

Article

Understanding of Customer Decision-Making Behaviors Depending on Online Reviews

Yeo-Gyeong Noh ¹, Junryeol Jeon ¹ and Jin-Hyuk Hong ^{1,2,*}¹ School of Integrated Technology, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea² Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea

* Correspondence: jh7.hong@gist.ac.kr

Abstract: With a never-ending stream of reviews propagating online, consumers encounter countless good and bad reviews. Depending on which reviews consumers read, they get a different impression of the product. In this paper, we focused on the relationship between the text and numerical information of reviews to gain a better understanding of the decision-making process of consumers affected by the reviews. We evaluated the decisions that consumers made when encountering the review structure of star ratings paired with comments, with respect to three research questions: (1) how consumers compare two products with reviews, (2) how they individually perceive a product based on the corresponding reviews, and (3) how they interpret star ratings and comments. Through the user study, we confirmed that consumers consider reviews differently according to product presentation conditions. When consumers were comparing products, they were more influenced by star ratings, whereas when they were evaluating individual products, they were more influenced by comments. Additionally, consumers planning to buy a product examined star ratings by more stringent criteria than those who had already purchased the product.

Keywords: consumer decision-making; star rating; review comment; sentiment analysis



Citation: Noh, Y.-G.; Jeon, J.; Hong, J.-H. Understanding of Customer Decision-Making Behaviors Depending on Online Reviews. *Appl. Sci.* **2023**, *13*, 3949. <https://doi.org/10.3390/app13063949>

Academic Editor: Christos Bouras

Received: 13 February 2023

Revised: 13 March 2023

Accepted: 18 March 2023

Published: 20 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Numerous reviews are being made and uploaded through the Internet today. This applies not only to tangible products, such as fans and vacuum cleaners, but also to intangible consumer goods, such as concerts and movies. These reviews are generated by consumers and have a huge impact on consumers' purchasing decisions [1,2]. Consumers perceive reviews as more reliable and engaging than information provided by vendors because they are created by fellow consumers who have already purchased and have been using the products [3]. From these reviews, companies obtain valuable information from the consumers' perspectives of their products [4]. In addition, online reviewing has been integrated by many companies because it is an effective way to not only engage consumers directly with the products, but also influence consumers [5].

There is a large diversity of reviews due to reflecting consumers' perceptions, and there is usually a variety of consumer perspectives even for a single product. In a related study [6], it was confirmed that the tendency to be affected by review data varies depending on the consumer type of how many expectations are made for the product. Depending on the type of consumer, there are differences, such as being more vulnerable to negative reviews or being affected by the selection. In the case of user types in which the reliability of reviews may be seen as a measure, factors, such as sources, are also affected [7]. From a seller's point of view, naturally, they want their products to have a positive perception. However, it is not advantageous to consumers for only the large number of good reviews to have high visibility, which is often the case. Reviews should be transparent and unbiased.

User reviews implemented by companies have two main characteristics. The first is the numerical review, such as a one to five-star rating, and the second is the text review, such as a comment. In most cases, star ratings and comments are shown together. Company sites that provide consumer goods with reviews, such as Amazon, Walmart, and Movietickets provide star ratings and comments together. Yet, it is unclear which review component, either star ratings or comments, holds more weight to the consumer. Do consumers prefer consumer goods with high star ratings but poor comments or vice versa? What do consumers think when they look at the star rating to make a purchase decision? This paper explores the decisions that consumers make when presented with both review components simultaneously. Thus, three tasks were devised through our user study to simulate consumers' reading and interpretation of reviews. As a result of the user study, we had the following observations; (1) Differences in the importance of the review components according to the purpose of reading them, and (2) arbitrary interpretation of reviews by consumers that were different from the intention of the reviewers.

2. Background

2.1. Attempts to Understand Consumers through Online Reviews

Online reviews provide valuable data that reflect the interactions between consumers and products. They comprise feedback and viewpoints of consumers with respect to products, serving as a form of electronic word-of-mouth [8], based on users' willingness to share their opinions. Understanding consumers through online reviews can help researchers and practitioners in the product-related field to enhance consumer purchasing intentions and loyalty [9,10]. Based on the better understanding of consumers, recommendation systems have potential to not only augment consumers' purchasing experience, but also boost sales revenue [11].

For this, the foremost step is to gain insight into the consumer experience. Consumer experience refers to an individual's sensory, emotional, cognitive, behavioral, and relational experiences as a consumer with a product or service [12]. Depending on the situation, this experience can be remembered positively or negatively, and if it is remembered positively, the consumer will feel satisfaction [13]. By identifying the key features that contribute to consumer experience, product and service providers can identify areas where they should focus their efforts [14,15]. Text-based review data analysis is a common method used to analyze key features. Latent Semantic Analysis [16,17] can be utilized to extract hidden semantic patterns of words and phrases that make up the document corpus [18].

Based on user experience and preference information that have been identified in previous research, the recommendation system has the ability to suggest more appropriate products [19]. The recommendation system incorporates a content-based filtering technique that assesses the similarity between a product and consumer preference [20]. However, in recent times, the collaborative filtering technique has gained prominence [21]. This is largely due to the abundance of online reviews, which facilitates the learning of a model that can predict an individual's interests. The most representative methods for predicting user interests include the neighborhood method and latent-factor model [22]. The neighborhood method involves finding the group of users and items that are the most similar to the individual's interests using algorithms such as the K-nearest neighbors [22,23]. The latent-factor model method involves vectorizing users and items and modeling the relationship between them using neural networks in the latent space [24,25].

2.2. Studies on Helpfulness of Reviews

It is important to understand the helpfulness of reviews to allow consumers to be exposed to more holistic and better summarized review information, despite the excess of review information in the modern age [26]. The helpfulness of reviews can be identified by analyzing which characteristics of reviews are essential to consumers [27–29]. Quantitative characteristics, such as the numerical rating of reviews or the length of a review [30], or qualitative characteristics, such as sentiment or review complexity [31], affect the ability

to infer the quality of a consumer product. In addition, factors affecting review helpfulness include reviewer-related factors (e.g., writer reputation), reader-related factors (e.g., reader identification), and environmental factors (e.g., website reputation) [27]. Through techniques such as natural language processing (NLP), machine learning, and deep learning, review helpfulness is regressed or classified by the factors mentioned above [32–36]. According to past studies, negative reviews tend to be more impactful than positive reviews [37,38] and reputation deterioration caused by negative reviews has an adverse effect on profits [39]. For both the product providers and consumers, it is necessary to present negative reviews, positive reviews, and an overall picture of consumer reviews all together, appropriately sorted.

Talton et al. studied a method of rating sorting and observed a user's choice when presented with two different rating scenarios [40]. Rating scenarios were a combination of a thumbs-up/thumbs-down ratio and a sample size. For example, there are comparisons where one distribution has a higher positive ratio but a lower number of total votes. To compare the two distributions, they calculated inter-rater reliability, a percentage value obtained by dividing the absolute value of the difference in the number of selections between two combinations by the total number of selections. Additionally, by measuring the time-to-rank, the time taken to decide which combination is more preferred, observations were made on the difficulty of ranking these scenarios.

