Mechanisms of Negativity Bias: An Empirical Exploration of App Reviews in Apple's App Store

Completed Research Paper

Dezhi Yin
Trulaske College of Business
University of Missouri
Columbia, MO 65211
yind@missouri.edu

Sabyasachi Mitra
Scheller College of Business
Georgia Institute of Technology
Atlanta, GA 30308
saby.mitra@scheller.gatech.edu

Han Zhang
Scheller College of Business
Georgia Institute of Technology
Atlanta, GA 30308
han.zhang@scheller.gatech.edu

Abstract

Researchers in many diverse areas have consistently found that we are unduly influenced by negative information. In electronic commerce, this negativity bias is evident in the effect of product reviews on consumer behavior in the information systems literature. While the negativity bias is well documented, there has been little systematic and empirical research on its underlying causes. Utilizing a novel data set collected from Apple's App Store, we examine three probable causes of the negativity bias: that negative reviews are more specific, that they have higher surprise value, and that they increase our ability to avoid losses. The empirical analysis revealed that while all three mechanisms contribute to the negativity bias, the 'surprise' factor and the ability to avoid losses play a more prominent role when consumers process and integrate positive and negative review information. Our findings also carry important practical implications for review platforms and online companies.

Keywords: negativity bias, online word-of-mouth, product review, review rating, review helpfulness

Introduction

In many diverse areas of human activity, researchers have consistently found that we are unduly influenced by negative information when forming impressions and making judgments. For example, when subjects are provided with two seemingly opposite descriptions of a person, their final impression is closer to the more negative of the two descriptions (Fiske 1980). More recently, researchers have observed higher electrical activity in the brain in response to negative stimuli compared to positive stimuli that are equally probable, extreme and arousing (Ito et al. 1998; Schupp et al. 2004). This bias for negative information develops early in childhood and remains consistent through the later stages in life (Vaish et al. 2008). In the psychology literature, there is abundant evidence of a generalized negativity bias in information processing, whereby "bad things will produce larger, more consistent, more multifaceted or more lasting effects than good things" (Baumeister et al. 2001, p. 325). The tendency to overweigh negative information has been established as a general principle in the domains of perception, memory, impression formation, emotional response, and marketing (Ahluwalia 2002; Kahneman and Tversky 1979; Rozin and Royzman 2001). In summary, there is ample empirical evidence that humans rely on negative information more than positive information in making sense of the world around them – this preference for the negative is termed the negativity bias.

In electronic commerce, the negativity bias is evident in the effect of product reviews on consumer behavior. Research in information systems (IS) and marketing demonstrate that consumers find negative product reviews more helpful than positive reviews. For example, Chevalier and Mayzlin (2006) study online book reviews at Amazon.com and Barnesandnoble.com, and find that the impact of negative reviews on sales outweighs the impact of positive reviews. Likewise, Basuroy, Chatterjee, and Ravid (2003) find that negative reviews in the film industry hurt box office performance more than positive reviews help performance. More direct evidence of negativity bias is also observed concerning the helpfulness perceptions of reviews. For instance, examining online product reviews and using laboratory experiments, Sen and Lerman (2007) find a negativity bias for utilitarian products such that more negative reviews are more influential than positive reviews in shaping consumer perceptions. Cao, Duan, and Gan (2011) also observe evidence of a negativity bias in their study of CNET software reviews.

While the existence of a negativity bias in online product reviews is well established, the reasons behind this bias have not been systematically examined. It is not immediately clear why consumers find negative product reviews more helpful. The question is important because online product reviews are a significant determinant of sales (Chevalier and Mayzlin 2006; Forman et al. 2008). Many online merchants, such as Amazon, display helpfulness metrics for each review that are determined through a voting mechanism. With a large number of reviews available online, there is evidence that consumers focus more on reviews that others perceive to be more helpful (Ghose and Ipeirotis 2011; Zhu and Zhang 2010). Thus, understanding the mechanisms that underlie the negativity bias is important in designing more effective online product review systems.

