and D'Haultfoeuille 2020) argues that the fixed effects estimation is a weighted sum of the average treatment effect (ATE) in all treated cells, and a negative weight could be generated when the ATEs are heterogeneous across groups or periods. To mitigate the concern that the negative effect in our model is because of the negative weights, we conduct an analysis using the new estimator, DID_M, based on pairwise outcome comparisons of the treatment and control groups (de Chaisemartin and D'Haultfoeuille 2020). A corresponding placebo test is also conducted to confirm the validity of the DID model. The estimated average treatment effect is −0.071, remaining significant and negative. The placebo test results show that there are no significant differences between the treatment and control groups before treatment.

Furthermore, because the fixed effects estimation in our main model is a weighted average of all the possible 2×2 DID estimators, the weights depend on group sizes and the variance of the treatment dummy with each pair. We give an explicit interpretation of the weights and estimates according to the decomposition theorem proposed by Goodman-Bacon (2021) by plotting the 2×2 DID estimates with their corresponding weights as shown in Figure 6. The DID estimators can be grouped into three categories: (1) treated units as treatment versus units never treated as control units. It can be observed that this category receives the largest weight in our model as high as 0.782 (Table 6). (2) Then, there are units treated at different times: units treated earlier as treatment versus units treated later before its treatment begins as control (weight = 0.120). (3) Next, there are units treated at different times: units treated later as treatment versus units treated earlier after its treatment begins as control (weight = 0.098). The second and third categories receive a much smaller weight compared with the first category, which further indicates that the negative weight issue (de Chaisemartin and D'Haultfoeuille 2020) is not a big concern as negative weights only happen in the third category. Table 6 shows the estimated coefficient in each category, and the overall DID estimate combining the three categories' results is −0.052, consistent with our previous estimate.

Figure 6. (Color online) Distribution of DID Estimators with Corresponding Weights



4.4.2. Relative Time Model. One of the key assumptions of the DID model is the parallel trend assumption. That is, the product ratings in the treatment group should have a similar trend to the control group before being treated. Following extant literature (Deng et al. 2021, Zhang et al. 2021), a relative time model is employed to test the parallel trend assumption. In addition to the DID estimator, a series of time dummy variables ($Pre_{it}(j)$, $Post_{it}(k)$) are added, indicating the relative chronological distance between observation time and the time a CGI is posted. The specification of the model is shown in Equation (2):

$$Rating_{ijt} = \sum_j \gamma_j (Pre_{it}(j) \cdot Treat_i) + \sum_k \lambda_k (Post_{it}(k) \cdot Treat_i) + Controls_{it} + T_t + \alpha_i + \epsilon_{ijt}. \quad (2)$$

The newly added variable $Pre_{it}(j)$ is an indicator variable that equals one if month t is j months prior to the treatment. For instance, $Pre_{it}(1) = 1$ means that the review is generated one month before the treatment, and $Pre_{it}(5) = 1$ means five months before the treatment. Similarly, $Post_{it}(k)$ is an indicator variable that equals one if one review is posted k months after the treatment. $Post_{it}(1) = 1$ represents that the review is generated one month after the treatment happens, and $Post_{it}(5) = 1$ means five months after the treatment. Consistent with prior literature (Deng et al. 2021), we put all the pretreatment periods that are greater than or equal to six months prior to treatment into one dummy,

Table 6. DID Estimates by Goodman-Bacon (2021)

DID comparison	Weight	Average DID estimate
Treatment versus never treated	0.782	−0.050
Earlier treatment versus later control	0.120	−0.089
Later treatment versus earlier control	0.098	−0.022

Table 7. Robustness Check Results

Variables	(1) *RelativeTime*	(2) *Unobservables*	(3) *LagEffect*	(4) *OrderedLogit*	(5) *PSW*	(6) *LinearTrend*	(7) *NonlinearTrend*
Pre1	(Omitted)						
Pre2	−0.0442 (0.0418)						
Pre3	−0.0118 (0.0414)						
Pre4	−0.0061 (0.0457)						
Pre5	−0.0543 (0.0457)						
Pre6	0.0138 (0.0471)						
Post1	−0.1134** (0.0489)						
Post2	−0.0755 (0.0470)						
Post3	−0.0581 (0.0530)						
Post4	−0.1174** (0.0595)						
Post5	0.0051 (0.0674)						
Post6	−0.0344 (0.0592)						
AfterTreat		−0.0536***		−0.0979***	−0.0369**	−0.0789***	−0.0753***
		(0.0163)		(0.0248)	(0.0176)	(0.0247)	(0.0235)
Lag5_Treat			−0.0596***				
			(0.0201)				
Observations	47,610	119,761	79,396	79,396	126,619	79,396	79,396
Controls	Y	Y	Y	Y	Y	Y	Y
Product fixed effects	Y	Y	Y	Y	Y	Y	Y
Month fixed effects	Y	Y	Y	Y	Y	Y	Y
Product-specific trend	N	N	N	N	N	Linear	Linear + Nonlinear
Adjusted R^2	0.128	0.109	0.128		0.120	0.137	0.141

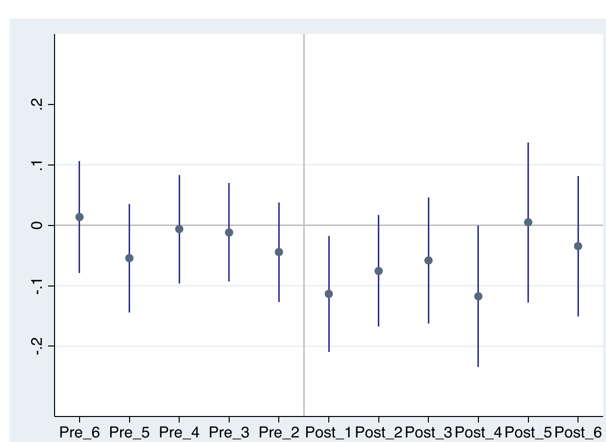
Note. Robust standard errors in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

$Pre_{it}(6)$, and all the posttreatment periods that are greater than or equal to six months after treatment into one dummy, $Post_{it}(6)$. Therefore, $\gamma_j (j = 1, 2, 3 \ldots 6)$ measures the pretreatment trend of the effect of CGIs, and $\lambda_k (k = 1, 2, 3 \ldots 6)$ measures the posttreatment trend of CGIs' impact on subsequent ratings. To prevent collinearity, the coefficient of $Pre_{it}(1)$ is normalized to zero.[8] $Controls_{it}$ represents all the control variables that are discussed in Equation (1) for abbreviation.

The estimation results are shown in column (1) of Table 7. As expected, the coefficients of the pretreatment variables are not significantly different from zero, validating the parallel trend assumption. After treatment, the negative effect is most significant in the first several months, and the coefficient demonstrates an overall increasing trend with the negative effect becoming smaller and not significant after four months. Figure 7 plots the dynamic effect of CGIs before and after treatment with 95% confidence intervals.