Koji et al. developed Review Spotlight which summarizes text reviews and presents them as adjective-noun pairs [41]. Review Spotlight uses sentiment analysis to select adjacent adjectives and nouns in the texts and matches them as pairs. The pairs shown to the user have different colors depending on the three cases, where their sentiment is interpreted as positive, neutral, or negative. If the user clicks on the adjective-noun pair, the original text review of the pair is shown.

2.3. Two Types of Reviews: Numeric and Text

User reviews typically have a numerical and text portion. A rating scale is a method that requires the rater to assign a value to a measure for an assessed attribution. Several rating scales have been developed for numerical review systems. According to Chen's study, high traffic websites mainly adopt one of two types of rating schemes [42]. One is a binary thumbs-up/thumbs-down system and the other is the one to five-star rating system. Binary thumbs-up/thumbs-down rating is a rating scale that doesn't allow for neutrality. This forces the rater to assign a good or bad designation, but does not provide insight into the degree of how good or bad [43]. The one to five-star is widely used in commercial online reviews. It is scored on a scale of one to five, with more stars equated to a more positive rating. Numerical reviews are advantageous in assigning a scaled system of scoring, providing greater insight [44]. The standard of each score in the star ratings may be different for each individual rater, but since the deviation of the difference is relatively small, the distribution of scores roughly follows a normal distribution. Therefore, a star rating system is regularly averaged to provide a numeric rating.

A text review includes qualitative information that is basically impossible to provide through a rating scale review. Being subjective, variation is high for text reviews [45]. Raters will often leave the same numerical ratings, but rarely have the same text reviews. Text reviews are also more subject to subjective variation. A rater leaving a four-star review may leave a comment, "this product was very good", while another may leave the same four-star review with the comment, "this product was somewhat disappointing." Therefore, a diversified number of text reviews for a product cannot be easily summarized or averaged, so the method currently employed by most companies is to list text reviews by the number of likes or by recency. However, listing by number of likes or recency can be too random or biased.

2.4. Research Question

Based on the importance and availability of the aforementioned online reviews, we studied user selection patterns considering multiple combinations of a star rating and comments. This paper aims to provide a better understanding of the consumer's decision-making process in an environment where they need to introduce attractive reviews to consumers or give them trust when providing reviews. To this end, three research questions were specifically designed according to the purpose: (1) First, in the choice between two things, we investigated if the star rating or the comment had a stronger impact on consumers. (2) When a single product is given with a star rating and comments, how the user makes choices was also studied. (3) Finally, when a star rating was given, the reliability interval of credible comments was investigated; for example, how a review with a three-point star rating affects the trustworthiness of the comment. We formulated two specific hypotheses with the aim of understanding how consumers interpret the combination of horoscopes and reviews, based on the research questions at hand.

Hypothesis H1. *In a binary selection situation, when star rating and reviews are oppositely given, the option with a high star rating will be chosen to prevail.*

Hypothesis H2. *In a single selection situation, deciding on a single subject alone, review data will be considered to be more important than star rating.*

3. Review Data Construction

3.1. Why Movie Review

We used movie reviews as our dataset in this study. Among numerous consumer product reviews, movie reviews were selected because the price of movies is not variable and emotional language is usually used in movie reviews.

The pricing of a product is considered to be one of the primary factors that significantly impacts the outcomes of consumer reviews. The price–quality inference posits that consumers perceive price and quality as strongly correlated [46]. Moreover, a product's quality level is often objectively reflected in its performance, as perceived by consumers [47,48]. Consequently, as the price of a product increases, customers are more likely to be satisfied [49]. In light of the fact that review data itself is indicative of consumer satisfaction, the various ways in which prices are formed in the review data may act as a confounding factor in our data design. However, movie review data is relatively free from this factor since all movies are equally priced.

Sentiment analysis is a technique utilized for information extraction [50] in which the objective is to gather and categorize emotions or attitudes towards a specific topic or entity [51,52]. In our study, we opted to apply sentiment analysis to review comments as a means of classification. We specifically chose to analyze movie reviews, as they are known to be more informal and emotionally charged, often displaying personal tendencies [53]. We deemed this strategy to be effective for our research purposes. We crawled information from a search engine that provides free reviews and scores written by users on movies.

3.2. Experimental Data Preparation

The crawled movie reviews of individual users consist of a comment with a star rating mapped to it. Since each individual has different criteria for grading comments and stars, we could not immediately use all the collected comments for the experiment. Therefore, we selected comments to represent a specific star rating score range. This was done by first mapping the sentiment frequently used in comments to each specific star rating section. Then, we selected comments containing a sentiment with high frequency of use as a representative review of the star rating section. The overall process is shown in Figure 1. The GoEmotions dataset is the largest manually annotated dataset of 58,000 English Reddit comments, labeled for 27 emotional categories [54]. To perform sentiment analysis on the collected review comments, we translated the GoEmotions dataset to Korean and then trained it

through KoELECTRA [55] to create a sentiment analysis model. We applied the collected Korean movie review comment data to this model to analyze and identify sentiments.

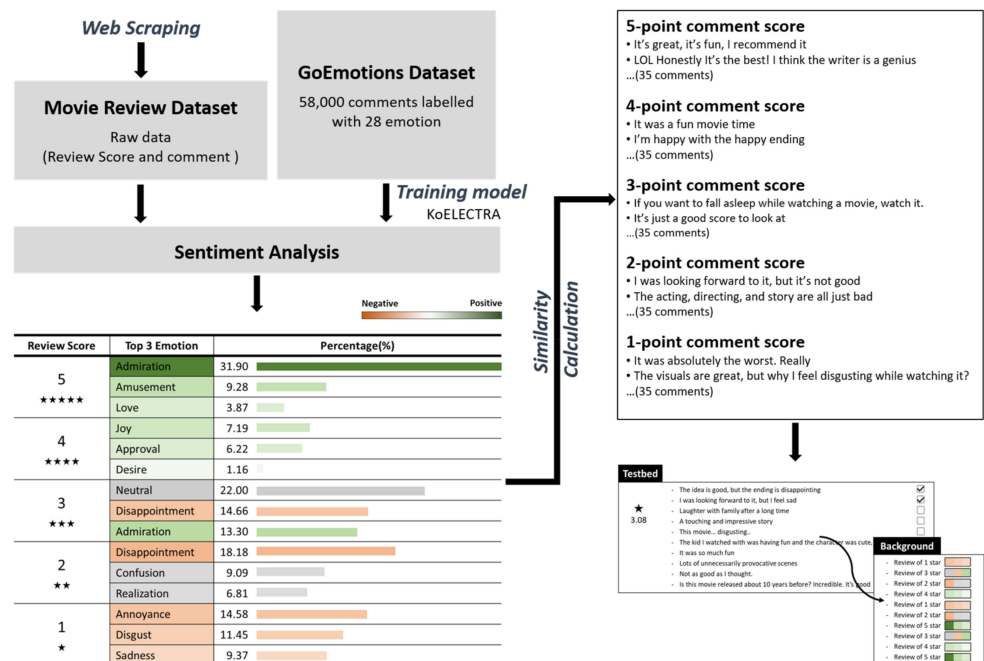


Figure 1. Overall process of experimental data preparation.