The psychology literature indicates that there can be at least three different reasons why consumers perceive negative product reviews to be more helpful than positive product reviews (Feldman 1999; Skowronski and Carlston 1987). First, negative product reviews may be more specific and convey more information than positive product reviews because consumers may describe negative experiences in greater detail than positive experiences. Second, most online product reviews describe positive experiences with the product, and thus negative reviews may have greater "surprise" value to readers. The greater surprise value increases the perceived helpfulness of the review. Finally, humans are risk averse, and they may pay closer attention to negative information to avoid and mitigate the risks associated with product purchase. Thus, they may find negative reviews that inform them of the risks associated with the product more helpful than positive reviews that highlight the benefits.

In this paper, we describe a large-scale empirical study that examines the helpfulness of over 400,000 reviews of over 62,000 apps in Apple's app store for which we could calculate a helpfulness rating based on the votes each review received from consumers. Following existing literature (Cao et al. 2011; Mudambi and Schuff 2010; Sen and Lerman 2007), we first demonstrate the existence of a negativity bias in the helpfulness ratings. More specifically, we show that reviews that assign lower ratings to products are perceived to be more helpful by consumers, after controlling for other factors that can influence perceived helpfulness. Next, through a combination of mediation and moderation tests, we investigate the reasons behind the negativity bias in the data. We develop measures to capture the information specificity and surprise value of online product ratings, and we demonstrate through standard mediation tests that these two variables mediate the relationship between product rating of a review and its perceived helpfulness. Since the purchase risk associated with a product is not defined at the review level (they are defined at the product level instead), we demonstrate that the negativity bias is lower for products that are free (lower purchase risk) through standard moderation tests based on interaction effects. In summary, we empirically demonstrate that all three mechanisms described above contribute to the negativity bias in online product reviews.

Our primary theoretical contribution to the emerging literature on online product reviews is that we identify and empirically demonstrate three underlying reasons behind the negativity bias. While the existence of the negativity bias in online reviews is well documented in the literature (Cao et al. 2011; Mudambi and Schuff 2010; Sen and Lerman 2007), there has been no systematic evaluation of its underlying causes, and we fill that gap in the literature. Our research has two implications for practice. First, many online retailers (such as Amazon) sort user reviews of products and sellers according to their helpfulness ratings by default. In many cases, consumers may only read a small set of helpful reviews, and sorting based on helpfulness enables them to shorten the information search, evaluate alternatives more efficiently, and make better purchase decisions (Cao et al. 2011; Mudambi and Schuff 2010). However, the negativity bias in online reviews can cause more negative reviews to appear higher in the sort order, perhaps unduly affecting the sales of the product. Understanding the mechanisms that underlie the negativity bias provides additional ways to sort, to emphasize and to highlight those reviews that consumers may find useful but may not have received high helpfulness scores due to the negativity bias. Second, online retailers can increase the helpfulness of all reviews by encouraging reviewers to adopt the characteristics of negative reviews that consumers find helpful (the underlying causes of the negativity bias), such as encouraging reviewers to be more specific, to provide more distinctive information in the reviews, and to discuss the risks associated with products and sellers. By providing these additional cues to reviewers, online retailers can improve the helpfulness of all reviews and level the playing field for negative and positive reviews, thereby reducing the negativity bias.

The rest of the paper is organized as follows. In the next section, we describe our theoretical model. Section 3 describes the data used in the analysis, while section 4 describes the empirical results. Section 5 concludes the paper.

Model and Hypotheses

Theoretical Model

A prominent information processing framework, the information-diagnosticity perspective, argues that receivers utilize incoming information about a target to classify the target into one or more behavioral domains (Skowronski and Carlston 1987). The weight attached to a piece of information is dependent on its diagnosticity, defined as "the degree to which a piece of information implies or determines one's response to a given question or other circumstance requiring a judgment or behavior" (Feldman 1999, p. 48). In other words, a piece of information is diagnostic if it is useful and informative for judgment. A central tenet of this approach is that in general, negative information is more diagnostic than positive information. In the context of online product reviews, we argue that there are three distinct reasons why negative reviews are more diagnostic than positive reviews, and hence more helpful to the consumer in making her decision.