4.4.3. Robustness Checks. In addition to the covariates that we consider in the matching process, there may be some unobserved variables that determine whether a product is treated (Zhang et al. 2021). Therefore, inspired by Goodman-Bacon (2021), we further exploit the heterogeneous timing of treatment and investigate the effect of CGIs among all the treated products to better address the selection problem caused by unobservables. Specifically, there are some products that get treated in the first month (receives the first and only CGI in the first month) and remain treated later, and other products get treated in later months. The early treated product group remain treated throughout the data observation period except the first month and, therefore, can serve as the control group, and the later treated products serve as the treatment group. In this circumstance, it is possible to investigate the effect of CGI appearance on subsequent ratings after neglecting the reviews in the first month for all the treated products. The estimated coefficient based on this approach is −0.054 and consistent with previous results as shown in column (2) of Table 7.

Figure 7. (Color online) Dynamic Effect of CGIs in Pretreatment and Posttreatment Periods



Moreover, because it would normally take several days for an online purchase to be shipped to the customer, a review rating appearing after a CGI may not necessarily suggest the purchase decision being influenced by the CGI because the purchase may happen even before the CGI is posted. Therefore, allowing the time lag to be five days, a review is regarded as being treated only if the interval between the review posting date and the CGI posting date is longer than the specified time interval. Results in column (3) of Table 7 remain robust and negative (the results of the seven- and nine-day lags are similar).

Furthermore, as our outcome variable ranges in a five-star scale and only takes integer values, we further consider an ordered logit model (column (4) of Table 7) in which we assume reviewers have a latent evaluation for the products and there is a mapping between the latent evaluations and the star ratings. The marginal effect is calculated based on the estimated coefficient. The appearance of CGIs could, on average, increase the likelihood of one, two, three, and four stars by 0.81%, 0.49%, 0.56%, and 0.32%, and decrease the likelihood of a five-star rating by 2.19%, which is equivalent to an overall decrease of product ratings by 0.062 stars. This is also consistent with previous estimated results. In column (5) of Table 7, we use the propensity score weighting method to make sure the treatment and control groups are balanced. The propensity score function estimates the propensity of receiving treatment for each review, and a weight is generated according to the propensity score. Then, a weighted DID estimation is conducted, and the result shows the consistency despite a smaller coefficient. We further consider the possibility that each product has a specific time trend, which could be linear or nonlinear. This could control for any product-specific rating trends that might be driving the results (Wolfers 2006). The model specification is in Equation (3), and λ_i and δ_i capture the effects of linear and nonlinear time trends of product i. The result remains negative and significant as shown in columns (6) and (7) of Table 7.

$$Rating_{ijt} = \beta_1 \cdot AfterTreat_{it} + Controls_{it} + \lambda_i t + \delta_i t^2 + \alpha_i + T_t + \epsilon_{ijt}. \quad (3)$$

Reviews on the platform are displayed in an order calculated based on the platform's ranking algorithm, and the relative positions of the reviews could affect customers' probability of reading them. Therefore, we further estimate the platform's review ranking algorithm and estimate the effect of CGIs that appear on the first review page. The algorithm estimation and estimated results are all shown in Online Appendix C. We also conducted other robustness checks, including alternative matching methods, log transformation of variables, and alternative sample regression, and the results are presented in Online Tables C3 and C4.

5. Heterogeneity Analyses

Previous results show that CGIs lead to lower subsequent product ratings. In this section, we further explore potential factors that may affect the relative strength of CGIs' information effect and disconfirmation effect. Specifically, we choose two prominent visual factors of a CGI, namely, CGI aesthetics and reviewer face disclosure, and the main factor leading to expectation disconfirmation, CGI review rating, to investigate whether the effect of CGIs is dependent on these factors. In the review sequence of a product, there could be multiple CGIs appearing at different time points, making the heterogeneous effect of CGIs complicated. To disentangle the effect of CGI properties for those products with more than one CGI, we only keep reviews generated before the second CGI appears. In this case, the identified effect of CGI heterogeneity comes solely from the first and only CGI. Column (1) of Table 8 shows a

Table 8. Heterogeneous Effects of CGI Aesthetic Quality and Face Disclosure

Variables	(1) *Base*		(2) *LowAes*		(3) *HighAes*		(4) *Face*	
AfterTreat	−0.0730***	(0.0242)	−0.0496	(0.0348)	−0.1000***	(0.0322)	−0.1024***	(0.0256)
AfterTreat × *Face*							0.0685*	(0.0360)
Volume	−0.0001	(0.0003)	−0.0001	(0.0003)	−0.0004	(0.0005)	−0.0001	(0.0002)
Valence	−0.1457***	(0.0090)	−0.1438***	(0.0122)	−0.1521***	(0.0133)	−0.1459***	(0.0092)
Variance	0.0915***	(0.0144)	0.0685***	(0.0189)	0.1275***	(0.0224)	0.0921***	(0.0121)
TextLen	0.0014**	(0.0006)	0.0005	(0.0009)	0.0027***	(0.0009)	0.0014**	(0.0006)
TitleLen	0.0150**	(0.0065)	0.0180**	(0.0087)	0.0115	(0.0097)	0.0151**	(0.0059)
ReviewerExpe	0.1432***	(0.0191)	0.1622***	(0.0245)	0.1173***	(0.0300)	0.1432***	(0.0178)
Observations	47,610		27,614		19,996		47,610	
Product fixed effects	Y		Y		Y		Y	
Month fixed effects	Y		Y		Y		Y	
Adjusted R^2	0.148		0.158		0.131		0.149	

Note. Robust standard errors in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

consistent and negative overall effect of CGI (−0.0730) when only considering ratings before the second CGI. In Section 5.1, we first describe how to obtain the aesthetic quality and detect reviewer faces from a CGI. Then, we discuss the heterogeneous effect of CGIs and results on other product categories.

5.1. CGI Aesthetics Assessment

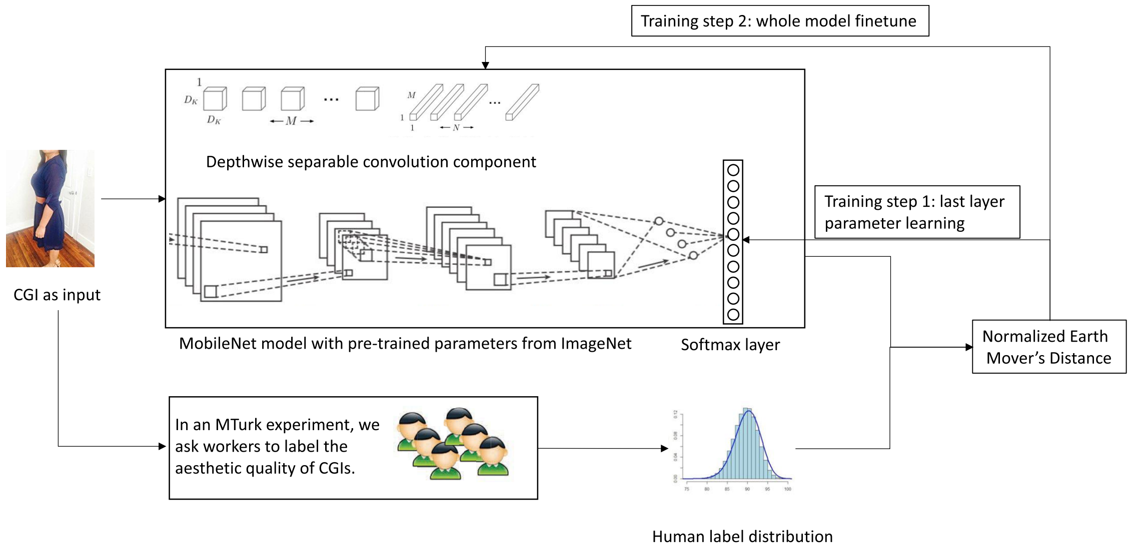
With recent progress in computer vision, deep learning–based models are proposed aiming to predict the aesthetic quality of images (Tian et al. 2015, Talebi and Milanfar 2018). Despite their remarkable performance, these models are trained and evaluated on professional photography data sets and may not apply well to CGI aesthetics assessment in our context as the CGIs are taken by amateur reviewers and the evaluation criteria in a shopping context are different from those in a photography contest. Therefore, building on research progress in computer vision, we develop an image aesthetics prediction framework, named CGI-MobileNet, as depicted in Figure 8.

