Article

Are eBay's Feedback Ratings Consistent with the Sentiments Embedded in Textual Comments? An Empirical Study

Xubo Zhang [1]**, Yanbin Tu** [2,*] **and Ke Zhong** [3]

1 School of Economics, Wuhan Polytechnic University, Wuhan 430048, China; zhangxubo@whpu.edu.cn
2 Rockwell School of Business, Robert Morris University, Pittsburgh, PA 15108, USA
3 College of Business, Central Washington University, Lynnwood, WA 98036, USA; zhongk@cwu.edu
* Correspondence: tu@rmu.edu

Abstract: eBay's feedback rating system is currently widely used. In this study, we examine if eBay's feedback rating types (+, 0, −) are consistent with the sentiments reflected in the textual comments posted by buyers. Using the datasets collected from eBay, we test the hypotheses associated with the research questions at three levels: individual, group, and total. Overall, the types of feedback ratings are consistent with the sentiments embedded in the textual comments. However, there are some issues with eBay's current feedback rating system: (1) at the individual level, the correlation coefficient between the ratings and the comments' sentiments is low at 0.4311 (<0.5). While the three types of ratings are symmetric, like (−1, 0, +1), buyers' textual comments have asymmetric distributions of sentiments among these three types. The three simple feedback ratings (+, 0, −) are not fully aligned with the sentiments revealed in the textual comments posted by buyers. We propose expanding the current three ratings into five ratings such as (−2, −1, 0, +1, +2), which might help remedy the issue. We contribute to the literature by tapping into this less-studied area vital to improving the online marketplace's efficiency.

Keywords: e-commerce; feedback rating system; sentiment analysis; system improvement

1. Introduction

The biggest online auction platform, eBay, attracts millions of buyers and sellers to trade various goods. According to the quarterly report of eBay in Quarter 4, 2022, there were 134 million active eBay users by 31 December 2022, and the eBay marketplace generated a gross merchandise volume (the total value of all successfully closed transactions) of USD 18.2 billion in Quarter 4 of 2022 (Source: https://investors.ebayinc.com/overview/default.aspx, accessed on 10 December 2023). The business success of eBay can be attributed to many factors, one of which is its feedback rating system (also called the reputation system) [1,2]. eBay's feedback rating system is the most critical and successful mechanism to boost online users' trust [3–6]. A winning bidder/buyer can rate the purchase experience as positive, negative, or neutral. A positive feedback rating (+) indicates that buyers are satisfied with the auction transaction. A neutral feedback rating (0) indicates that buyers are not fully satisfied and have some reservations. A negative feedback rating (−) shows that buyers are strongly dissatisfied with their purchase experiences. Buyers can only rate an order as positive, neutral, or negative. Otherwise, they can leave no feedback. For an eBay user, feedback ratings are aggregated into a summary score displayed next to his/her ID. Over time, an eBay user develops a feedback profile, or reputation, based on other people's textual comments and ratings. The feedback score is an essential piece of a feedback profile. The higher an eBay user's feedback score is, the more likely it is that other users will want to conduct business with them [3]. Buyers also can write brief textual comments or explanations with their posted feedback ratings. The comments typically provide reasons for a rating and reflect buyers' sentiments, ranging from appreciation and

pride to frustration and anger. Once posted, all these comments, regardless of positive, negative, or neutral feedback ratings, are publicly available to eBay users. Potential eBay bidders or buyers could use this information to judge sellers' reputations and decide if they should bid for those auction listings. According to prior research [2,7], a good reputation could bring in a price premium to products listed at online auction marketplaces, while negative or neutral feedback impedes sellers from gaining customer trust and even forces new sellers to exit from the online auction markets at eBay.

Studies on online auctions in the literature can be grouped into four streams: (1) online auction transactions and their determinants; (2) online auction mechanisms and designs; (3) online trust and information asymmetry; and (4) online reputation and its role [8]. For the online feedback/reputation system, many studies focus on buyers' ratings without addressing the textual comments associated with these ratings. For example, some studies show the validation of eBay rating profiles to build trust among online auction users [9] or the potential problems associated with rating frauds [10]. To our best knowledge, few studies examined buyers' textual comments and embedded sentiments associated with their feedback ratings and checked if the sentiments are consistent with the ratings. This study aims to fill in the gap in the literature. We mainly try to answer the following research questions by using text mining and sentiment analytical techniques:

> *Are the types of eBay feedback ratings (positive, neutral, and positive) consistent with the sentiments (positive, neutral, and positive) reflected in the textual comments posted by buyers? How to improve eBay feedback rating system from the perspective of sentiment consistency?*

Investigating the answers to these questions can help validate eBay's feedback system, figure out its potential issues, and shed light on improving it. All these efforts could help boost the online auction market's sustainability and efficiency.

The paper is organized as follows: we provide a literature review after the introduction. Next, we outline the research methodology and analytical outcomes. We also discuss our findings and their managerial implications. Finally, we conclude the study and discuss the limitations and future studies.

2. Literature Review

One big challenge in online shopping is information asymmetry between buyers and sellers [11]. Many platforms and vendors use buying recommendation systems or WOM (word-of-mouth) reputation systems to facilitate information asymmetry, build trust, and help buyers choose the right products [12–14]. At eBay, its feedback system plays an important role in facilitating trust between buyers and sellers on the online marketplace. Prior studies explored the impact of eBay's feedback system on buyer and seller behaviors and concluded that eBay's online feedback system was generally effective in improving trust between buyers and sellers [3,5,15,16]. These studies suggested that positive feedback ratings correlated with higher selling prices, increased sales volume, attracted new customers, and retained existing ones. The effectiveness of eBay's feedback rating system can also be seen through the negative impact on sellers of unfavorable (such as negative or neutral) ratings. Negative feedback, on the other hand, can significantly damage sellers' reputations and deter potential buyers [17–19]. Ref. [20] indicated that negative feedback was associated with lower final auction prices on eBay, and potential buyers were more sensitive to negative feedback when buying used or refurbished products. Ref. [21] indicated that sellers' selling reputation was positively associated with the probability of sale and selling prices. Ref. [1] examined the factors that affect the long-term survival rate of auction ventures on eBay. These works showed that sellers' reputations with favorable ratings are vital to their long-term survival.

Although eBay's feedback system works well overall, challenges and issues exist, including retaliation and subsequent revoking, reciprocation, changing of online identity, cheap positive feedback through purchases, and "market for feedback" with the sole purpose of manufacturing positive feedback [4,6,22–24]. Under these drawbacks and issues

existing in the feedback system, some researchers explored potential improvements. For example, Ref. [25] proposed solutions to user collusion and short-lived online identity problems; Ref. [26] proposed a new product impact evaluation method by utilizing customer feedback data following user purchases.