As a result of extracting emotional information from movie review scores, we correlated specific emotions to scores, as shown in Figure 1, marked with the three highest percentage emotions per each star rating. In the five-star range, admiration (31.9%) was the highest by far, followed by amusement (9.28%) and love (3.87%). The emotions that appear in the four-star range were joy (7.19%), approval (6.22%), and desire (1.16%). The three most common emotions in the three-star range were neutral (22%), disappointment (14.7%), and admiration (13.3%). Based on the general acceptance of the three-star range, we decided to only consider neutral sentiments in the three-star range. (Note that admiration and disappointment were included in the five-star and two-star ranges.) Emotions, such as disappointment (18.18%), disapproval (11.36%), and confusion (9.09%) were the most frequent in the two-star range, with emotions such as curiosity (4.54%) and optimism (2.27%) being rare. Curiosity and optimism were emotions extracted from mostly sarcastic reviewers. Lastly, annoyance (14.58%), disgust (11.45%), and sadness (9.37%) were the most common in the one-star range. We calculated the similarity of the Euclidean distance, as shown in formula (1), for the three selected emotions from the one to five-star ratings and selected the 165 comments closest to the five-star ranges (165 comments = 35 comments × 5 star ranges) as follows:

$$\text{similarity} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (1)$$

Emotion Type (x, y, z), Representative Emotion Value (x_2, y_2, z_2), Review Emotion Value (x_1, y_1, z_1)

Finally, we constructed a pool of comments, consisting of 165 comments for five-star ranges, and redefined the ranges as comment scores (CSs) from one to five, i.e., each CS corresponded with 35 comments.

3.3. Star Rating Score Presentation

Unlike comments, star ratings are presented as the average value of all ratings. However, for the average star rating, the collected scores ranged from 2.5 to 5 points. We divided

the range of collected star ratings into five and applied it to the user study. We decided to call these five ranges the star rating ranges (SRRs) as follows. (SRR #1: $2.5 < x \leq 3.0$, SRR #2: $3.0 < x \leq 3.5$, SRR #3: $3.5 < x \leq 4.0$, SRR #4: $4.0 < x \leq 4.5$, SRR #5: $4.5 < x \leq 5.0$, x is average star rating score).

4. User Study

4.1. Overview

We designed three tasks corresponding to each research question (shown in Figure 2). Participants were selected by non-probabilistic sampling (chain-referral sampling), considering the difficulty of finding subjects representative of the experiment population. We recruited a total of 72 participants by posting announcements on social networking services and through the university community. The experiment took on average about 23 min for each participant, 6.80 USD was paid as compensation, and the experiment was conducted online. Participants' demographic information are summarized in Appendix A. In addition, a survey was conducted to understand the background of the user's movie consumption, as shown in Appendix B. This study did not include personal preferences or traits related to products. In order to exclude preferences according to age, gender, and social status, information such as genres or actors, which are characteristics of the movie, were not provided in the experiment. The development of testbed and clinical study for the progress of the experiment were conducted in the first half of 2021.

Task1. Choosing between two options

Instruction : Please choose your preferred movie from the following two suggested movies

★ 4.59 - It was a rough ride - Except for great acting...I don't think so... - It's just a little boring. - It's just a normal fantasy movie.	★ 2.89 - Great! I enjoyed it! - I watched the movie with my heart fluttering from start to finish. Anyway, I recommend it. - It's really fun.
---	--

Task2. Decision on a single option

Instruction : Please rate the presented movie in Good/Bad

★ 3.29 - It was pity that there were many cruel scenes - I was really bored - The story was boring and the running time was too long - It was rather disappointing. A trite story - It was more boring than I expected	Positive ☐ Negative ☒
---	--

Task3. Acceptable comments to star rating

Instruction : Please select all reviews that may written on the movie with the following star ratings

★ 3.08 - The idea is good, but the ending is disappointing - I was looking forward to it, but I feel sad - Laughter with family after a long time - A touching and impressive story - This movie... disgusting.. - The kid I watched with was having fun and the character was cute, too - It was so much fun - Lots of unnecessarily provocative scenes - Not as good as I thought. - Is this movie released about 10 years before? Incredible. It's good	☐ ☒ ☐ ☐ ☐ ☐ ☐ ☒ ☐
--	---

Figure 2. The interface and instruction of the testbed used in the user study; Task1 (top): Choosing between two options; Task2 (middle): Decision on a single option; Task3 (bottom): Acceptable comments to star rating.

4.2. Task Design

4.2.1. TASK1: Choosing between Two Things

Every day, consumers compare products by star rating and comments. The selection between two things is an insightful, simplified version of the actual selection process for multiple options. The experiment task was to decide between two movies, each with a review (one one-to-five-star rating, plus three comments), as shown in Figure 2-Task1. We compared Type A (a low star rating, high scoring comments) and Type B (a high star rating, low scoring comments), as shown in Figure 3 (left). Selection of Type A can be interpreted as indicating that comments have a stronger influence than star rating, and conversely, for selection of Type B, vice versa. To account for all combinations of Type A and Type B scenarios, a total of 100 sample spaces were compared, as shown in Figure 3 (right). Look again at the example in Figure 2-Task1. The Type B (left) combination consists of a SRR #5 ($4.5 < x \leq 5.0$) and comments of CS 3, and the Type A (right) combination consists of a SRR #1 ($2.5 < x \leq 3.0$) and comments of CS 5. This example scenario corresponds with the red cell shown in Figure 3 (right).

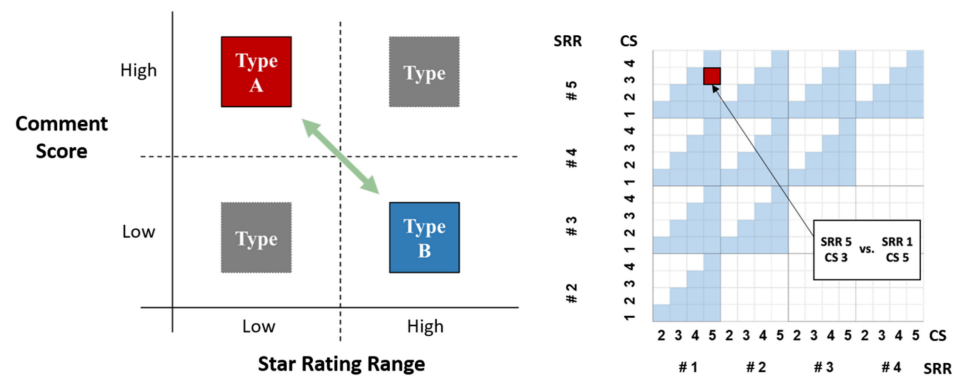


Figure 3. Subject type description (left); Sample space description (right): Red-colored cell represents the example comparison of Task1.

The Type A and Type B scenarios were randomly generated for specified ranges. For the 100 sample spaces, participants were divided evenly and randomly into two groups, and each performed selections for a half of the entire sample spaces to maintain a high level of attention. For counterbalance, a sequence of 50 selections was randomly presented to the participants. Participants chose one of two reviews/movies given this query. We computed the inter-rater reliability (I), as shown in formula (2), to measure ‘contention’ for each comparison, an absolute value calculated by the following expression.

$$I = |Na - Nb| / n \quad (2)$$

Na = number of participants who selected Type A, Nb = number of participants who selected Type B, n = total number of participants who selected either Type A or Type B ($n = 36$).