In CGI-MobileNet, we adopt a deep learning framework for our aesthetics assessment context, which is based on an efficient and light-weighted CNN architecture: MobileNet (Howard et al. 2017). The structure details of CGI-MobileNet are provided in Online Appendix A. To obtain the labels for the training set, we deploy the image-labeling task on Amazon Mechanical Turk and ask workers to rate each CGI's aesthetics with an integrated evaluation of image clarity, brightness, foreground–background relationship, contrast, composition, etc. Each rating ranges from one to five stars and is evaluated by 10 master workers. Eighty percent of the labels are randomly selected into the training set, and 20% into the test set. Detailed instructions and examples are provided to further ensure the quality of the evaluation results (see Online Appendix A).

The 10 human-annotated labels constitute an aesthetic quality distribution for each CGI. Our proposed prediction model's last layer generates a five-dimension vector, representing the predicted distribution, and the objective is to minimize the distance between the true and predicted distribution. Earth mover's distance is chosen as the loss function (Talebi and Milanfar 2018), which measures the distance of moving one cumulative distributions to another cumulative distribution, $EMD_i = \left(\frac{1}{N}\sum_{k=1}^{N}(CDF_{GT}(k) - CDF_{PR}(k))^2\right)^{1/2}$, where N represents the number of classes (five in our case) and CDF_{GT} and CDF_{PR} denote the cumulative distributions of the ground truth and predicted labels.

Considering the small scale of our training set, instead of training the whole network from scratch, a fine-tuned strategy is employed to avoid the overfitting problem. Specifically, the weights are initialized using a pretrained model on the large-scale ImageNet data set, and in the training process, we first train the last Softmax layer with a large learning rate to initialize the parameters of the last layer and keep the other parameters fixed. Then, in the second stage, we fine-tune the whole CNN with a lower learning rate and save the model for later evaluation and prediction. In this process, we simultaneously utilize the capabilities of the deep neural network to extract visual features and human intelligence to provide supervision for the aesthetics prediction task. We compare our model with baseline models and demonstrate the superior prediction performance compared with previous image

Figure 8. (Color online) Aesthetic Quality Assessment Framework (CGI-MobileNet)



assessment models (Online Appendix A provides the results of performance comparison and illustrative examples).

For the task of face detection, a state-of-the-art face detection algorithm, multitask cascaded convolutional networks (MTCNN) (Zhang et al. 2016), is utilized to detect faces from CGIs. MTCNN is a deep convolutional neural network comprising three subnetworks. First, candidate windows are generated through the fast proposal network (P-Net). Second, the candidates are refined through a refinement network (R-Net). Third, the final bounding box positions are generated through the output network (O-Net). Validation accuracy of MTCNN can achieve 95.4% (Zhang et al. 2016). In the whole data set, faces are detected in 40% of the CGIs. Next, we discuss the effect of CGI aesthetic quality and reviewer face disclosure.

5.2. CGI Aesthetic Quality and Reviewer Information Disclosure

We expect that higher aesthetics both increase customers' prepurchase expectation and reduce information uncertainty; therefore, the net effect can be uncertain. To empirically examine this question, we divide all of the treated products into two groups: high and low aesthetics groups based on the aesthetic quality scores of CGIs obtained in Section 5.1, and we investigate the effects separately for each group. From the results in columns (2) and (3) in Table 8, the effect of CGI is significant at the 1% level and negative only in the high-aesthetics group, indicating that the disconfirmation effect is larger than information effect, and therefore, CGIs with high aesthetics hurt subsequent customer's postpurchase satisfaction because of the high expectation level elicited before purchase. In the low-aesthetics group, there's no significant effect. Because the aesthetic scores come from the predictions of our proposed CGI-MobileNet framework, they may contain measurement errors that may affect the accuracy of the estimated coefficients. To mitigate this concern, we conduct the estimation using only the images with ground truth labels, and the results are consistent and presented in Online Table C3.

For products such as women's dresses, some reviewers choose to disclose their private information (such as face identity) in the CGIs for self-presentation or providing information to others (Sung et al. 2016, Li and Xie 2020). In Section 2.3, we argue that reviewer personal information disclosure brings about a positive effect as the identity disclosure in CGIs provides a reference for potential buyers to reduce product fit uncertainty. An interaction term of face disclosure and the DID estimator is added to the main model to examine the impact of face disclosure in CGIs, and the results are shown in column (4) of Table 8. As expected, face disclosure positively affects product ratings, emphasizing the informative role of reviewer personal information disclosure.

5.3. CGI Reviewer Subjectivity

Although reviews with CGIs have a higher credibility and are deemed more useful by customers (Hong et al. 2020, Zinko et al. 2020, Li et al. 2022), customers may be susceptible to CGI reviewer subjectivity, especially when reviewers tend to post CGIs at the time when they are satisfied with the product (as illustrated in Figure 3(a)). CGIs with subjective evaluation may have misleading effects on subsequent customers, making them fail to form an objective opinion toward the product, constituting one major source of expectation disconfirmation.

To validate the theoretical argument and the existence of expectation disconfirmation, we conduct subsample regression on different rating groups. Specifically, we divide all the products with CGIs into high and low rating groups and investigate if the outcomes are different in these two groups. Because the average rating on the platform is around four stars, products with five-star CGI reviews are classified into the high-rating group, and products with CGI reviews lower than four stars are classified into the low-rating group. In each group, we conduct the same DID analysis as in Equation (1). From the results in columns (2) and (3) of Table 9, it can be observed that CGIs lead to a significant decline of ratings by as high as 0.15 stars in the high-rating group, whereas there's no significant effect in the low-rating group. The equivalence of the coefficients in these two groups are tested with the Wald test, and the difference is significant at the 1% level.