While the above studies focused on the feedback ratings, the sentiments embedded in textual comments associated with the ratings have not been fully analyzed in the literature. Sentiment analysis constitutes retrieving sentiments from text comments posted by buyers. Sentiment analysis (also called opinion mining) is a text-mining technique to process language-related content. Refs. [27–30] pointed out that sentiment analysis could help understand consumers and make effective marketing strategies and decision-making policies. Sentiment analysis is an important tool for tracing customers' or investors' sentiments, significantly impacting product sales and stock markets [31]. For example, Refs. [32,33] conducted social media analytics using Twitter data on cruise travel; Ref. [34] used sentiment changes to predict movie box office revenues; Ref. [35] used Twitter sentiment analysis to capture visitors' sentiments on resorts.

Regarding some existing studies on textual comments at eBay, Ref. [14] created a survey on how the textual reviews associated with positive, negative, or neutral feedback ratings were utilized to drive users' interest. Ref. [36] used sentiment analysis to detect the polarity for positive or negative reviews. Textual sentiment analysis can also be used to tune up the numerical ratings. For example, by using various online textual and numerical reviews, Ref. [37] developed a system that used aspect-based sentiment analysis to generate more accurate and reliable reputation scores. Ref. [28] used NLP techniques, including sentiment analysis, to detect and remove fake reviews on Amazon so that more trustworthy ratings were available. However, in the literature, few studies investigate if the types of feedback ratings are consistent with the sentiments revealed from the textual comments posted by buyers or address possible improvements to the feedback system from the perspective of sentiment consistency. This study aims to contribute to the literature by studying this consistency and seeks to find a new avenue which can help improve eBay's feedback ratings system.

3. Research Methodology and Analytical Outcomes

As we mentioned previously, the feedback ratings at eBay are classified as positive (+), neutral (0), and negative (−). Since eBay adds 1 to a seller's profile if they receive a positive rating, adds nothing if they receive a neutral rating, and deducts 1 if they receive a negative rating, understandably, people take a negative rating as −1, a neutral rating as 0, and a positive rating as +1. Thus, this paper uses −1, 0, and +1 to represent the negative, neutral, and positive ratings. Whatever specific type of feedback rating is used, it should represent buyers' evaluations and sentiments (satisfaction or frustrations).

There are two ways to conduct sentiment analysis: corpus-based and lexicon-based [38]. Corpus-based analysis uses the texts/corpora representing a paragraph to determine the sentiment types (negative, neutral, or positive) derived from text contents. A lexicon-based method uses an existing lexicon or dictionary to determine text contents' sentiment types (negative, neutral, or positive). Ref. [39] showed that the two approaches produced similar accuracy in some cases. There are various lexicon methods available for sentiment analysis at present. In this study, we chose to use four lexicon methods: SentimentR, Bing, Loughran, and NRC, since they are widely used in both academia and industries.

Sentiment analysis presents challenges such as domain and event dependency [40]. One word might have opposite sentiment implications under different contexts. Ref. [41] proposed a domain-adapted sentiment classification (DA-SC) technique to reduce the domain-dependency problem. As the algorithm evolves, sentiment analytical methods such as SentimentR can better handle these challenges. Moreover, with a large sample size, we can be more confident with the reliability of sentiment outputs to reflect the true feelings and moods of content generators. We believe that sentiment analysis is suitable to reveal buyers' feelings by mining the textual comments associated with their ratings.

The research framework of this study is depicted in Figure 1. Under a well-designed feedback rating system, the positive, neutral, or negative ratings should effectively represent buyers' corresponding positive, neutral, or negative sentiments. Such sentiments can be traced from the textual comments of buyers. In this study, we will study the consistency between ratings and sentiments in three dimensions: (1) check if signs (+, 0, −) of ratings and sentiments are the same for each rating type; (2) measure the correlation coefficient between ratings and sentiments; (3) measure the distances among three types of ratings and sentiments, respectively, and see if distance symmetry exists for both ratings and sentiments.
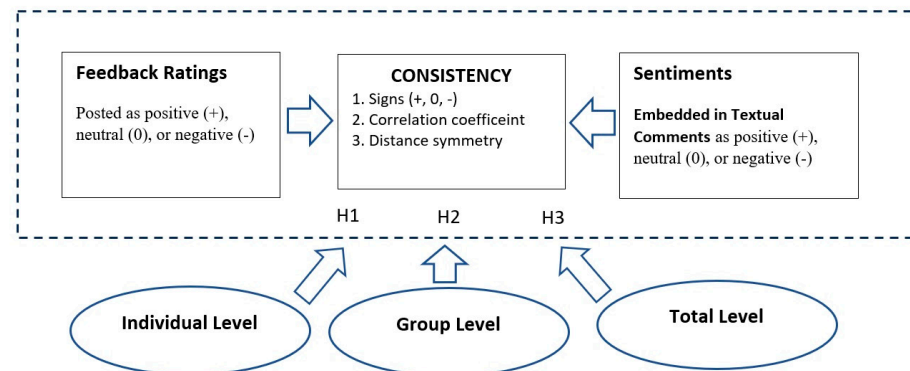

Figure 1. Research framework.

Theoretically, an effective eBay feedback rating system should maintain the consistency between the types of ratings and the sentiments embedded in the textual comments. Inconsistency might lead to users' confusion and erode their confidence, and, in the long run, the feedback rating system would become less useful. As mentioned above, our task is to test if the types of feedback ratings are consistent with the sentiments revealed in the comments posted by buyers. To better address this topic and obtain robust results, we will conduct tests at three levels: (1) the individual level: we examine whether each rating is consistent with the buyer's sentiment in the associated comments. The individual level is the most important to test the consistency between the types of ratings and sentiments. (2) The group level: we divide the dataset into 10 groups for each feedback rating. For example, a total of 1000 positive ratings will be randomly divided into 10 groups, and each group has 100 ratings. Then, we will test if the average sentiment values embedded in the comments are positive for these 10 positive groups. The group level helps us further test consistency with various lexicon methods. (3) The total level: using the total dataset, we test if the ratings are consistent with the sentiment embedded in the comments across various lexicon methods. This total level helps us understand if any differences exist among these lexicons.

Previous studies showed that eBay's feedback system did help reduce information asymmetry and boost trust between buyers and sellers [3,5,15,16]. The effectiveness of the feedback system implies that buyers' ratings represent their true evaluations and sentiments. Buyers could correctly express their assessments of their purchasing experience in both ratings (such as +, 0, −), and sentiments (such as positive, neutral, and negative) in words. Even though the numeric ratings cannot fully capture the polarity of information in the textual reviews/comments [42], it is generally understood that ratings are a numeric representation of textual comments and that the sentiments of ratings and textual comments are consistent [40]. Text sentiments significantly influence how users perceive the interestingness and trustworthiness of ratings [43]. More important, an interaction between textual comments and ratings would affect a rating's trustworthiness [40]. Because, in the literature, it is widely concluded that eBay's feedback system is working well (as we introduced in the literature review section), we would assume that the sentiments from textual comments are consistent with the associated ratings. Therefore, we would like to list the following hypotheses:

Hypothesis 1. *At the individual level, the sign of feedback ratings in each type (positive, neutral, and negative) is the same as that of sentiments embedded in associated textual comments.*

Hypothesis 2. *At the group level, the sign of feedback ratings in each type (positive, neutral, and negative) is the same as that of sentiments embedded in associated textual comments.*

Hypothesis 3. *At the total level, the sign of feedback ratings in each type (positive, neutral, and negative) is the same as that of sentiments embedded in associated textual comments.*

To test the above hypotheses, we will check if the sentiment score associated with negative feedback ratings is negative, if the sentiment score associated with neutral feedback ratings is equal to zero, and if the sentiment score associated with positive feedback ratings is also positive.