This value gets closer to zero when people have higher contention (participants are equally split between the two choices) and closer to one with lower contention (when participants completely agree on one choice). To indicate bias for the calculated inter-rater reliability, we used the positive and negative scale. When Type B is more preferred, inter-rater reliability is a positive value; when Type A is more preferred, inter-rater reliability is a negative value.

4.2.2. TASK2: Decision on a Single Product

Unlike the choice between two things, this task is designed to determine whether a product is preferred or not when given a star rating and comments for a standalone product, mirroring the like/dislike rating system. This reproduces the situation that a consumer

reads the reviews on a single product presented on a single page and takes a closer look at the product. To implement the task, we provided the participants with a one-star rating and five comments on an arbitrary movie, as shown in Figure 2-Task2. After examining the reviews, the participants had to decide if they had a desire to get it or not and click the corresponding button. For Figure 2-Task2, the SRR is #2 and the CS of the comments corresponds with 2. The sample space of the experiment consisted of all combinations of SRRs and CSs. The total sample spaces were 25 (5 possible SRRs $\times$ 5 possible CSs = 25) and each participant performed the selections for the entire sample space. For counterbalance, the sequence of 25 selections was randomly presented to the participants. During this task, we asked participants to select 'thumbs-up' or 'thumbs-down' for the movie, based on the movie's star rating and comments.

4.2.3. TASK3: Acceptable Comments to Star Rating

This task was conducted to determine the scoring range of 'acceptable' user comments for a star rating. We presented a star rating score along with 10 comments of varying scores, as shown in Figure 2-Task3. The comments presented in each task (scenario) were randomly selected, two from each possible CS (1 to 5). Therefore, 10 comments (2 comments $\times$ 5 CSs) were given for each star rating. The star rating score was randomly generated, twice for each possible SSR (#1 to #5). The entire sample space consisted of 10 total scenarios per participant due to running twice for each SRR. During this task, we asked participants to check which comments might match the star rating, with the query "Choose all comments that you think were written for the movie with the following star rating." The task provided insight into the perception of star ratings by consumers and the relationship between star ratings and comments.

5. Result

5.1. Result of TASK1

We collected 3600 selections (3600 selections = 36 participants $\times$ 2 groups $\times$ 50 selections) for 100 unique comparison scenarios. Each scenario was performed by 36 participants. Type A was selected 1745 times and Type B 1855 times, indicating that Type B was selected slightly more often. The overall trend is represented as a heatmap, shown in Figure 4. Each cell (comparison scenario) is composed of signed inter-rater reliability and indicates which type (Type A or Type B) of bias of the users. The darker the color, the more biased the selection is, and the lighter the color, the more contentious (evenly split) the selection is. The darker the red color, the more dominant Type A was for the comparison scenario; the darker the blue color, the more dominant Type B was for the comparison scenario. In general, a darker blue color appeared in the upper left quadrant and a darker red color appeared for the lower right quadrant. For example, when comparing a scenario with a SRR #5 and a CS of 3 (Type B) with a SRR #1 and a CS of 5 (Type A) (same as compared in Figure 2-Task1.), the participants chose Type B more and the inter-rater reliability at this time was +44.4%. In Figure 4, consider Type A (SRR #3, CS of 5) vs. Type B (SRR #4, CS of 1): the color is the darkest red, indicating a -100% inter-rater reliability score and that all participants chose Type A. There were 45 cells each biased for Type A and Type B. However, the median value of inter-rater reliability values was skewed for Type B (-22.2% for Type A, 33.3% for Type B). Therefore, Type B was chosen more predominantly. Both in terms of the number of choices and in terms of median values of inter-rater reliability, Type B has a slight predominance over Type A in general. Considering that Type B is a combination with a high star rating, when participants performed this task, they prioritized the star rating when making their decision. Therefore, when interpreting the results compared with the established assumption (H1), although the assumption prevailed, it was not acceptable in all cases. This result means that, when comparing two or more products and choosing one, the participants were more influenced by the quantitative review of a star rating.

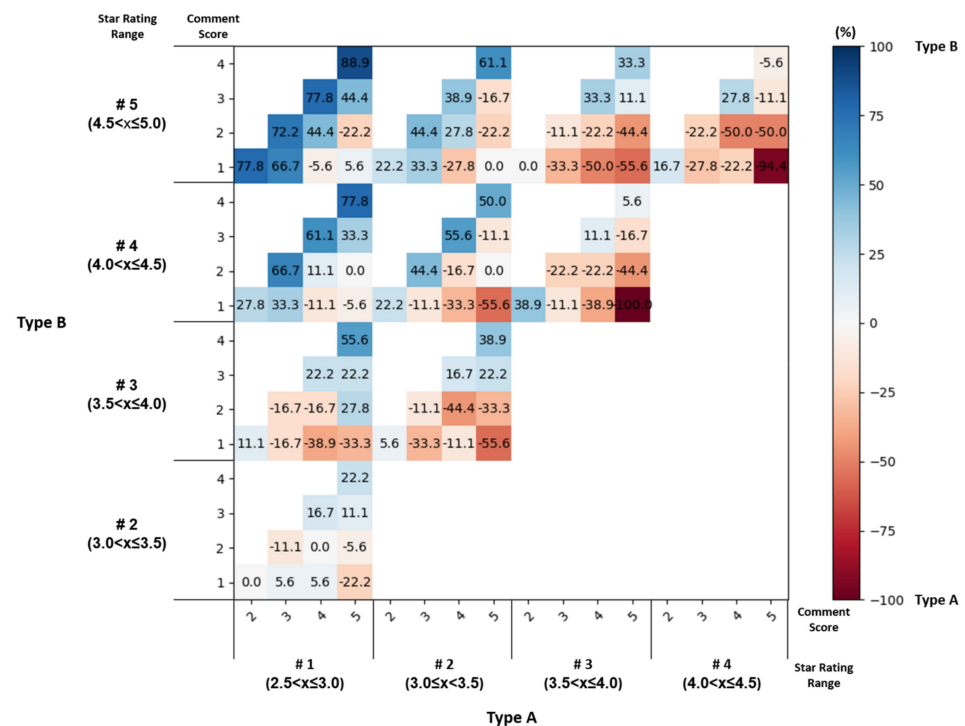


Figure 4. Heatmap of Task1 (Choosing between two things); A darker color corresponds to a higher percentage (%); Type A: red, Type B: blue.

5.2. Result of TASK2

A total of 1800 selections (1800 selections = 72 participants × 25 selections) were collected. Preference value refers to the percentage of how many thumbs-up buttons are pressed for that selection. Preference value for each cell is shown in Figure 5 (left). The darker the color of the cells in this heatmap, the higher the preference. For example, for the CS of 4 vs. SRR #2, the preference value is 69.4%. That means 69.4% of participants chose the ‘thumbs-up’ button and 30.6% of participants chose the ‘thumbs-down’ button.