From another perspective, if the rating of CGI is much higher than the average rating of a specific product, then the negative effect is more significant. To test this argument, we calculate the deviation of the CGI rating from product quality (product quality is shown on the platform and calculated according to a machine learning algorithm by the platform with a weighted average of previous ratings) as an alternative measure of reviewer subjectivity and examine the moderating effect of CGI reviewer subjectivity on subsequent product ratings. The model specification and results are shown in Equation (4) and column (4) of Table 9. Consistent with previous results, the higher the deviation (meaning that CGI rating is much higher than the product quality), the more negative the effect of CGIs.

$$\begin{aligned} Rating_{ijt} = \beta_1 \cdot AfterTreat_{it} + \delta \cdot AfterTreat_{it} \cdot RatingDeviation_i \\ + Controls_{it} + T_t + \alpha_i + \epsilon_{ijt}, \end{aligned} \tag{4}$$

and these findings imply the existence of expectation disconfirmation in high-rating CGIs. These five-star CGIs generate a high prepurchase expectation in subsequent

Table 9. Heterogeneous Effects of CGI Review Rating

Variables	(1) *Overall*		(2) *LowRating*		(3) *HighRating*		(4) *Rating_Deviation*	
AfterTreat	−0.0730***	(0.0242)	0.0738	(0.0553)	−0.1545***	(0.0318)	−0.0554**	(0.0241)
AfterTreat × Deviation							−0.0551***	(0.0192)
Volume	−0.0001	(0.0003)	−0.0007*	(0.0004)	0.0006	(0.0005)	−0.0002	(0.0003)
Valence	−0.1457***	(0.0090)	−0.1767***	(0.0226)	−0.1347***	(0.0111)	−0.1445***	(0.0090)
Variance	0.0915***	(0.0144)	0.0064	(0.0347)	0.1448***	(0.0178)	0.0916***	(0.0145)
TextLen	0.0014**	(0.0006)	0.0021	(0.0015)	0.0009	(0.0007)	0.0014**	(0.0006)
TitleLen	0.0150**	(0.0065)	0.0190	(0.0146)	0.0126	(0.0082)	0.0142**	(0.0065)
ReviewerExpe	0.1432***	(0.0191)	0.2074***	(0.0437)	0.1222***	(0.0242)	0.1434***	(0.0191)
Observations	47,610		9,612		27,920		47,610	
Product fixed effects	Y		Y		Y		Y	
Month fixed effects	Y		Y		Y		Y	
Adjusted R^2	0.148		0.175		0.142		0.148	

Note. Robust standard errors in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

customers, whereas these high expectations may not be met after their own product experiences. With a disconfirmation of expectation, the subsequent ratings and satisfaction go down. By analyzing the heterogeneous effect of CGIs, we shed new light on the underlying mechanisms. The user experiment results presented in Section 6 further provide direct evidence regarding the mechanisms.

5.4. Different Numbers of CGIs

In this section, we further examine the moderating effect of the number of CGIs on subsequent ratings. Following previous literature (Brynjolfsson et al. 2019, Qiao et al. 2020), we add an interaction term of the DID variable and number of CGIs in the model specification as shown in Equation (5), and $CGICount_{it}$ means the number of CGIs that appear for product i at time t.

$$Rating_{ijt} = \beta_1 \cdot AfterTreat_{it} + \gamma \cdot AfterTreat_{it} \cdot CGICount_{it} + Controls_{it} + T_t + \alpha_i + \epsilon_{ijt}. \quad (5)$$

The results in column (2) of Table 10 show that the coefficient of the interaction term is marginally significant and negative, meaning that more CGIs have a more negative effect on subsequent ratings. The reason could be that more CGIs exaggerate the biased expectation and the additional information provided is not significant, therefore leading to a further decline of ratings.

5.5. Results on Search Goods

In the previous analyses, we select women's dresses as the focal product category and discuss the effect of CGIs on subsequent customers' postpurchase satisfaction. In this section, we extend the analyses to a type of search goods. Specifically, we collected data from the same platform and chose a type of typical search goods, lightning cable, which belongs to a subcategory of electronic goods, to reestimate the DID model and conduct heterogeneity analyses. The summary statistics are shown in Online Table B1. Similarly, we conduct propensity score matching before estimation (balance checks before and after matching are shown in Online Table B2). After matching, we estimate the effect of CGIs following Equation (1). The estimation results are shown in Table 11.

For search goods such as lightning cables, as they are relatively standardized products and relevant information can be transmitted effectively through textual content, customers generally do not rely on examining the CGIs much. Therefore, CGIs have less information effect as well as less disconfirmation effect compared with experience goods such as women's dresses. Column (1) of Table 11 shows that the overall effect is insignificant, different from the results on women's dresses. The data are further divided into different groups based on aesthetic quality and CGI review rating, and results show that image aesthetic quality has no significant effect in both groups (columns (2) and (3)), which makes sense as people generally do not care

Table 10. Heterogeneous Effects of CGI Count

Variables	(1) Base		(2) CGI count	
AfterTreat	−0.0602***	(0.0207)	−0.0616***	(0.0201)
AfterTreat × CGI Count			−0.0050*	(0.0030)
Volume	−0.0002	(0.0002)	0.0000	(0.0002)
Valence	−0.1266***	(0.0091)	−0.1272***	(0.0092)
Variance	0.0580***	(0.0140)	0.0573***	(0.0140)
TextLen	0.0009	(0.0006)	0.0010*	(0.0006)
TitleLen	0.0096	(0.0062)	0.0095	(0.0062)
ReviewerExpe	0.1505***	(0.0150)	0.1503***	(0.0150)
Observations	79,396		79,396	
Product FEs	Y		Y	
Month FEs	Y		Y	
Adjusted R^2	0.126		0.126	

Note. Robust standard errors in parentheses.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 11. Estimation Results of Lightning Cables

		Image aesthetics		CGI review rating	
Variables	(1) *Overall*	(2) *LowAes*	(3) *HighAes*	(4) *LowRating*	(5) *HighRating*
AfterTreat	−0.0995	−0.0275	−0.1601	0.1531	−0.2679**
	(0.0836)	(0.1252)	(0.1242)	(0.1314)	(0.1213)
Volume	−0.0029***	−0.0048**	−0.0046***	−0.0041***	−0.0033*
	(0.0008)	(0.0023)	(0.0012)	(0.0015)	(0.0019)
Valence	−0.1549***	−0.1120**	−0.2097***	−0.2050***	−0.1335***
	(0.0277)	(0.0463)	(0.0394)	(0.0523)	(0.0380)
Variance	0.0063	−0.0071	0.0253	−0.0147	0.0341
	(0.0369)	(0.0520)	(0.0537)	(0.0662)	(0.0475)
TextLen	−0.0006	0.0001	−0.0022	−0.0001	0.0003
	(0.0017)	(0.0028)	(0.0026)	(0.0023)	(0.0030)
TitleLen	0.0941***	0.0705	0.1004***	0.0939**	0.0970***
	(0.0252)	(0.0447)	(0.0328)	(0.0408)	(0.0327)
ReviewerExpe	0.2856**	0.1636	0.4002**	0.1051	0.3676**
	(0.1291)	(0.1665)	(0.1930)	(0.2257)	(0.1727)
Observations	6,276	2,935	2,854	2,739	3,299
Product fixed effects	Y	Y	Y	Y	Y
Month fixed effects	Y	Y	Y	Y	Y
Adjusted R^2	0.144	0.137	0.142	0.180	0.0993

Note. Robust standard errors in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

about the photography quality of lightning cable images. In columns (4) and (5), the results show that, in the high-rating group, the effect of CGIs becomes significant and negative, implying that the expectation disconfirmation caused by reviewer subjectivity is quite robust. We also discuss another special type of experience goods for which CGIs contain little product-related information. The descriptive statistics and estimation results are also presented in Online Appendix B. One thing noteworthy about the results is that the expectation disconfirmation effect is further confirmed, meaning that CGIs can affect customer expectation even when the image is irrelevant to the genuine product quality.