Review Specificity: A probable reason behind the negativity bias is that negative information is more specific and detailed, and carries a narrower range of potential implications (Birnbaum 1974; Wyer 1973). When a consumer has a negative experience with a product, they are more likely to spend the time to write a longer review and provide greater details about the experience. The greater information content of the review makes it less ambiguous and enables prospective consumers to make product judgments more confidently, thereby enhancing the diagnosticity of the review, and consequently its helpfulness.

Review Surprise: Negative cues are perceived to be more helpful because of their contrast with internal standards or reference points that are typically positive (Helson 1964; Sherif and Sherif 1967). In the online environment, each product usually has a large number of reviews, and positive reviews are generally more common and frequent than negative reviews (Chevalier and Mayzlin 2006). Consumers are used to positive reviews, and they are likely to have a positive impression of products in their consideration set prior to a negative experience. Thus, a negative review may evoke greater surprise, attract more attention, and cause elaboration to be sought from the consumer (Fiske 1980), resulting in a perception that such reviews are more diagnostic and thus more helpful.

Loss Avoidance: Loss avoidance refers to the tendency that people are more concerned about negative consequences (e.g., monetary loss, reputational damage, etc.) than about positive consequences (Kahneman et al. 1991). In general, humans tend to be risk averse (Kahneman and Tversky 1979). In their evolutionary history, humans have been attuned to process negative information more carefully to avoid risky situations. Recent research has shown that even in early childhood, humans pay more attention to negative information and that this bias continues through adulthood (Vaish et al. 2008). Since negative reviews alert us to product risks, we may pay greater attention to negative reviews compared to positive reviews that highlight the product benefits.

Figure 1 (Panel 1) shows our theoretical model that captures the three underlying reasons behind the negativity bias described above. Higher *Review rating* (the product rating assigned by the review) decreases *review specificity*, *review surprise*, and *loss avoidance*. Increased review specificity, review surprise and the ability to avoid purchase risk increases *review helpfulness* as perceived by the consumer.