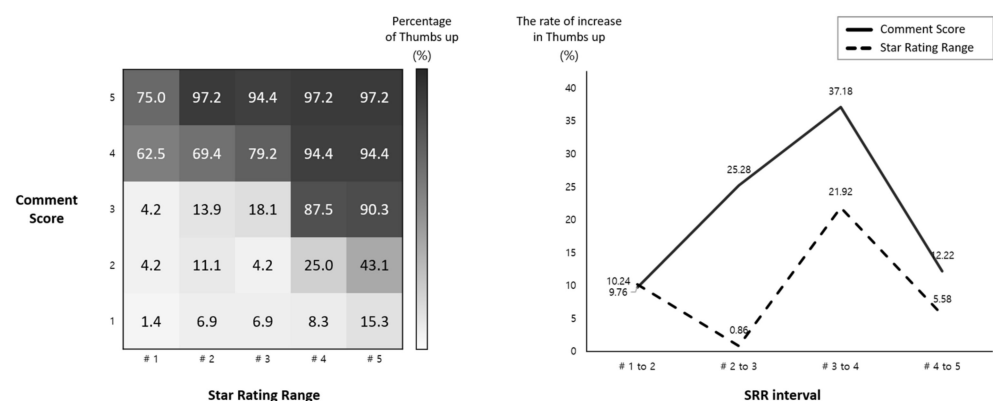


Figure 5. Heatmap of Task2 (left); Comparison of Thumbs up growth rates by SRR interval of comment score and star rating range (right).

Preference values of 50% or higher were found for: SRRs of #1, #2, #3 with comments of 3 or higher and SRRs of #4 and #5 with CSs of 3 or higher. All cells with CSs of 4 and 5 showed a preference value of 50% or higher, regardless of the star rating. For SRRs of #4 or higher with CSs of 2 or lower, the preference value was lower than 50%. Comparing these findings, CSs seem to have more influence for this type of single product evaluation.

Preference values increased more dramatically when the CS increased than when star rating increased. Figure 5 (right) shows how people's desire to get increased each time that the SRR or CS increased by one step. Preference value change was calculated by taking the difference between the averaged preference values for the designated range (#1 to #2, etc. Each average preference score value was calculated by averaging each column (SRR) or row (CS), shown on the heatmap of Figure 5 (left). The preference score change when the SRR changed from #1 to #2 was 10.24%. The preference score change when CS changed from 1 to 2 was 9.76%. Except for this range, the CS change had a higher preference score change for all ranges. Thus, for evaluating a single product, participants were influenced more by comments than star ratings, contrary to the choice between two products. For the (#)3 to (#)4 score change, the preference change amount of SRR and CS showed their highest values, 21.92% and 37.18%, respectively. For the #2 to #3 change, the SRR had its lowest change of 0.86%. This result indicates that the participants barely changed their behavior, despite the star rating changing from #2 to #3 when they considered the selection on a single product.

Comments had a greater effect than star rating for consumer preference in single product evaluation. Despite a high star rating, the movie was not 'liked' if the CS was low; despite a low star rating, if the CS was high, the movie had more chances to be 'liked.' Additionally, the preference increase due to the increase of CS was higher than the preference increase due to the increase of the star rating. As a result, the hypothesis proposed for a single selection (H2) has been supported, as the review data was found to have a greater impact on decision-making. It seems that preference (liking or disliking) of a single product is more influenced by subjective information in the form of comments, rather than star ratings.

5.3. Result of TASK3

A total of 720 data points for Task 3 were collected (72 participants $\times$ 10 scenarios). Figure 6 shows the resulting range of CSs suitable for each SRR. We conducted the Mann–Whitney U test to check whether the ranges of CSs assigned to each SRR were significantly different from each other. In the Mann–Whitney U test, there was a significant difference ($*** p < 0.001$) in all cases, except when comparing SRR #4 and #5 ($p = 0.056$). The difference between SRRs #4 and #5 was the smallest, and the difference between SRRs #2 and #3 was the largest.

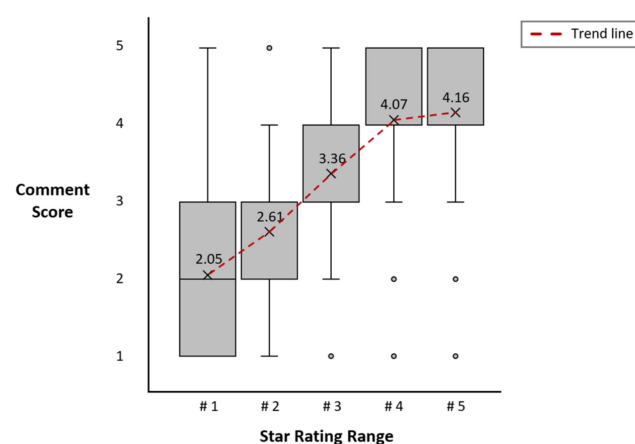


Figure 6. Acceptable comment score according to star rating range.

The average calculated CS for each SRR was 2.05, 2.61, 3.36, 4.07, and 4.16 for #1 to #5, respectively. When we substitute the representative sentiment, according to the CS that we obtained through sentiment analysis into the result of Task 3, the participants selected a negative sentiment for SRRs #1 and #2, a neutral sentiment for SRR #3, and a positive sentiment for SRRs #4 and #5.

Interestingly, when converting SRR #3 to its actual rating range, it corresponds to 3.5 to 4.0 stars for a one-to-five-star scale. The sentiment collected for this SRR #3 was neutral. Thus, despite a one-to-five-star scale indicating mathematical neutrality at the three-star rating, our sentiment analysis shows that consumers view 3.5 to 4.0 stars as a neutral rating. Thus, consumers who are planning to purchase a product view the star rating with a stricter standard than reviewers who have already made a purchase and given a review. Both SRRs #4 and #5 were viewed with positive sentiment, with no significant difference in consumer sentiment between the two. SRRs #4 and #5 are accepted by consumers in a positive way, but the difference in the positive level is not that big compared with the difference in star rating. It seems that people do not notice the difference between the high star ratings.

CSs changed most significantly when SRRs changed from #2 to #3. The average CS increased by 0.56, 0.75, 0.71, and 0.09 when the SRR changed from #1 to #5, respectively. Thus, for SRRs #2 to #3, CS increased by the largest margin. From a sentiment point of view, consumers seem to have a jump from the low 3.0 values to the high 4.0 values as the largest increase in perceived value.

5.4. Star Rating Inflation

From the results of Task 1, we determined that consumers were more influenced by star ratings when making decisions between two products. Based on Task 3, consumers were highly critical of star ratings, viewing a 3.5 to 4.0 star rating with a neutral sentiment, despite 3.0 being the mathematically neutral rating. Additionally, consumers did not demonstrate a difference in sentiment for star ratings between 4.0 and 5.0. Thus, ratings in the 4.0 to 5.0 range lose discriminating power. When the rating scale loses discrimination, consumers are less able to differentiate between good and bad products. Maintaining an exceedingly high star rating, such as one higher than 4.9, becomes a necessity for sellers who are trying to differentiate their products. As a result, sellers may create some fake reviews to increase their star rating, as revealed in the study by Luca et al. [56]. This behavior leads to a vicious cycle in which star ratings lose their discriminating power and the star rating scale becomes inflated. This cycle is detrimental to both consumers and sellers. Consumers cannot discriminate between products and sellers compete for a volatile and often inaccurate rating. To ameliorate this issue, the sentiment of consumers reading reviews and the sentiment of reviewers writing them need to be better aligned. In addition, comments can be better integrated and structured to give insight into products and guide consumers.