To summarize, these analyses show that the overall negative effect is not significant for search goods such as lightning cables. However, the expectation disconfirmation brought by CGIs with high ratings is robust throughout different product categories. These findings enrich our understanding of CGIs' complex effects for different product types. The results across product categories allow us to conclude that the negative effect of CGIs is more prominent for experience goods as people rely more on CGIs and suffer more from the reviewer subjectivity embedded in them, whereas people generally do not care much about CGIs when purchasing search goods.

6. User Study

Our extensive analyses show that CGIs introduce significant impacts on subsequent product ratings, especially for experience goods. To further validate the underlying mechanisms with respect to its effect on expectation disconfirmation and uncertainty reduction, a randomized controlled laboratory experiment was conducted. Specifically, the experiment aims to investigate if CGIs cause expectation disconfirmation by affecting users' expectation toward the product and if CGIs bring about uncertainty reduction by reducing perceived product uncertainty faced by users. Furthermore, through participants' real experiences of the product, we can accurately measure the degree of disconfirmation in different conditions.

6.1. Experiment Design

A laboratory experiment is designed to investigate customers' perception in different CGI conditions. We aim to simulate an online shopping environment in which the potential buyers read the product review information, report their attitudes toward the product, and experience the product by themselves. We bought some women's dresses of a certain style and recruited female participants from a university to participate in the experiment. Specifically, the experiment consists of two stages. In the first stage, a 2 (CGI dummy) × 2 (review valence) between-subjects experiment design was adopted. We manipulated whether a product review contains a CGI, and we manipulated the review valence as either a one- or five-star review, indicating the valence of the review. The related information shown to different groups are presented in Online Appendix D1. The participants were randomly assigned to different groups and exposed to different sets of information depending on which group they were in. They were asked questions about their expectation and degree of uncertainty perceived toward the product. In the second stage, which is three days later, the participants were invited to the laboratory to try on the clothes. All the clothes are the same except for size, considering participants' heterogeneous needs on clothing size. The clothes they experienced in the second stage were exactly the same as the product shown in the review content in the first stage. This ensures that the whole process resembles the online shopping environment. After experiencing the product, participants were asked to complete a second survey, indicating their perceived quality of the product after experiencing it. Referring to previous literature (Diehl and Poynor 2010, Lin et al. 2018), disconfirmation is measured by the difference between perceived quality and prepurchase expectation.

In addition to the main experiment, we also constructed another group to investigate the effect of aesthetic quality, keeping everything else unchanged. Specifically, we lower the brightness and adjust the composition of the image (ratio of foreground and background) without affecting the main object of the CGI to create a low-aesthetics group, treating the original image as the high-aesthetics group. The manipulation of different aesthetic conditions is presented in Online Appendix D2.

There are three main constructs that we measure in this experiment: expectation, product uncertainty, and perceived quality, respectively. Referring to previous literature (Kim et al. 2009, Dimoka et al. 2012), the measurements of these constructs are provided in Online Table D2. All of the items use seven-point Likert scales anchored from "strongly disagree" to "strongly agree." Manipulation questions were set to ensure the quality of the answers, and we restricted the participants to be female as the products were targeted at a female group. One hundred thirty-four effective responses were collected in total. The distributions of different groups' demographic information are presented in Online Table D1.

6.2. Reliability, Validity, and Manipulation Check

Several tests are conducted to make sure the measurement of the constructs is effective. Cronbach's alpha and composite reliability are calculated and shown in Table 12 and are both indicative of strong reliability. Confirmatory factor analysis shows that each item has a much higher loading on its intended construct than on the other construct, suggesting good convergent and discriminative validity. Additionally, the average variances extracted (AVEs) for the three constructs are all above 0.5, and the square root of AVEs of each construct is higher than the cross-correlations with other latent variables, indicating that the variance explained by each construct is larger than the measurement error variance, demonstrating discriminant validity (Fornell and Larcker 1981).

A manipulation check is conducted to ensure the effectiveness of the experiment design. First, manipulation of the CGI review rating is tested through two questions asking about whether the reviewer in the product review is satisfied with the product. There is a significant difference between the two groups ($M_{high\_rating} = 6.32$, $M_{low\_rating} = 1.32, t = 28.62,\ p < 0.0001$). Second, the effectiveness of image aesthetic quality manipulation is tested, which is also successful ($M_{high\_aes} = 4.90$, $M_{low\_aes} = 4.06, t = 2.63,\ p < 0.01$). The related manipulative questions are displayed in Online Table D2.

6.3. Experiment Results

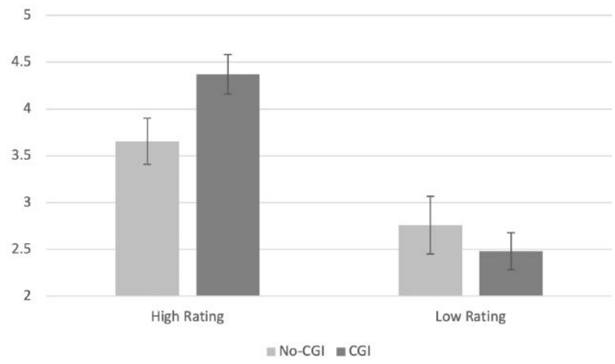
As elaborated in Section 2.3, to validate the existence of expectation disconfirmation, there should be significant changes in customers' prepurchase expectation in the CGI group compared with the no-CGI condition. Similarly, to validate the effect of uncertainty reduction, there should be significant reduction in perceived product uncertainty with CGIs. Furthermore, through comparison of perceived quality and expectation, degree of disconfirmation can be concluded. As mentioned, one of the main sources of expectation disconfirmation is reviewer subjectivity reflected by CGI review rating. Therefore, changes in expectation and product uncertainty for the high- and low-rating groups are separately discussed in this section.

The first question to investigate is how expectation changes in the CGI group compared with the no-CGI group. From Figure 9, *t*-test results show that, in the high-rating group, the appearance of CGIs increases prepurchase expectation compared with the group without CGIs ($m_{no} = 3.65$, $m_{CGI} = 4.37$, $p = 0.0004$). The 95% confidence interval of the mean value in each group is also shown in the figure. These results validate that CGIs increase people's expectation. For the low-rating group, however, the result is opposite in that reviews with CGIs lower people's expectation compared with the group without CGIs ($m_{no} = 2.76$, $m_{CGI} = 2.48$, $p > 0.1$) with no significant difference. These results validate our

Table 12. Reliability and Validity Test Results

	Cronbach's alpha	Composite reliability	AVE	Correlation of factors: Expectation	Perceived quality	Uncertainty
Expectation	0.914	0.939	0.793	0.890		
Perceived quality	0.871	0.908	0.711	0.144	0.843	
Uncertainty	0.771	0.849	0.585	0.320	0.210	0.765

Figure 9. Expectation in No-CGI and CGI Groups



theoretical arguments that CGIs alter people's expectation upward or downward because of CGI reviewers' subjective evaluation. Further, because of the two-stage experimental design, we can measure participants' attitudes after experiencing the product. Specifically, in the high-rating group, participants in the CGI group experience a much higher expectation than perceived quality, therefore leading to a significant negative disconfirmation ($m_{CGI} = -0.45$), meaning that prepurchase expectation exceeds perceived true quality. For the no-CGI group, there is a positive disconfirmation ($m_{no} = 0.36$), meaning that perceived quality exceeds prepurchase expectation. The difference of disconfirmation between the CGI and no-CGI groups is significant ($m_{no} = 0.36$, $m_{CGI} = -0.45$, $p = 0.021$). In the low-rating group, CGI and no-CGI groups both demonstrate positive disconfirmation with their differences insignificant ($m_{no} = 0.76$, $m_{CGI} = 1.33$, $p > 0.1$) as shown in Figure 10. This further confirms that the high expectation caused by CGIs in the high-rating group causes expectation inflation and, subsequently, a negative expectation disconfirmation.