We collected auction data related to video game consoles from eBay. Video game console sales are a big business industry. According to [44], video game console sales reached USD 60 billion in 2021 and would reach USD 61 billion in 2022. The video game console market has three key brands: Sony PlayStation, Microsoft Xbox, and Nintendo. We use online auctions for one industry, which could reduce the noise in ratings and sentiments due to different product categories. As examples in Table 1, the comments could provide information regarding the reasons behind buyers' feedback ratings and reveal their complaints, praises, wishes, pain points, suggestions, and warnings to potential buyers. Therefore, sentiment analysis on these comments will be helpful for eBay users to gain a complete picture of buyers' experiences. Please see the data summarized in Table 1.

R (4.2.2) and RStudio (2022.07.2+576.pro12) software applications were used to conduct data analysis in this study. At the individual level, we used the R package, SentimentR, to calculate sentiment scores for the textual comments associated with each rating. SentimentR is a lexicon-based sentiment analysis package that can quickly calculate text polarity sentiment at the sentence level. At the group level, we divided the dataset into 10 groups for each type of feedback rating. Then, we used three lexicon methods—Bing, Loughran, and NRC—to analyze the buyers' comments in each group. Each lexicon method checks all the words in the comments and identifies sentimental words as negative, positive, uncertainty, anger, and disgust. Then, it summarizes each category and produces sentiment scores equal to the total number of positive words minus the total number of negative words. All the lexicon methods ignore the words which do not reflect the sentiments of human beings. These three lexicon methods are widely used and well known in text sentiment mining applications. Three lexicon methods were used together because each has advantages and disadvantages. Furthermore, combining them might help us take full advantage of lexicon sentiment analysis. At the total level, we also used Bing, Loughran, and NRC to mine the dataset for each type of feedback ratings. We will test whether the sentiment scores are the same under these three lexicon methods, which will help us check if these lexicons generate similar results.

Table 1. Data summary and examples.

#	# of Auction Listings for Sony PlayStation, Xbox, and Nintendo Consoles	26,465	100%
1	# of negative (−) feedback posted by buyers	1063	4%
2	# of neutral (0) feedback posted by buyers	481	1.8%
3	# of positive (+) feedback posted by buyers	24,921	94.2%

Table 1. *Cont.*

|	Examples of Buyers' Feedback Ratings and Comments	
Rating	Comments	Buyer ID
−	AFTER 3 WKS OF LIES ABOUT SENDING SYSTEM I FILED A DISPUTE---- BUYERS BEWARE---	alinick
−	i still have not received my purchase!! I want it sent to me,	shammol
−	will not answer emails or phone calls. paid 6/24 it is now 7/6, no item yet	schneidermrs
−	FRAUD! NEVER SENT ITEM, NO CONTACT..AVOID! *** TOOK MY $$!! LOSER!!!!!!!!!!!!!!!!!!	kcscollections
−	SENT ME A BROKEN UNIT, VERY UNHAPPY WITH THIS TRANSACTION!	rocubs1m0j
−	Seller Bid on own auction, Items not correct, Buyer complaint filed	daniellesells
0	Received what I ordered, but shipping delays were a problem-	ltjboy
0	Refunded my payment. . . Item destroyed when friend met with an accident. . .!!?-	abbey486
0	4 weeks to come in and 2 of the 3 games were missing!-	sanchez101786
0	over 2 wks. late on delivery-	bammer1283
0	Great product. Exactly as stated. Problem with shipping costs, but seems honest-	schnear25
0	took a bit of time to send, poorly packed, but it works so i cant complain 2 much-	djvlee
+	Smooth sale. . .Very tentative ..cant beat that	magnetikdemise
+	Good to deal with. Honest, upfront. Great Ebayer. AAAA++++	dewans
+	Very friendly staff helped me when I needed it! Highly recommended!	briteyedbear
+	Thanks again!!!!! Works great!!!!!	shaffl
+	Great seller great communication A++++ Thanks	patch0909
+	great item and fast service A ++++++++++++	cowgirl_always

3.1. Hypothesis Testing: Individual Level

At the individual level, we used the R package SentimentR to assign sentiment scores to textual comments associated with each rating. Table 2 lists a few examples of negative, neutral, and positive ratings. For the first row, the comment associated with a negative rating has a sentiment score of -0.124, the one associated with a neutral rating is -0.08, and the one associated with a positive rating is $+0.632$.

The sentiment scores for all the negative, neutral, and positive ratings are described as boxplots in Figure 2. We can observe that some buyers express positive segments but still post negative or neutral ratings. We also observe that some buyers express negative sentiments but still post positive ratings. Overall, the mean values of sentiment scores generally align with the types of feedback ratings. That is, negative ratings have a negative mean value of sentiment scores, neutral ratings have a close to 0 mean value of sentiment scores, and positive ratings have a positive mean value of sentiment scores. More formally, we first conducted a one-way ANOVA test and examined whether the mean values of the three types of feedback ratings were the same. The ANOVA testing suggests that these three mean values are not the same. The *t*-tests show that the types of feedback ratings are consistent with the sentiments revealed from the comments posted by buyers. This suggests that at the individual level, we accept H1. The *t*-tests and one-way ANOVA tests are summarized in Table 3.

Table 2. Sentiment scores and textual comment samples at individual level.