6. Discussion

6.1. Limitation

In this paper, we selected star ratings and comments as the main influencing factors to understand a consumer's decision-making behaviors towards movies. For this reason, we inevitably excluded information, such as genre, story, and directing or visual style, and consistently designed review datasets representing each score. Additionally, analysis without considering the preference for numerical and text review according to the user's characteristics and user's state data (e.g., gaze and biometric) can be pointed out as a limitation. For instance, it appears that the gaze data of the participant executing the task can be utilized for secondary analysis. By analyzing the area of star ratings or comments through gaze analysis, it becomes feasible to scrutinize which component garners greater attention from the user [57,58]. Furthermore, when inspecting the relevant component, the user's psychological state may also be inferred [59,60]. We analyzed representative emotions in each score range through sentiment analysis and constructed a dataset with comments that represent those emotions well. However, constructing comment data based on emotional similarity does not sufficiently eliminate all confounding factors. For example, from the review comment, "I watched a movie with my children, and they enjoyed it a lot", the participant might assume that the movie was of the animation genre. One

participant commented that he chose movies that were similar to his tastes by inferring genres. Consequently, given the availability of data that can be used to infer cinematic characteristics, it is challenging to generalize this finding to all other products.

In addition, the diversity of expressions was not accounted for. More specifically, reviews using sarcasm or expressing emotions other than the representative emotions were excluded from our dataset. This situation often occurs when some common idioms are used in comments. As has been discussed in other studies [61–63], it is very challenging to accurately read the implied intentions of the user. This is because emotions can be expressed in reverse through paradoxical expressions or irony. For example, the comment “I’m okay with not eating sweet potatoes for a while”, was mapped to a one-star rating. In Korea, the expression ‘eating sweet potatoes’ is negative. However, this comment could not be used because it was analyzed as a positive emotion of ‘caring’ through sentiment analysis. It seems possible to experiment with more abstract comments if the natural language processing technique, such as sarcasm detection [64], that analyzes various expressions beyond the superficial expression of emotions, is applied. Therefore, detecting an unusual emotion or sarcasm is a very challenging issue.

6.2. Future Work

In this study, the movie review dataset was used because of zero price variability for movies and the frequent use of emotional language in movie reviews. These characteristics reduce the variables that we need to control, but at the same time, there are a number of areas that we need further analysis on in the future. Since most consumer goods do not have the same price, it seems necessary to study how consumers behave when reviews and prices are provided at the same time in future research. Additionally, we set a comment score in this study and selected three representative sentiments for it through sentiment analysis. However, this sentiment is generated from movie review data. If sentiment analysis is applied to review data for other consumer goods, it is expected that different representative sentiments will be selected even with the same comment score. To solve this problem, it seems necessary to extend the sentiment analysis of the movie domain to other domains. Transfer learning techniques can be used to extend domains [65]. After collecting review data of the domain to be applied, additional training is performed on the pre-trained model. This enables sentiment analysis for other review domains. It seems possible to find sentiment expressions that are similarly used for each domain through word embedding [66]. Through this, sentiment analysis can be applied to multiple domains with only one model without transfer learning.

7. Conclusions

This paper aimed to identify how people perceive and behave against reviews of star ratings paired with comments. A choice between two products, single product evaluation, and the relationship between comments and star ratings were investigated by user behavioral and sentiment analysis. First, we confirmed the usability of movie review data through exploratory data analysis and designed three viable user studies with 72 participants. Consumers value star ratings and comments differently, based on the scenario (experiment 1 and 2) in which they are provided. For the first study, a choice problem between two products, star rating had more influence than comments when choosing a more preferred movie. Yet, if movie ratings and comments are presented for a single movie, a ‘like’ or ‘dislike’ decision was more influenced by comments. Additionally, consumers were strict when assigning positive sentiment towards star ratings when viewing them as consumers, rather than when reviewing products themselves. Our study aims to contribute insights into identifying consumer perception patterns concerning star ratings and comment reviews. These findings could potentially aid in the development of more effective review structures.

Author Contributions: Conceptualization, Y.-G.N. and J.-H.H.; methodology, Y.-G.N.; software, Y.-G.N.; validation, Y.-G.N. and J.J.; formal analysis, Y.-G.N.; writing—original draft preparation, Y.-G.N.; writing—review and editing, Y.-G.N. and J.J.; visualization, Y.-G.N. and J.J.; supervision, J.-H.H.; project administration, J.-H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (2021R1A4A1030075).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Demographic Information.

		N (n = 72)	Percentage (%)
Gender	Male	42	58.33
	Female	30	41.67
Age	18–23	17	23.61
	24–29	23	31.94
	30–35	13	18.06
	36–41	10	13.89
	42+	9	12.50
Occupation	High school student	4	5.56
	University student	17	23.61
	Graduate student	23	31.94
	Office worker	12	16.67
	Practitioner	9	12.50
	Freelance	4	5.56
	Others	3	4.17
Major	Engineering	34	47.22
	Nature	13	18.06
	Liberal arts	6	8.33
	Education	11	15.28
	Society	3	4.167
	Entertainments and sports	1	1.39
	others	34	47.22

Appendix B

Table A2. Background of Participants' Movie Consumption.

		N (n = 72)	Percentage (%)
Frequency of watching movies	4 or more per month	14	19.44
	2 or more per month	16	22.22
	1 or more per month	20	27.78
	1 movie 2 month	19	26.39
	Rarely watch	3	4.17

Table A2. Cont.

		N (n = 72)	Percentage (%)
Major genre of watching (Multiple choice)	SF/Fantasy/Adventure	43	59.72
	Thriller/Mystery	34	47.22
	Romantic comedy	29	40.28
	Action	27	37.50
	Drama	24	33.33
	Comedy	21	29.17
	Animation	18	25.00
	Regardless of genre	18	25.00
	Others	3	4.167
Criteria for selecting movies	Story	62	86.11
	Evaluation of online rating	43	59.72
	Genre	41	56.94
	Recommendation of acquaintance	33	45.83
	Companion's taste	24	33.33
	Actor	22	30.56
	Director	21	29.17
	Film festival entry or award	19	26.39
	Others	4	5.56
How to get information (Multiple choice)	Movie information tv program	48	66.67
	OTT service	31	43.06
	Customer review information	30	41.67
	Social network service (SNS)	23	31.94
	Advertisement	21	29.17
	others	19	26.39