The second question to investigate is how customers perceive product uncertainty in different CGI and rating conditions. Results in Figure 11 show that, in the high-rating group, there are no significant changes in the CGI group compared with the no-CGI group ($m_{no} = 4.84$, $m_{CGI} = 4.51$, $p > 0.1$).[9] For the low-rating group, however, the difference is significant ($m_{no} = 5.45$, $m_{CGI} = 4.99$, $p = 0.047$). In other words, CGIs only demonstrate their information effect when the rating is low. The *t*-test results on expectation, uncertainty, perceived quality, and disconfirmation are also shown in Table 13. Regression analyses results controlling participant demographic information are presented in Online Tables D4 and D5.

Next, the experimental design also allows us to investigate the information and disconfirmation effects of different aesthetic quality conditions. With respect to expectation, in the high-rating group, the high-aesthetics group has a higher expectation compared with the low-aesthetics group ($m_{high\_aes} = 4.37$, $m_{low\_aes} = 3.81$, $p = 0.015$), validating the statement that high aesthetic quality significantly increases customers' prepurchase expectation. However, the difference of disconfirmation in the two groups is not significant ($m_{high\_aes} = 0.45$, $m_{low\_aes} = 0.23$, $p > 0.1$). With respect to product uncertainty, there is no significant difference between the high- and low-aesthetics groups ($m_{high\_aes} = 4.55$, $m_{low\_aes} = 4.82$, $p > 0.1$), showcasing the limited effect of aesthetic quality on product uncertainty reduction.

To summarize, these results indicate that, in the high-rating group, the disconfirmation effect plays a more prominent role than the information effect. Previous literature documents the negative relationship between disconfirmation and satisfaction (Anderson 1973, Oliver 1980), which explains the decline of subsequent postpurchase satisfaction in the high-rating group because of the CGI disconfirmation effect. Because most of the CGI reviews have a high rating from our observational data (Figure 3(a)), it causes the overall expectation to be higher than products without CGIs, and when the higher expectation is not met, unsatisfactory experiences happen, explaining the negative effect on subsequent product ratings in column (3) of Table 5. Considering that there's a time interval between when a customer observes the CGI and receives the product, we further designed another online experiment to validate that customers still remember the CGI-related information after

Figure 10. Disconfirmation in No-CGI and CGI Groups

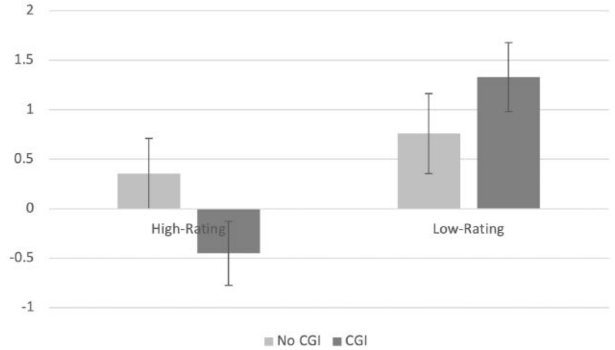


Figure 11. Uncertainty in No-CGI and CGI Groups

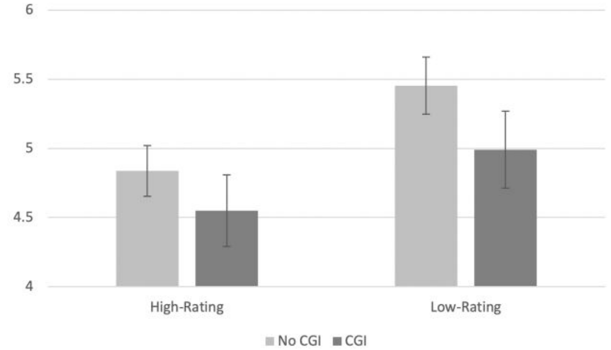


Table 13. *T*-Test Results in High- and Low-Rating Groups

	High-rating group				Low-rating group			
	CGI	No-CGI	*t* statistic	*p*-value	CGI	No-CGI	*t* statistic	*p*-value
Expectation	4.37	3.65	2.77	0.004	2.48	2.76	−0.98	0.166
Uncertainty	4.51	4.84	−1.28	0.102	4.99	5.45	−1.71	0.047
Perceived quality	3.92	4.01	−0.26	0.396	3.81	3.52	0.77	0.223
Disconfirmation	−0.45	0.36	−2.09	0.021	1.33	0.76	1.30	0.100

three days' time with the detailed experimental design and results discussed in Online Appendix E.

7. Conclusions and Implications

Although extensive research investigates the persuasive advantages of images, our study focuses on the images generated by customers and their effect on subsequent product ratings. Rigorous econometric models, the latest progress in computer vision, and randomized controlled experiments are leveraged to conduct in-depth and extensive investigations. Our main conclusion is that CGIs lead to a decline in subsequent product ratings. Additionally, CGIs with high ratings have a significantly more negative effect than CGIs with low ratings, providing initial evidence of the disconfirmation effect brought by reviewer subjectivity. Differently, face disclosure in CGIs has a more positive effect on subsequent product ratings with more information provision especially for products such as women's dresses. To validate the causal mechanism, a laboratory experiment was conducted, and the effect of CGIs on expectation disconfirmation is further confirmed.

Theoretically, this research adds to recent discussion on the potential value and impact of a new type of UGC, CGIs. To the best of our knowledge, our study is among the first to demonstrate the complex effects of CGIs in online review systems, and a new factor is identified that could influence rating dynamics. Specifically, drawing upon ECT and image-related literature, CGIs have both information and disconfirmation effects that could potentially affect customers' product evaluation. Results from secondhand data and a randomized controlled experiment confirm that expectation disconfirmation brought by CGIs plays a larger role than uncertainty reduction. These findings are novel and enlightening as they are different from previous findings that suggest the positive influences of UGC, such as boosting sales (Chen and Xie 2008), increasing trust (Goh et al. 2013), etc. This study also differs from previous research efforts that mainly demonstrate the positive implications of images (So and Oh 2018, Zhang et al. 2021). Rather, this research demonstrates that, because of CGIs' specific characteristics, postpurchase satisfaction could be hampered with its distortive role in affecting customers' prepurchase expectation. Moreover, several salient factors are detected that could affect the strength of the information and disconfirmation effects, which extend the boundary of existing literature on review images. For instance, reviewer personal information disclosure is identified as a new factor that affects postpurchase satisfaction in addition to its effect on trust and review helpfulness (Cyr et al. 2009, Karimi and Wang 2017). Aesthetic quality and CGI review rating are identified as two main factors that cause disconfirmation, providing alternative insights to existing research on the positive effect of aesthetic quality in promoting sales and user engagement (Li and Xie 2020, Zhang et al. 2021).