Score	Negative Rating	Score	Neutral Rating	Score	Positive Rating
−0.124	"item still not received, paid 9 days ago. run-around on the emails"	+0.043	"Dont like leaving bad feedback, but this guy deserves it."	+0.695	"Great seller. Fast delivery. Thanks!"
−0.080	"This seller did not provide a shipping date and was not punctual not recomended!"	−0.083	"Tried calling seller, phone # he gave doesnt work, doesn"	+0.150	"Great seller!! product as described A+A+A+A+A+A+A+A+ would deal with again!!!"
−0.200	"Lost my payment, late shipment, accused of not paying"	0.000	"no comment-"	+0.583	"quick book in great shape, will deal again AAA+++"
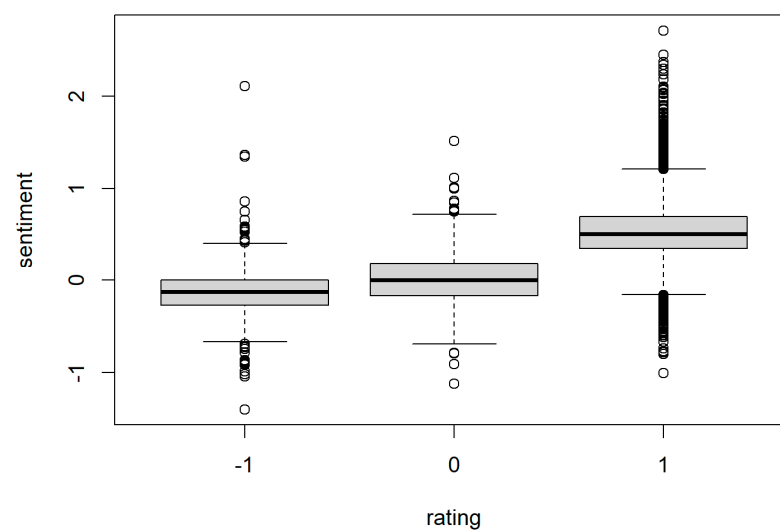
Figure 2. Boxplots of sentiment scores.

Table 3. Individual-level test.

	Negative Rating	Neutral Rating	Positive Rating
Mean	−0.1326 ***	0.0205	0.5344 ***
t-value	−19.1910	1.3896	279.22
p-value	<0.01	>0.10	<0.01
Testing Result		Accept H1	
ANOVA Test		F-value = 6042 ***, $p < 0.01$	

*: Significant at 10%; **: significant at 5%; ***: significant at 1%.

3.2. Hypothesis Testing: Group Level

At the group level, we evenly divided the negative, neutral, and positive ratings datasets into 10 groups. Then, we mined each group to obtain sentiment scores by using three lexicons: Bing, Loughran, and NRC, respectively. The rationale for using three lexicons is that each one has unique strengths and weaknesses. Using three also allows for cross-verification. First, we carried out a one-way ANOVA test to see if the mean values of sentiments for the three types of feedback ratings were the same. The testing shows that the *p*-value is <0.01 for each case of three lexicons, which suggests that the mean values of sentiments under the three-type ratings differ. Then, we conduct *t*-tests for each type of ratings, and the testing summary is listed in Table 4. We can see that for negative ratings, the mean sentiment score associated is also negative. For neutral ratings, the testing under

the Bing case suggests that the mean sentiment score associated with neutral feedback ratings is equal to zero. The testing under Loughran suggests the mean sentiment score associated with neutral ratings is negative, but NRC suggests that it is positive. Overall, at the group level, we can accept Hypothesis 2.

Table 4. Group-level test.

	Bing	**Loughran**	**NRC**	**Testing Result**
Negative rating	−0.4698 ***	−0.4518 ***	−0.2226 ***	
t-value	−7.6786	−10.8230	−4.9324	
p-value	<0.01	<0.01	<0.01	
Neutral rating	0.1375	−0.1875 **	0.2062 ***	
t-value	2.1754	−2.8980	3.4883	
p-value	>0.10	<0.05	<0.01	Accept H2
Positive rating	1.7396 ***	0.8472 ***	0.6965 ***	
t-value	91.5120	112.7600	118.7100	
p-value	<0.01	<0.01	<0.01	
ANOVA test	F = 287.3 $p < 0.01$	F = 154.7 $p < 0.01$	F = 233.1 $p < 0.01$	

*: Significant at 10%; **: significant at 5%; ***: significant at 1%

3.3. Hypothesis Testing: Total Level

At the total level, we used the dataset as a whole for each type of feedback rating. We also used three lexicons to retrieve the sentiments for each type. Firstly, we carried out one-way ANOVA testing to see if the mean values of sentiments for the three types of feedback ratings were the same. The testing result suggests that the mean values of sentiments for the three types of ratings are not the same. Then, we conducted *t*-tests for each type of ratings. The testing summary is listed in Table 5. Overall, at the total level, the types of feedback ratings are consistent with the sentiments revealed from the comments posted by buyers. This suggests that we accept H3.

Table 5. Total-level test.

	Negative Rating	**Neutral Rating**	**Positive Rating**
Mean	−0.3803 **	0.0201	1.0944 ***
t-value	−4.7934	0.1316	3.3623
p-value	<0.05	>0.10	<0.1
Testing result	Accept H3		
ANOVA test	F-value = 21.95 ***, $p < 0.01$		

*: Significant at 10%; **: significant at 5%; ***: significant at 1%.

3.4. How Are Ratings Correlated to Sentiments? Correlation Coefficient Analysis

The above analyses show that the three types of feedback ratings (negative, neutral, and positive) align with the types of sentiments (negative, neutral, and positive). This finding further validates the usefulness and effectiveness of eBay's feedback rating system. This finding might boost users' confidence in using the rating system and encourage more users to utilize it. Beyond this, we move to investigate the consistency further: Do the three types of feedback ratings fully reflect the sentiments embedded in comments? If not, can we do better with eBay's feedback rating system?

To check if the three types of feedback ratings fully reflect the sentiments embedded in comments, we used correlation coefficient analysis and examined whether both are highly

correlated. We conducted the correlation coefficient analysis at three levels: (1) individual level, (2) group level, and (3) total level. The analytical outcomes are summarized in Table 6.

Table 6. Correlation coefficient analysis.

Level	Coefficient	*t*-Value	*p*-Value
Individual	0.4311 ***	77.7294	<0.01
Group—Bing	0.9545 ***	16.9510	<0.01
Group—Loughran	0.9201 ***	12.4361	<0.01
Group—NRC	0.9448 ***	15.2683	<0.01
Total	0.8707 ***	4.6851	<0.01

*: Significant at 10%; **: significant at 5%; ***: significant at 1%.

According to Table 6, the correlation between the types of feedback ratings and sentiments is significantly positive. However, the magnitude of the positive correlation at the individual level is relatively low (0.4311 < 0.5). One potential reason is that the feedback ratings have just three types: negative (-1), neutral (0), and positive (+1). If eBay wants feedback ratings to represent sentiments better, it probably needs to extend the feedback ratings from three types (-1, 0, +1) to five types (-2, -1, 0, +1, +2). At the group level, the correlation between the types of feedback ratings and sentiments is also significantly positive. The magnitude of the positive correlation is relatively high (>0.92). At the group level, the sentiments are more stable than at the individual level because we combine a number of feedback ratings into one group and use the mean values of each group to calculate the correlation coefficients. At the total level, the correlation between the types of feedback ratings and sentiments is significantly positive at 0.8707.