References

- Chen, A.; Lu, Y.; Wang, B. Customers' purchase decision-making process in social commerce: A social learning perspective. *Int. J. Inf. Manag.* **2017**, *37*, 627–638. [\[CrossRef\]](#)
- Lee, J.; Park, D.-H.; Han, I. The effect of negative online consumer reviews on product attitude: An information processing view. *Electron. Commer. Res. Appl.* **2008**, *7*, 341–352. [\[CrossRef\]](#)
- Blazevic, V.; Hammedi, W.; Garnefeld, I.; Rust, R.T.; Keiningham, T.; Andreassen, T.W.; Donthu, N.; Carl, W. Beyond traditional word-of-mouth: An expanded model of customer-driven influence. *J. Serv. Manag.* **2013**, *24*, 294–313. [\[CrossRef\]](#)
- Dellarocas, C.; Zhang, X.M.; Awad, N.F. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interact. Mark.* **2007**, *21*, 23–45. [\[CrossRef\]](#)
- Masłowska, E.; Malthouse, E.C.; Viswanathan, V. Do customer reviews drive purchase decisions? The moderating roles of review exposure and price. *Decis. Support Syst.* **2017**, *98*, 1–9. [\[CrossRef\]](#)
- Tsao, W.-C. Which type of online review is more persuasive? The influence of consumer reviews and critic ratings on moviegoers. *Electron. Commer. Res.* **2014**, *14*, 559–583. [\[CrossRef\]](#)
- Bae, S.; Lee, T. Product type and consumers' perception of online consumer reviews. *Electron. Mark.* **2011**, *21*, 255–266. [\[CrossRef\]](#)
- Donthu, N.; Kumar, S.; Pandey, N.; Pandey, N.; Mishra, A. Mapping the electronic word-of-mouth (eWOM) research: A systematic review and bibliometric analysis. *J. Bus. Res.* **2021**, *135*, 758–773. [\[CrossRef\]](#)
- Ismagilova, E.; Rana, N.P.; Slade, E.L.; Dwivedi, Y.K. A meta-analysis of the factors affecting eWOM providing behaviour. *Eur. J. Mark.* **2021**, *55*, 1067–1102. [\[CrossRef\]](#)
- Dessart, L.; Pitardi, V. How stories generate consumer engagement: An exploratory study. *J. Bus. Res.* **2019**, *104*, 183–195. [\[CrossRef\]](#)
- Oestreicher-Singer, G.; Sundararajan, A. The visible hand? Demand effects of recommendation networks in electronic markets. *Manag. Sci.* **2012**, *58*, 1963–1981. [\[CrossRef\]](#)
- Ye, Q.; Law, R.; Gu, B. The impact of online user reviews on hotel room sales. *Int. J. Hosp. Manag.* **2009**, *28*, 180–182. [\[CrossRef\]](#)
- Yang, Y.; Pan, B.; Song, H. Predicting hotel demand using destination marketing organization's web traffic data. *J. Travel Res.* **2014**, *53*, 433–447. [\[CrossRef\]](#)
- Zhou, L.; Ye, S.; Pearce, P.L.; Wu, M.-Y. Refreshing hotel satisfaction studies by reconfiguring customer review data. *Int. J. Hosp. Manag.* **2014**, *38*, 1–10. [\[CrossRef\]](#)
- Dong, J.; Li, H.; Zhang, X. Classification of customer satisfaction attributes: An application of online hotel review analysis. In Proceedings of the Digital Services and Information Intelligence: 13th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, Sanya, China, 28–30 November 2014; pp. 238–250.

16. Evangelopoulos, N. Citing Taylor: Tracing Taylorism's technical and sociotechnical duality through latent semantic analysis. *J. Bus. Manag.* **2011**, *17*, 57–74.
17. Visinescu, L.L.; Evangelopoulos, N. Orthogonal rotations in latent semantic analysis: An empirical study. *Decis. Support. Syst.* **2014**, *62*, 131–143. [\[CrossRef\]](#)
18. Kitsios, F.; Kamariotou, M.; Karanikolas, P.; Grigoroudis, E. Digital marketing platforms and customer satisfaction: Identifying eWOM using big data and text mining. *Appl. sci.* **2021**, *11*, 8032. [\[CrossRef\]](#)
19. Gartrell, M.; Alanezi, K.; Tian, L.; Han, R.; Lv, Q.; Mishra, S. SocialDining: Design and analysis of a group recommendation application in a mobile context. *Comput. Sci. Tech. Rep.* **2014**, *10*, 1034.
20. Vanetti, M.; Binaghi, E.; Carminati, B.; Carullo, M.; Ferrari, E. Content-based filtering in on-line social networks. In Proceedings of the International Workshop on Privacy and Security Issues in Data Mining and Machine Learning, Barcelona, Spain, 24 September 2010; pp. 127–140.
21. Srifi, M.; Oussous, A.; Ait Lahcen, A.; Mouline, S. Recommender systems based on collaborative filtering using review texts—A survey. *Information* **2020**, *11*, 317. [\[CrossRef\]](#)
22. Sánchez, C.N.; Domínguez-Soberanes, J.; Arreola, A.; Graff, M. Recommendation System for a Delivery Food Application Based on Number of Orders. *Appl. Sci.* **2023**, *13*, 2299. [\[CrossRef\]](#)
23. Roy, A.; Banerjee, S.; Sarkar, M.; Darwish, A.; Elhoseny, M.; Hassanien, A.E. Exploring New Vista of intelligent collaborative filtering: A restaurant recommendation paradigm. *J. Comput. Sci.* **2018**, *27*, 168–182. [\[CrossRef\]](#)
24. Da'u, A.; Salim, N.; Idris, R. An adaptive deep learning method for item recommendation system. *Knowl.-Based Syst.* **2021**, *213*, 106681. [\[CrossRef\]](#)
25. Wu, H.; Zhang, Z.; Yue, K.; Zhang, B.; He, J.; Sun, L. Dual-regularized matrix factorization with deep neural networks for recommender systems. *Knowl.-Based Syst.* **2018**, *145*, 46–58. [\[CrossRef\]](#)
26. Karimi, S.; Wang, F. Online review helpfulness: Impact of reviewer profile image. *Decis. Support Syst.* **2017**, *96*, 39–48. [\[CrossRef\]](#)
27. Rietsche, R.; Frei, D.; Stöckli, E.; Söllner, M. Not all Reviews are Equal—a Literature Review on Online Review Helpfulness. In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm/Uppsala, Sweden, 8–14 June 2019; p. 58.
28. Otterbacher, J. 'Helpfulness' in online communities: A measure of message quality. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09), Boston, MA, USA, 4–9 April 2009; pp. 955–964.
29. Huang, A.H.; Yen, D.C. Predicting the helpfulness of online reviews—A replication. *Int. J. Hum.-Comput. Interact.* **2013**, *29*, 129–138. [\[CrossRef\]](#)
30. Huang, A.H.; Chen, K.; Yen, D.C.; Tran, T.P. A study of factors that contribute to online review helpfulness. *Comput. Hum. Behav.* **2015**, *48*, 17–27. [\[CrossRef\]](#)
31. Salehan, M.; Kim, D.J. Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decis. Support Syst.* **2016**, *81*, 30–40. [\[CrossRef\]](#)
32. Chua, A.Y.; Banerjee, S. Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Comput. Hum. Behav.* **2016**, *54*, 547–554. [\[CrossRef\]](#)
33. Singh, J.P.; Irani, S.; Rana, N.P.; Dwivedi, Y.K.; Saumya, S.; Roy, P.K. Predicting the “helpfulness” of online consumer reviews. *J. Bus. Res.* **2017**, *70*, 346–355. [\[CrossRef\]](#)
34. Ma, Y.; Xiang, Z.; Du, Q.; Fan, W. Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. *Int. J. Hosp. Manag.* **2018**, *71*, 120–131. [\[CrossRef\]](#)
35. Chen, C.; Qiu, M.; Yang, Y.; Zhou, J.; Huang, J.; Li, X.; Bao, F.S. Multi-domain gated CNN for review helpfulness prediction. In Proceedings of the WWW '19: The World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2630–2636.
36. Mitra, S.; Jenamani, M. Helpfulness of online consumer reviews: A multi-perspective approach. *Inf. Process. Manag.* **2021**, *58*, 102538. [\[CrossRef\]](#)
37. Filieri, R. What makes an online consumer review trustworthy? *Ann. Tour. Res.* **2016**, *58*, 46–64. [\[CrossRef\]](#)
38. Filieri, R.; Raguseo, E.; Vitari, C. Extremely negative ratings and online consumer review helpfulness: The moderating role of product quality signals. *J. Travel Res.* **2021**, *60*, 699–717. [\[CrossRef\]](#)
39. Chen, P.-Y.; Wu, S.-Y.; Yoon, J. The impact of online recommendations and consumer feedback on sales. In Proceedings of the ICIS 2004, Washington, DC, USA, 12–15 December 2004; p. 58.
40. Talton, J.O., III; Dusad, K.; Koiliaris, K.; Kumar, R.S. How do people sort by ratings? In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '19), Glasgow, UK, 4–9 May 2019; pp. 1–10.
41. Yatani, K.; Novati, M.; Trusty, A.; Truong, K.N. Review spotlight: A user interface for summarizing user-generated reviews using adjective-noun word pairs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11), Vancouver, BC, Canada, 7–13 May 2011; pp. 1541–1550.
42. Chen, C.-W. Five-star or thumbs-up? The influence of rating system types on users' perceptions of information quality, cognitive effort, enjoyment and continuance intention. *Internet Res.* **2017**, *27*, 478–494. [\[CrossRef\]](#)
43. Maharani, W.; Widiantoro, D.H.; Khodra, M.L. Discovering Users' Perceptions on Rating Visualizations. In Proceedings of the 2nd International Conference in HCI and UX Indonesia 2016 (CHIuXiD '16), Jakarta, Indonesia, 13–15 April 2016; pp. 31–38.
44. Zhao, Y.; Xu, X.; Wang, M. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *Int. J. Hosp. Manag.* **2019**, *76*, 111–121. [\[CrossRef\]](#)