This research provides rich managerial implications. Customer-generated images are becoming more prevalent nowadays. Some platforms are totally composed of user-generated multimedia content, such as Pinterest. Therefore, investigating the effect of CGIs has substantial managerial implications considering its prevalence and wide adoption in today's business world. Even on some traditional e-commerce platforms, such as Amazon, reviews with images are more likely to be placed in salient positions signaling the emphasis of the platforms on customer-generated images. However, a CGI is a double-edged sword. Although CGIs usually serve as an efficient promotional tool to persuade customers into purchases, they can also backfire, triggering unsatisfactory purchase experiences. Platforms or online retailers should use caution in the practice of taking advantage of CGIs to boost sales as they may bring about a negative effect on future ratings, damaging the long-term relationship with customers and hurting the platform's reputation. Despite the downside of CGIs, some practical measures can be taken to alleviate or eliminate this negative effect. As illustrated by these findings, both time and CGI content matter. The negative effect of CGIs generated in earlier periods gradually disappears with more new reviews appearing, and therefore, customers should be more careful with newly posted CGIs. With respect to the content of CGIs, platforms should encourage customers to post CGIs with less subjectivity and more usefulness that truly benefit potential buyers. Moreover, a better privacy-protection policy could be considered as it could alleviate reviewers' privacy concerns and give reviewers more incentives to disclose personal information (such as faces) on the platform. Finally, because CGIs' effect is dependent on the specific product type, personalized measures can be taken to maximize the value of CGIs corresponding to different products.

Future research efforts can be extended in several directions. First, we mainly discuss the effect of CGIs on rating dynamics in this study, whereas it may also affect customers' purchase behavior in some different manner, especially the case when some CGIs with low ratings or low aesthetics prevent customers from purchasing, which may explain why we did not observe significant effect when CGIs have a low rating in our estimation results. This could be more complex than prior findings that CGIs can increase purchase intention (Zinko et al. 2020) and provides a promising direction to explore if related data are accessible. Second, our laboratory experiment sheds light on the mechanism of how CGIs affect people's perception, and we drew on prior literature to illustrate the relationship between disconfirmation (uncertainty) and satisfaction. Future research can strengthen our research findings with more extensive experimental results if possible. Third, the reviewers who choose to post CGIs may possess certain characteristics, and investigating the CGI posting behavior itself could provide meaningful managerial implications as well. Specifically, by detecting the probability of a reviewer posting CGIs, retailers could launch customized marketing campaigns in advance. Furthermore, text indeed plays a role in affecting the influence of CGIs as prior research points out that the imagery-provoking ability of text (Unnava and Burnkrant 1991) affects recall of information in advertisements. On e-commerce platforms, there could be interactions between review texts and CGIs, and it is worthy of investigation in future research work. Finally, as multimedia content generation and consumption have penetrated every aspect of everyday life, we believe this research paves a new way to encourage more related research efforts to deepen our understanding of new forms of UGC and their social and economic impacts.

Acknowledgments

The authors thank the senior editor, associate editor, and three anonymous reviewers for their insightful comments and constructive suggestions throughout the review process.

Endnotes

[1] See https://www.statista.com/forecasts/997048/social-media-activities-in-the-us.

[2] See https://www.emarketer.com/content/online-shoppers-expectations-for-visual-merchandising-rises-dramatically.

[3] See https://www.trendreports.com/article/imagebased-social-media.

[4] Positive disconfirmation means product quality is better than the prepurchase expectation, and negative disconfirmation means that product quality is worse than the prepurchase expectation.

[5] All of the collected data are public information. Additionally, in conducting this research, we fully protected platform users' privacy and did not disclose their identity information in this research.

[6] Three-day moving average is performed to increase the smoothness of the plot in Figure 4.

[7] Because 86% of the reviews are posted by reviewers without previous review experiences, we think that reviewer strategic rating behavior is not a big concern in this paper.

[8] For those products with more than one CGI, we only keep reviews generated before the second CGI appears to eliminate the effect of multiple CGIs.

[9] To further exclude the effect of review text length, we conducted a supplemental online experiment with different review lengths and found that, whether the review text has a short or long length, CGI's effect on uncertainty reduction is not significant in the high-rating group.

References

Anderson RE (1973) Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *J. Marketing Res.* 10(1):38–44.

Aydinoğlu NZ, Cian L (2014) Show me the product, show me the model: Effect of picture type on attitudes toward advertising. *J. Consumer Psych.* 24(4):506–519.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quart. J. Econom.* 119(1): 249–275.

Brynjolfsson E, Hui X, Liu M (2019) Does machine translation affect international trade? Evidence from a large digital platform. *Management Sci.* 65(12):5449–5460.

Chen P, Hitt LM, Hong Y, Wu S (2021) Measuring product type and purchase uncertainty with online product ratings: A theoretical model and empirical application. *Inform. Systems Res.* 32(4):1470–1489.

Chen Y, Xie J (2008) Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Sci.* 54(3):477–491.

Chen-Yu HJ, Kincade DH (2001) Effects of product image at three stages of the consumer decision process for apparel products: Alternative evaluation, purchase and post-purchase. *J. Fashion Marketing Management* 5(1):29–43.

Childers TL, Houston MJ (1984) Conditions for a picture-superiority effect on consumer memory. *J. Consumer Res.* 11(2):643–654.

Cyr D, Head M, Larios H, Pan B (2009) Exploring human images in website design: A multi-method approach. *Management Inform. Systems Quart.* 33(3):539–566.

de Chaisemartin C, D'Haultfoeuille X (2020) Two-way fixed effects estimators with heterogeneous treatment effects. *Amer. Econom. Rev.* 110(9):2964–2996.

Deng Y, Loy CC, Tang X (2017) Image aesthetic assessment: An experimental survey. *IEEE Signal Processing Magazine* 34(4):80–106.

Deng Y, Zheng J, Khern-am-nuai W, Kannan KN (2021) More than the quantity: The value of editorial reviews for a UGC platform. *Management Sci.* 68(9):6865–6888.

Diehl K, Poynor C (2010) Great expectations?! Assortment size, expectations, and satisfaction. *J. Marketing Res.* 47(2):312–322.

Dimoka A, Hong Y, Pavlou PA (2012) On product uncertainty in online markets: Theory and evidence. *Management Inform. Systems Quart.* 36(2):395–426.

Forman C, Ghose A, Wiesenfeld B (2008) Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Systems Res.* 19(3):291–313.

Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *J. Marketing Res.* 18(1):39–50.

Freeman M (2007) *The Photographer's Eye: Composition and Design for Better Digital Photos* (Routledge, New York).

Godes D, Silva JC (2012) Sequential and temporal dynamics of online opinion. *Marketing Sci.* 31(3):448–473.

Goh KY, Heng CS, Lin Z (2013) Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. *Inform. Systems Res.* 24(1):88–107.

Goodman-Bacon A (2021) Difference-in-differences with variation in treatment timing. *J. Econometrics* 225(2):254–277.

Hagtvedt H, Patrick VM (2008) Art infusion: The influence of visual art on the perception and evaluation of consumer products. *J. Marketing Res.* 45(3):379–389.

Ho YC, Wu J, Tan Y (2017) Disconfirmation effect on online rating behavior: A structural model. *Inform. Systems Res.* 28(3):626–642.

Hong W, Yu Z, Wu L, Pu X (2020) Influencing factors of the persuasiveness of online reviews considering persuasion methods. *Electronic Commerce Res. Appl.* 39:100912.