3.5. The Symmetry of Distance between Positive, Negative, and Neutral Sentiments

The three types of feedback ratings seem symmetric, with the same distance from 0 to -1 as that from 0 to +1. Put another way, the ratio of two distances should be 1 if the two distances are the same. To study if the three types of feedback ratings really reflect the sentiments embedded in comments, we measured if the sentiments were also symmetric. In Table 7, we report the ratio of the sentiment distance between positive feedback and neutral feedback over the distance between neutral feedback and negative feedback. For example, for SentimentR algorithm, the ratio = $(0.5344 - 0.0205)/(0.0205 - (-0.1326)) = 3.3566$. This is the ratio for sentiment scores at the individual level. This suggests that sentiment scores have an asymmetric distribution among negative, neutral, and positive feedback ratings. The ratio for NRC lexicon is close to 1. This indicates that NRC lexicon generates a relatively symmetric distribution of sentiment scores among negative, neutral, and positive feedback ratings. The ratio for Loughran lexicon is more than 6, indicating that Loughran lexicon generates a relatively asymmetric distribution of sentiments among negative, neutral, and positive feedback ratings. This asymmetrical distribution of sentiments is also depicted in Figure 3. The radar diagram demonstrates the asymmetry by showing that the triangles have a sharper angle from positive ratings.

Table 7. Distance by sentiment scores.

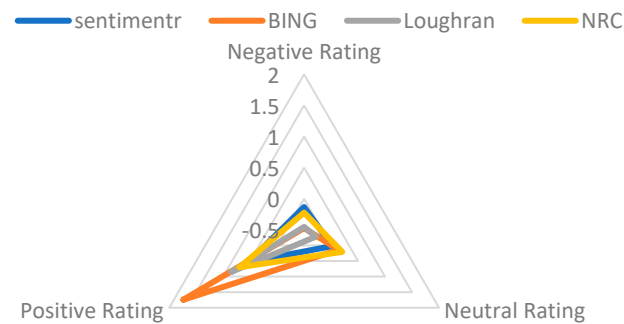| Algorithm	Negative Rating (n = 1063)	Neutral Rating (n = 481)	Positive Rating (n = 24,921)	Ratio
SentimentR	-0.1326	0.0205	0.5344	3.3566
Bing	-0.4685	0.1372	1.7396	2.6455
Loughran	-0.4506	-0.2827	0.8472	6.7314
NRC	-0.2220	0.2058	0.6965	1.1468

Figure 3. Radar diagram of sentiment scores.

At the individual level, the sentiment score distributions of negative, neutral, and positive feedback ratings are depicted in Figure 4a–c. While each one has a tall spike around the mean value, they also spread to both sides. To better view the distribution of sentiments for the three types of feedback ratings, we chose 481 samples of neutral ratings and randomly selected 481 negative and positive ratings, respectively. Then, we combined them into one dataset. The density chart of sentiments is shown in Figure 5. The distribution expands to both the right and left sides. The mean value of sentiments for this combo is 0.1423, and the *p*-value < 0.01, which suggests the density skews to the right. This is evidence that the three simple types of feedback ratings (−1, 0, +1) cannot fully align with the sentiments revealed in the comments posted by buyers.
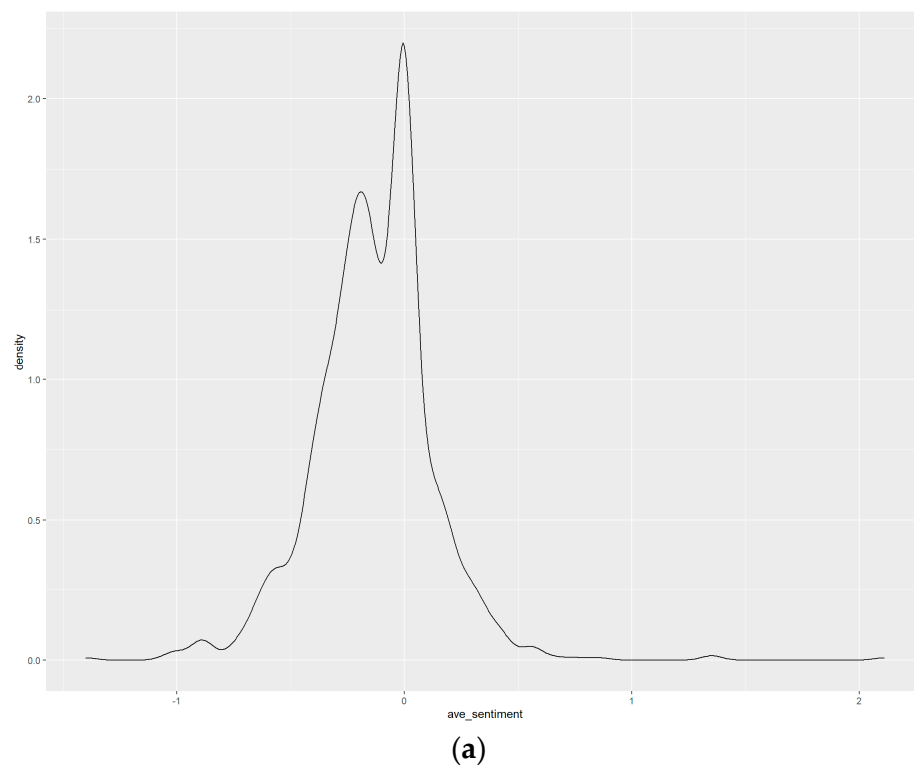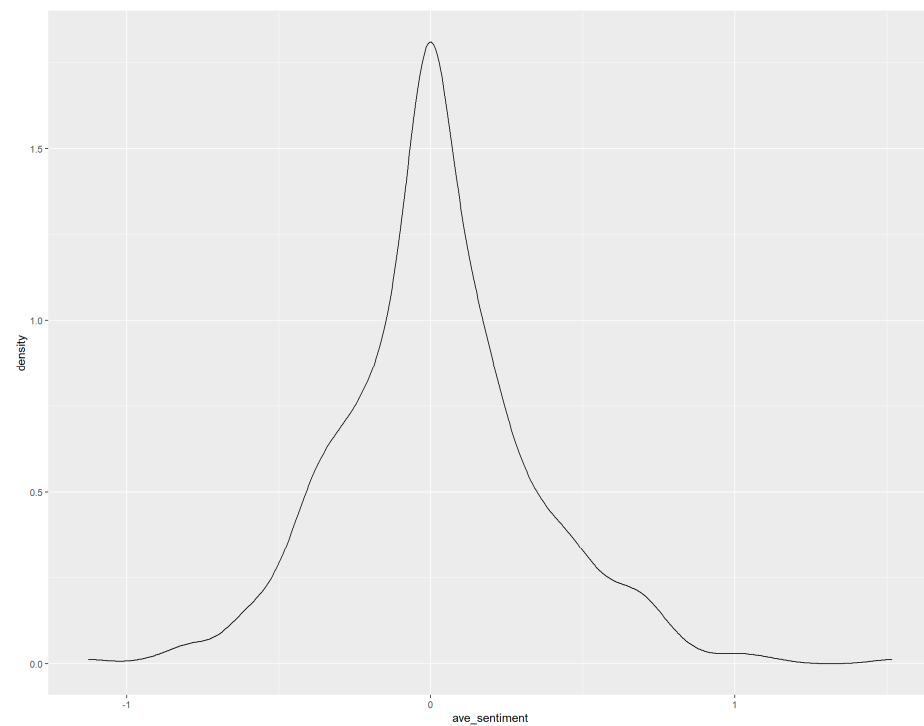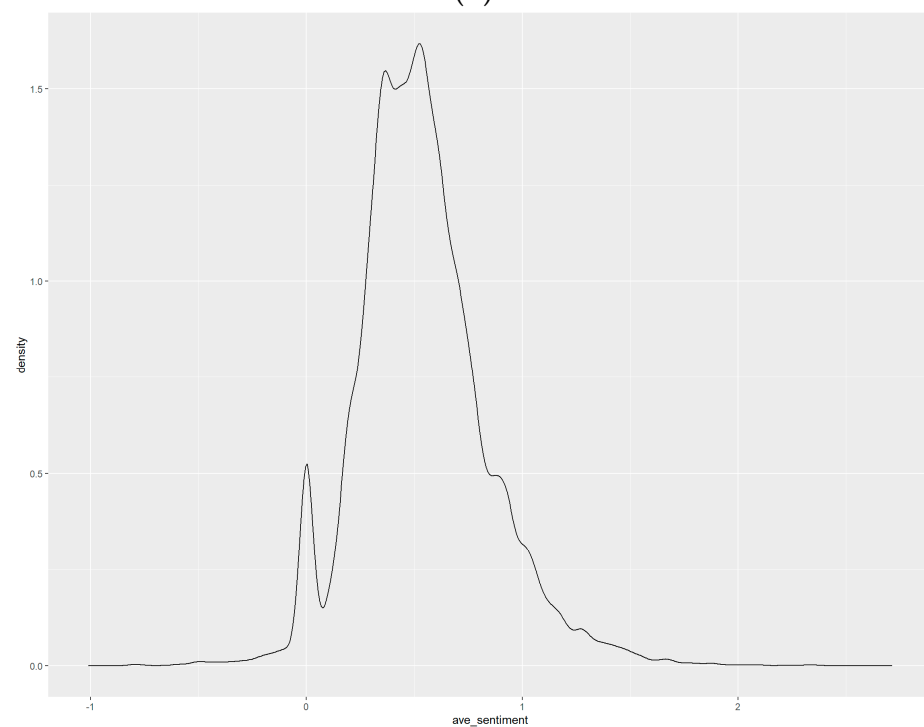