45. Hu, Y.; Kim, H.J. Positive and negative eWOM motivations and hotel customers' eWOM behavior: Does personality matter? *Int. J. Hosp. Manag.* **2018**, *75*, 27–37. [\[CrossRef\]](#)
46. Lichtenstein, D.R.; Burton, S. The relationship between perceived and objective price-quality. *J. Mark. Res.* **1989**, *26*, 429–443. [\[CrossRef\]](#)
47. Anderson, E.W.; Sullivan, M.W. The antecedents and consequences of customer satisfaction for firms. *Mark. Sci.* **1993**, *12*, 125–143. [\[CrossRef\]](#)
48. Fornell, C.; Johnson, M.D.; Anderson, E.W.; Cha, J.; Bryant, B.E. The American customer satisfaction index: Nature, purpose, and findings. *J. Mark.* **1996**, *60*, 7–18. [\[CrossRef\]](#)
49. Wang, J.-N.; Du, J.; Chiu, Y.-L.; Li, J. Dynamic effects of customer experience levels on durable product satisfaction: Price and popularity moderation. *Electron. Commer. Res. Appl.* **2018**, *28*, 16–29. [\[CrossRef\]](#)
50. Turney, P.D. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics: ACL '02, Stroudsburg, PA, USA, 7–12 July 2002; pp. 417–424.
51. Ptaszynski, M.; Dybala, P.; Rzepka, R.; Araki, K. An automatic evaluation method for conversational agents based on affect-as-information theory. *J. Jpn. Soc. Fuzzy Theory Intell. Inform.* **2010**, *22*, 73–89. [\[CrossRef\]](#)
52. Wankhade, M.; Rao, A.C.S.; Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif. Intell. Rev.* **2022**, *55*, 5731–5780. [\[CrossRef\]](#)
53. Na, J.C.; Thura Thet, T.; Khoo, C.S. Comparing sentiment expression in movie reviews from four online genres. *Online Inf. Rev.* **2010**, *34*, 317–338. [\[CrossRef\]](#)
54. Demszky, D.; Movshovitz-Attias, D.; Ko, J.; Cowen, A.; Nemade, G.; Ravi, S. GoEmotions: A dataset of fine-grained emotions. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 4040–4054.
55. KoELECTRA: Pretrained ELECTRA Model for Korean Home Page. Available online: <https://github.com/monologg/KoELECTRA> (accessed on 18 March 2023).
56. Luca, M.; Zervas, G. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Manag. Sci.* **2016**, *62*, 3412–3427. [\[CrossRef\]](#)
57. Noh, Y.-G.; Hong, J.-H. Designing Reenacted Chatbots to Enhance Museum Experience. *Appl. Sci.* **2021**, *11*, 7420. [\[CrossRef\]](#)
58. Kong, H.-K.; Zhu, W.; Liu, Z.; Karahalios, K. Understanding Visual Cues in Visualizations Accompanied by Audio Narrations. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; p. 50.
59. Choi, Y.; Kim, J.; Hong, J.-H. Immersion Measurement in Watching Videos Using Eye-tracking Data. *IEEE Trans. Affect. Comput.* **2022**, *13*, 1759–1770. [\[CrossRef\]](#)
60. Soleymani, M.; Mortillaro, M. Behavioral and Physiological Responses to Visual Interest and Appraisals: Multimodal Analysis and Automatic Recognition. *Front. ICT* **2018**, *5*, 17. [\[CrossRef\]](#)
61. Farias, D.H.; Rosso, P. Irony, sarcasm, and sentiment analysis. In *Sentiment Analysis in Social Networks*; Pozzi, F., Fersini, E., Messina, E., Liu, B., Eds.; Morgan Kaufmann: Burlington, MA, USA, 2017; pp. 113–128.
62. Agrawal, A.; Hamling, T. Sentiment analysis of tweets to gain insights into the 2016 US election. *Columbia Undergraduate Sci. J.* **2017**, *11*, 34–42. [\[CrossRef\]](#)
63. Maynard, D.; Greenwood, M. Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, 26–31 May 2014; pp. 4238–4243.
64. Verma, P.; Shukla, N.; Shukla, A. Techniques of sarcasm detection: A review. In Proceedings of the 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 4–5 March 2021; pp. 968–972.
65. Yoshida, Y.; Hirao, T.; Iwata, T.; Nagata, M.; Matsumoto, Y. Transfer learning for multiple-domain sentiment analysis—Identifying domain dependent/independent word polarity. In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI-11), San Francisco, CA, USA, 7–11 August 2011; pp. 1286–1291.
66. Giatoglou, M.; Vozalis, M.G.; Diamantaras, K.; Vakali, A.; Sarigiannidis, G.; Chatzisavvas, K.C. Sentiment analysis leveraging emotions and word embeddings. *Expert Syst. Appl.* **2017**, *69*, 214–224. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.