Hong Y, Pavlou PA (2014) Product fit uncertainty in online markets: Nature, effects, and antecedents. *Inform. Systems Res.* 25(2):328–344.

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. Preprint, submitted April 17, https://arxiv.org/abs/1704.04861.

Hu N, Pavlou PA, Zhang JJ (2017) On self-selection biases in online product reviews. *Management Inform. Systems Quart.* 41(2):449–471.

Huang N, Hong Y, Burtch G (2017) Social network integration and user content generation: Evidence from natural experiments. *Management Inform. Systems Quart.* 41(4):1035–1058.

Jiang Z, Wang W, Tan BC, Yu J (2016) The determinants and impacts of aesthetics in users' first interaction with websites. *J. Management Inform. Systems* 33(1):229–259.

Karimi S, Wang F (2017) Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems* 96:39–48.

Kim DJ, Ferrin DL, Rao HR (2009) Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. *Inform. Systems Res.* 20(2):237–257.

Kisielius J, Sternthal B (1984) Detecting and explaining vividness effects in attitudinal judgments. *J. Marketing Res.* 21(1):54–64.

Lee YJ, Hosanagar K, Tan Y (2015) Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Sci.* 61(9):2241–2258.

Li H, Ji H, Liu H, Cai D, Gao H (2022) Is a picture worth a thousand words? Understanding the role of review photo sentiment and text-photo sentiment disparity using deep learning algorithms. *Tourism Management* 92:104559.

Li X, Hitt LM (2008) Self-selection and information role of online product reviews. *Inform. Systems Res*. 19(4):456–474.

Li Y, Xie Y (2020) Is a picture worth a thousand words? An empirical study of image content and social media engagement. *J. Marketing Res.* 57(1):1–19.

Lim KH, Benbasat I, Ward LM (2000) The role of multimedia in changing first impression bias. *Inform. Systems Res.* 11(2):115–136.

Lin C, Wei YH, Lekhawipat W (2018) Time effect of disconfirmation on online shopping. *Behav. Inform. Tech.* 37(1):87–101.

Lin Z, Heng CS (2015) The paradoxes of word of mouth in electronic commerce. *J. Management Inform. Systems* 32(4):246–284.

Lin Z, Zhang Y, Tan Y (2019) An empirical study of free product sampling and rating bias. *Inform. Systems Res.* 31(1):260–275.

Liu L, Dzyabura D, Mizik N (2020) Visual listening in: Extracting brand image portrayed on social media. *Marketing Sci.* 39(4): 669–686.

Liu Y, Du R (2020) Examining the effect of reviewer socioeconomic status disclosure on customers' purchase intention. *J. Global Inform. Systems Management* 28(3):17–35.

Luo J, Ba S, Zhang H (2012) The effectiveness of online shopping characteristics and well-designed websites on satisfaction. *Management Inform. Systems Quart.* 36(4):1131–1144.

Ma Y, Xiang Z, Du Q, Fan W (2018) Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. *Internat. J. Hospitality Management* 71:120–131.

Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res*. 48(3):444–456.

Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Sci.* 341(6146):647–651.

Mudambi SM, Schuff D (2010) Research note: What makes a helpful online review? A study of customer reviews on Amazon.com. *Management Inform. Systems Quart.* 34(1):185–200.

Oliver RL (1977) Effect of expectation and disconfirmation on post-exposure product evaluations: An alternative interpretation. *J. Appl. Psych.* 62(4):480–486.

Oliver RL (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Marketing Res.* 17(4):460–469.

Peracchio LA, Meyers-Levy J (2005) Using stylistic properties of ad pictures to communicate with consumers. *J. Consumer Res.* 32(1):29–40.

Qiao D, Lee SY, Whinston AB, Wei Q (2020) Financial incentives dampen altruism in online prosocial contributions: A study of online reviews. *Inform. Systems Res.* 31(4):1361–1375.

Rishika R, Ramaprasad J (2019) The effects of asymmetric social ties, structural embeddedness, and tie strength on online content contribution behavior. *Management Sci.* 65(7):3398–3422.

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Shen W, Hu Y, Ulmer JR (2015) Competing for attention: An empirical study of online reviewers' strategic behavior. *Management Inform. Systems Quart.* 39(3):683–696.

Shin D, He S, Lee GM, Whinston AB, Cetintas S, Lee KC (2020) Enhancing social media analysis with visual data analytics: A deep learning approach. *Management Inform. Systems Quart.* 44(4):1459–1492.

So H, Oh W (2018) Picture perfect: An image mining of advertising content and its effects on social targeting. *Proc. 39th Internat. Conf. Inform. Systems* (Association for Information Systems, Atlanta), 1–17.

Somerfield K, Mortimer K, Evans G (2018) The relevance of images in user-generated content: A mixed method study of when, and why, major brands retweet. *Internat. J. Internet Marketing Advertising* 12(4):340–357.

Sunder S, Kim KH, Yorkston EA (2019) What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions. *J. Marketing* 83(6):93–112.

Sung Y, Lee JA, Kim E, Choi SM (2016) Why we post selfies: Understanding motivations for posting pictures of oneself. *Personality Individual Diffferences* 97:260–265.

Talebi H, Milanfar P (2018) NIMA: Neural image assessment. *IEEE Trans. Image Processing* 27(8):3998–4011.

Tian X, Dong Z, Yang K, Mei T (2015) Query-dependent aesthetic model with deep learning for photo quality assessment. *IEEE Trans. Multimedia* 17(11):2035–2048.

Truong QT, Lauw HW (2017) Visual sentiment analysis for review images with item-oriented and user-oriented CNN. *Proc. 25th ACM Internat. Conf. Multimedia* (ACM, New York), 1274–1282.

Unnava HR, Burnkrant RE (1991) An imagery-processing view of the role of pictures in print advertisements. *J. Marketing Res.* 28(2):226–231.

Wang A, Zhang M, Hann IH (2018) Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Inform. Systems Res.* 29(3):641–655.

Wang M, Li X, Chau PYK (2021) Leveraging image-processing techniques for empirical research: Feasibility and reliability in online shopping context. *Inform. Systems Frontiers* 23(3):607–626.

Wang M, Li X, Chau PYK (2016) The impact of photo aesthetics on online consumer shopping behavior: An image processing-enabled empirical study. *Proc. 37th Internat. Conf. Inform. Systems* (Association for Information Systems, Atlanta), 1–11.

Wheeler L, Martin R, Suls J (1997) The proxy model of social comparison for self-assessment of ability. *Personality Soc. Psych. Rev.* 1(1):54–61.

Wolfers J (2006) Did unilateral divorce laws raise divorce rates? A reconciliation and new results. *Amer. Econom. Rev.* 96(5):1802–1820.

Xu Q (2014) Should I trust him? The effects of reviewer profile characteristics on eWOM credibility. *Comput. Human Behav.* 33:136–144.

You Y, Vadakkepatt GG, Joshi AM (2015) A meta-analysis of electronic word-of-mouth elasticity. *J. Marketing* 79(2):19–39.

Zinko R, Stolk P, Furner Z, Almond B (2020) A picture is worth a thousand words: How images influence information quality and information load in online reviews. *Electronic Marketing* 30:775–789.

Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Lett.* 23(10):1499–1503.

Zhang S, Lee D, Singh PV, Srinivasan K (2021) What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Management Sci.* 68(8):5644–5666.