(a)

Figure 4. *Cont.*

Figure 4. (**a**) Density of sentiments for negativefeedback ratings. (**b**) Density of sentiments for neutral feedback ratings. (**c**) Density of positive feedback ratings.
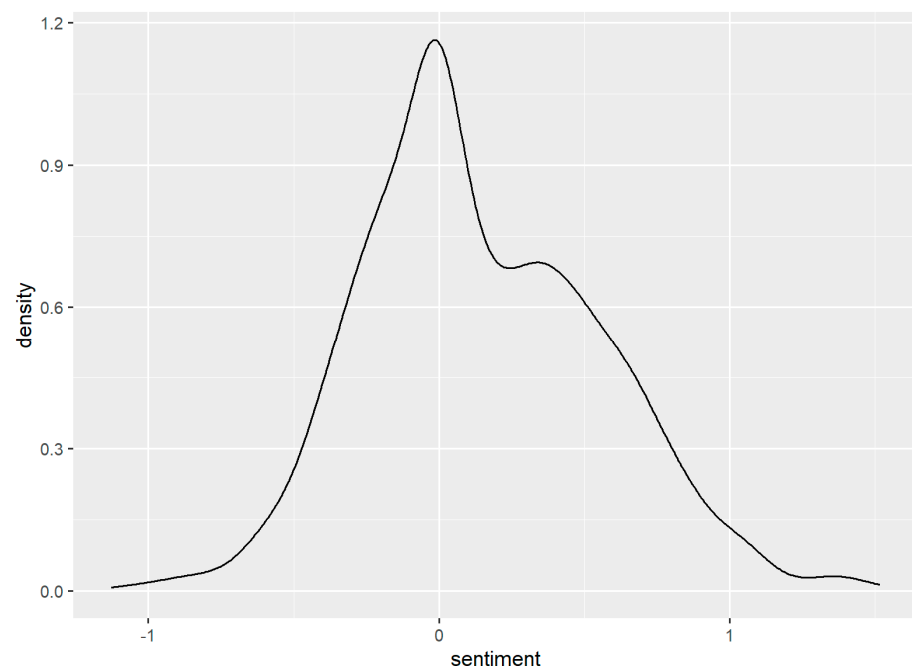
Figure 5. Density of sentiments for three types of ratings with even sample size.

4. Discussions and Managerial Implications

Compared with the existing literature, this study uses sentiment analysis and obtains new insights in the following areas: (1) checking the effectiveness of eBay feedback rating systems; (2) addressing certain drawbacks existing in the system; and (3) proposing a potential solution to improve the system.

In the literature, existing studies addressed the system's effectiveness from the perspectives of building trust, earning premium prices, reducing information asymmetry, etc. [3,12–15]. These studies mainly focused on the functions and benefits of the existing feedback system and concluded that the system was running effectively. Unlike past studies, this paper contributes to the literature by using sentiment analysis as a gauge to assess the eBay feedback rating system. Using a different perspective, this study addresses the effectiveness by checking if the ratings really align with the sentiments of buyers. More specifically, we study the consistency between ratings and sentiments by looking at three dimensions: (1) checking whether signs (+, 0, −) of ratings and sentiments are the same for each rating type or not; (2) measuring the correlation coefficients between ratings and sentiments; (3) determining the distances among three types of ratings and sentiments, respectively. We find that the types of feedback ratings have the same signs as the sentiments embedded in textual comments associated with them. Consistent with the existing studies in the literature at an aggregate level, our study also suggests the effectiveness of the eBay feedback rating system and that people can confidently use the system to choose appropriate trading partners. eBay might promote this finding and encourage more people to use its feedback system.

While eBay's feedback rating system is generally running effectively, eBay needs to improve and optimize the system further. In order to do so, any potential drawbacks in the feedback rating system should be carefully investigated. In the literature, existing studies found some issues and problems in the system, such as rating retaliation, changing of online identity, fake positive ratings, and "market for feedback" [4,6,22–24]. These studies mainly focus on some users taking advantage of and abusing the feedback rating system. In this study, we use a unique text mining approach—sentiment analysis—and investigate an area that has not been effectively addressed previously: are buyers' ratings and their sentiments highly correlated, and do their distributions match? From the perspective of sentiment consistency, we find certain new issues concerning eBay's feedback rating system. We

observe that the magnitude of the correlation coefficient at the individual level is relatively low (0.4311 < 0.5). This suggests that while the signs (+, 0, −) of ratings and sentiments are consistent, the ratings cannot fully reflect the sentiments of buyers. While the types of ratings are symmetric, we find the sentiments are not symmetric in terms of distances between positive, negative, and neutral sentiments. This means that the distributions of buyers' ratings and sentiments do not match. There might be many factors leading to these issues. One potential reason is that the feedback ratings have just three types (negative (−1), neutral (0), and positive (+1)), but the domain of sentiments is not limited to between −1 and +1. The simple three types (+, 0, −) of ratings cannot be well correlated with the sentiments derived in textual comments. At eBay, if buyers are not satisfied, they can only choose the negative rating without any other scale choices. The same case is also true for buyers who are satisfied: they can only choose the positive rating without any other scale choices. Those who post neutral ratings are not completely indifferent. Some of them have positive sentiments, and others have negative sentiments.

The existing studies focused on user behavior to propose their solutions to improving eBay feedback rating systems. For example, Ref. [25] suggested methods to reduce user collusion and short-lived online identity problems, and Ref. [26] recommended methods to improve a new product impact evaluation. This study focuses on the system building and proposes that eBay expand the rating choices from the three types of feedback ratings (+, 0, −) to five types, such as (−2, −1, 0, +1, +2) or (1, 2, 3, 4, 5). More choices would help buyers more accurately evaluate post-purchase experiences and better express the sentiments associated with their purchasing. All these revisions will improve the correlation between ratings and sentiments embedded in textual comments and permit the distributions of ratings and sentiments to match more closely.

Of course, expanding the current three-type ratings into five-type ratings might present the following challenges:

- How to transfer existing three-type ratings into five-type ratings. eBay has been using three-type feedback ratings for many years. If the system is changed from three-type ratings into five-type ratings, eBay will need to handle the past three-type ratings properly. Many sellers have used three-type ratings to accumulate high user profiles, and adjusting past ratings to compile new user profiles is a big challenge for eBay.
- Using expanded feedback ratings might confuse buyers. The three-type rating system might be simple for buyers to use. If the system is changed into a five-type rating system, buyers have more choices so they require extra involvement to differentiate and choose various ratings.
- Which five-type rating is better? (1, 2, 3, 4, 5) vs. (−2, −1, 0, +1, +2)? Mathematically, both have five levels or scales. Psychologically, people might have different perceptions of (1, 2, 3, 4, 5) vs. (−2, −1, 0, +1, +2). Further studies are needed to compare them and choose the better one in the context of online marketplaces.

While these challenges exist, we still believe that it would be worthwhile for eBay or academia to conduct an experiment or a pilot study. Expanding the current feedback ratings or other recommendations should be encouraged and supported in future studies.

5. Conclusions and Limitations

Unlike existing studies which explored the system's effectiveness from the perspectives of building trust, earning premium prices, and reducing information asymmetry, this paper studies if eBay's feedback ratings are consistent with the sentiments reflected in the comments posted by buyers. Using the data collected from eBay, we test the consistency in three dimensions at three levels. We obtained the following main findings: the sentiment scores associated with negative feedback ratings are negative, the sentiment scores associated with neutral feedback ratings are equal to zero, and the sentiment scores associated with positive feedback ratings are positive. The finding suggests the effectiveness of eBay's feedback rating system and that people can confidently use the system to make purchase decisions.

While some limitations in eBay's feedback rating system were discussed at length in the previous literature, this paper finds a few new limitations related to the correlation between buyers' ratings and sentiments and their asymmetry in distance and distribution. At the individual level, the correlation coefficient between the types of ratings and the sentiments revealed in the comments is relatively low. The three types of ratings are symmetric $(-1, 0, +1)$ but the textual sentiments are not. The sentiments from the textual comments have asymmetric distributions among negative, neutral, and positive feedback ratings. All these suggest that the three simple types of feedback ratings cannot fully reflect the sentiments from the comments posted by buyers. We propose that expanding the three types of ratings into five types, such as $(-2, -1, 0, +1, +2)$, might help resolve the issue. We also discuss some potential challenges associated with the proposal.

There are a few limitations to this study. First, we use the dataset on one product category, video game consoles, and our findings may not extend to other product categories. Second, we only use four lexicon methods to analyze sentiments from the comments of buyers. More lexicons or non-lexicon sentiment analysis might offer more robust findings. Other topics, such as comparative studies between different e-commerce platforms and comparative studies between periods, types of users, and types of products are also left for future research.

Author Contributions: Conceptualization, X.Z., Y.T. and K.Z.; methodology, K.Z. and Y.T.; software, Y.T.; validation, X.Z. and K.Z.; formal analysis, Y.T. and K.Z.; data curation, Y.T.; writing—original draft preparation, Y.T. and X.Z.; writing—review and editing, K.Z. and X.Z.; visualization, X.Z. and Y.T.; project administration, X.Z., Y.T. and K.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the paper.

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

References

1. Gregg, D.; Parthasarathy, M. Factors affecting the long-term survival of eBay ventures: A longitudinal study. *Small Bus. Econ.* **2017**, *49*, 405–419. [CrossRef]
2. Tadelis, S. Reputation and feedback systems in online platform markets. *Annu. Rev. Econ.* **2016**, *8*, 321–340. [CrossRef]
3. Bharadwaj, K.K.; Al-Shamri, M.Y. Fuzzy computational models for trust and reputation systems. *Electron. Commer. Res. Appl.* **2009**, *8*, 37–47. [CrossRef]
4. Masclet, D.; Pénard, T. Do reputation feedback systems really improve trust among anonymous traders? An experimental study. *Appl. Econ.* **2012**, *44*, 4553–4573. [CrossRef]
5. Resnick, P.; Zeckhauser, R.; Swanson, J.; Lockwood, K. The value of reputation on eBay: A controlled experiment. *Exp. Econ.* **2006**, *9*, 79–101. [CrossRef]
6. Ye, S.; Gao, G.; Viswanathan, S. Strategic Behavior in Online Population Systems: Evidence from Revoking on eBay. *MIS Q.* **2014**, *38*, 1033–1056. [CrossRef]
7. Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Manag. Sci.* **2003**, *49*, 1407–1424. [CrossRef]
8. Goes, P.; Tung, Y.A.; Tu, Y. Seller heterogeneity in electronic marketplaces: A study of new and experienced sellers in eBay. *Decis. Support Syst.* **2013**, *56*, 247–258. [CrossRef]
9. Przepiorka, W. Buyers pay for and sellers invest in a good reputation: More evidence from eBay. *J. Socio-Econ.* **2013**, *42*, 31–42. [CrossRef]
10. Malaga, R.A. Web-Based Reputation Management Systems: Problems and Suggested Solutions. *Electron. Commer. Res.* **2001**, *1*, 403–417. [CrossRef]
11. Goes, P.; Tung, Y.A.; Tu, Y. Online auction segmentation and effective selling strategy: Trust and information asymmetry perspectives. *J. Electron. Commer. Res.* **2017**, *18*, 189–211.
12. Anastasiei, B.; Dospinescu, N.; Dospinescu, O. Word-of-mouth engagement in online social networks: Influence of network centrality and density. *Electronics* **2023**, *12*, 2857. [CrossRef]

13. Fayyaz, Z.; Ebrahimian, M.; Nawara, D.; Ibrahim, A.; Kashef, R. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. *Appl. Sci.* **2020**, *10*, 7748. [CrossRef]

14. Srifi, M.; Oussous, A.; Lahcen, A.A.; Mouline, S. Recommender systems based on collaborative filtering using review texts—A survey. *Information* **2020**, *11*, 317. [CrossRef]

15. Ba, S.; Pavlou, P.A. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Q.* **2002**, *26*, 243–268. [CrossRef]

16. Resnick, P.; Zeckhauser, R. Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In *The Economics of the Internet and E-Commerce (Advances in Applied Microeconomics)*; Baye, M.R., Ed.; Emerald Group Publishing Limited: Bingley, UK, 2002; Volume 11, pp. 127–157.

17. Hortacsu, A.; Cabral, L. *The Dynamics of Seller Reputation: Theory and Evidence from eBay*; National Bureau of Economic Research: Cambridge, MA, USA, 2004; Available online: http://www.nber.org/papers/w10363 (accessed on 10 December 2023).

18. Houser, D.; Wooders, J. Reputation in auctions: Theory, and evidence from eBay. *J. Econ. Manag. Strategy* **2006**, *15*, 353–369. [CrossRef]

19. Khopkar, T.; Li, X.; Resnick, P. Self-selection, slipping, salvaging, slacking, and stoning: The impacts of negative feedback at eBay. In Proceedings of the 6th ACM Conference on Electronic Commerce, Vancouver, BC, Canada, 5–8 June 2005; pp. 223–231.

20. Lee, Z.; Im, I.; Lee, S.J. The effect of buyer feedback scores on Internet auction prices. *J. Organ. Comput. Electron. Commer.* **2006**, *16*, 51–64. [CrossRef]

21. Zhang, J. The roles of players and reputation: Evidence from eBay online auctions. *Decis. Support Syst.* **2006**, *42*, 1800–1818. [CrossRef]

22. Bauerly, R.J. Online auction fraud and eBay. *Mark. Manag. J.* **2009**, *19*, 133–143.

23. Brown, J.; Morgan, J. Reputation in online auctions: The market for trust. *Calif. Manag. Rev.* **2006**, *49*, 61–81. [CrossRef]

24. Dini, F.; Spagnolo, G. Buying reputation on eBay: Do recent changes help? *Int. J. Electron. Bus.* **2009**, *7*, 581–598. [CrossRef]

25. Swamynathan, G.; Almeroth, K.C.; Zhao, B.Y. The design of a reliable reputation system. *Electron. Commer. Res.* **2010**, *10*, 239–270. [CrossRef]

26. Lu, J.; Zhan, X.; Zhan, X.; Shi, L. A book-influence-evaluation method based on user ratings of e-commerce platform. *Electronics* **2022**, *11*, 4198. [CrossRef]

27. Kauffmann, E.; Peral, J.; Gil, D.; Ferrández, A.; Sellers, R.; Mora, H. Managing Marketing Decision-Making with Sentiment Analysis: An Evaluation of the Main Product Features Using Text Data Mining. *Sustainability* **2019**, *11*, 4235. [CrossRef]

28. Kauffmann, E.; Peral, J.; Gil, D.; Ferrández, A.; Sellers, R.; Mora, H. A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Ind. Mark. Manag.* **2020**, *90*, 523–537. [CrossRef]

29. Pathak, H.; Pathak-Shelat, M. Sentiment analysis of virtual brand communities for effective tribal marketing. *J. Res. Interact. Mark.* **2017**, *11*, 16–38. [CrossRef]

30. Younis, E.M.G. Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study. *Int. J. Comput. Appl.* **2015**, *112*, 44–48.

31. Hu, N.; Koh, N.S.; Reddy, S.K. Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decis. Support Syst.* **2014**, *57*, 42–53. [CrossRef]

32. Capriello, A.; Mason, P.R.; Davis, B.; Crotts, J.C. Farm tourism experiences in travel reviews: A cross-comparison of three alternative methods for data analysis. *J. Bus. Res.* **2011**, *66*, 778–785. [CrossRef]

33. Park, S.B.; Ok, C.M.; Chae, B.K. Using Twitter Data for Cruise Tourism Marketing and Research. *J. Travel Tour. Mark.* **2016**, *33*, 885–898. [CrossRef]

34. Asur, S.; Huberman, B.A. Predicting the future with social media. *Web Intell. Intell. Agent Technol.* **2010**, *1*, 492–499.

35. Philandera, K.; Zhong, Y.Y. Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *Int. J. Hosp. Manag.* **2016**, *55*, 16–24. [CrossRef]

36. Ibrahim, E.; Gherbi, E.A. Building sentiment analysis model and compute reputation scores in e-commerce environment using machine learning techniques. *Int. J. Organ. Collect. Intell. (IJOCI)* **2020**, *10*, 32–62.

37. Boumhidi, A.B.; Nfaoui, E.H. Cross-platform reputation generation system based on aspect-based sentiment analysis. *IEEE Access* **2022**, *10*, 2515–2531. [CrossRef]

38. Chiu, C.; Chiu, N.H.; Sung, R.J.; Hsieh, P.Y. Opinion mining of hotel customer-generated contents in Chinese weblogs. *Curr. Issues Tour.* **2015**, *18*, 477–495. [CrossRef]

39. Dhaoui, C.; Webster, C.M.; Tan, L.P. Social media sentiment analysis: Lexicon versus machine learning. *J. Consum. Mark.* **2017**, *34*, 480–488. [CrossRef]

40. Pang, B.; Lee, L. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]

41. Yang, C.S.; Chen, C.H.; Chang, P.C. Harnessing consumer reviews for marketing intelligence: A domain-adapted sentiment classification approach. *Inf. Syst. e-Bus. Manag.* **2015**, *13*, 403–419. [CrossRef]

42. Ghose, A.; Ipeirotis, P.G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 1498–1512. [CrossRef]

J. Theor. Appl. Electron. Commer. Res. **2023**, *18*

2272

43. Tsang, A.S.L.; Prendergast, G. Is a "star" worth a thousand words? The interplay between product-review texts and rating valences. *Eur. J. Mark.* **2009**, *43*, 1269–1280. [CrossRef]

44. Batchelor, J. Ampere Analysis: Console Market Grew to Record \$60bn in 2021. 2022. Available online: https://www.gamesindustry.biz/articles/2022-03-01-ampere-analysis-console-market-grew-to-record-usd60bn-in-2021 (accessed on 10 December 2023).