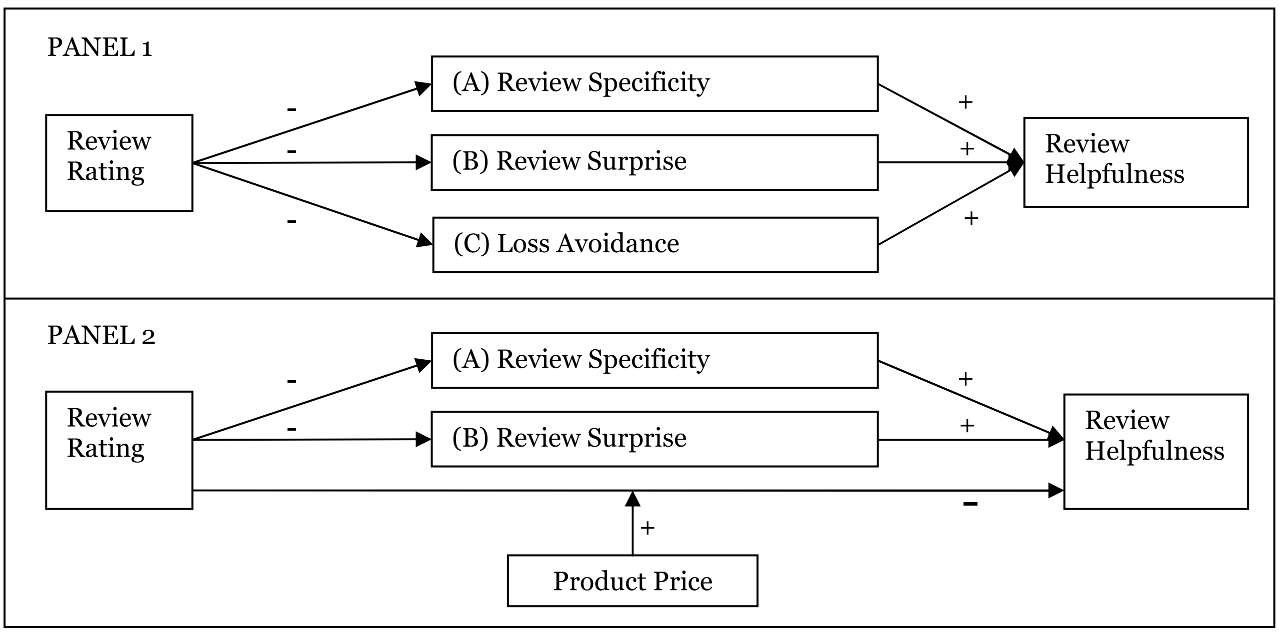

Figure 1: Theoretical and Empirical Models

Modified Model and Hypotheses

While *review specificity* and *review surprise* can be measured at the review level through appropriate proxies we describe later, *loss avoidance* due to the review is very difficult (if not impossible) to measure. However, we argue that the price of the product is an appropriate proxy for the purchase risk associated with the product. Consequently, the negativity bias will be greater for higher priced products, since the final pathway in our theoretical model in Figure 1 (Panel 1) will be stronger for such products.

Figure 1 (Panel 2) shows our empirical model based on the discussion above. Paths A and B in our theoretical model can be evaluated through standard mediation analysis (hypotheses 2 and 3 below). Path C in the theoretical model is evaluated through moderation analysis in the empirical model (hypothesis 4 below). The specific hypotheses we evaluate are shown below.

Hypothesis 1: Lower product rating assigned by a review is associated with higher perceived helpfulness of the review (negativity bias).

Hypothesis 2: (a) Lower product rating assigned by a review is associated with higher review specificity, and (b) higher review specificity is associated with higher perceived usefulness of the review.

Hypothesis 3: (a) Lower product rating assigned by a review is associated with higher review surprise, and (b) higher review surprise is associated with higher perceived usefulness of the review.

Hypothesis 4: The negative relationship between higher product rating assigned by a review and the perceived usefulness of the review (negativity bias) will be stronger for high priced products.

Data and Methods

To test our hypotheses and the mechanisms underlying the negativity bias in online product reviews, we collected and analyzed actual review data from the Apple's App Store, which provides user ratings and reviews for apps. Apps are rapidly becoming a critical way for users to spend time on the Internet. Gartner forecasts that worldwide downloads in mobile application stores will surpass 21.6 billion by 2013 (Gartner 2010). The App Store launched in mid-2008 and now has over 500,000 "apps" approved by Apple. At the time of data collection, the site had accumulated nearly two years of user reviews. In a review, existing users of an app can evaluate the app by leaving a rating on a scale of 1 to 5 stars. Additionally, they can write a text review to provide more details about their experience with the app. When a potential user evaluates the reviews of an app before making a purchase, the review page displays all reviews for that app chronologically, and the most recent reviews appear first by default. Readers of a review can also indicate whether they found a review helpful or not by clicking on the appropriate button.

Data Collection

We collected the data in early April, 2010, using individual reviews as the unit of analysis. We began by identifying 62,266 apps that appeared in the top-500 rankings of all app store categories (20 categories in total, including games, business, reference, social networking, etc.) in the first three months of 2010. Among these apps, 40,417 had at least one review, and we retrieved all their historical reviews. For each review, we collected the following information: rating, text review content, the number of "helpful" votes, and the number of total votes cast by readers ("helpful" and "not helpful"). We also collected app-level information, including the average rating, count of all ratings, app category, and whether the app is paid or free for the consumer.

In order to reduce noise in the reviews, the following steps were taken. First, we dropped 94,815 reviews that included non-ASCII characters (mostly from non-English languages). Next, we dropped reviews that contained no text content (2,743), and reviews that had system errors (38). These steps resulted in 1,623,497 reviews. Of this set, 418,415 reviews (over 25%) had received helpfulness votes (see below). Analysis was conducted on these 418,415 reviews.

Variables

The dependent variable of interest, *Review helpfulness*, was operationalized as follows. Below each review, Apple's App Store lists the question "Was this review helpful?", along with "Yes" and "No" options. A review that has received at least one vote will display the number of "helpful" votes and the total votes received immediately below the review content. Helpfulness was measured as the proportion of "helpful" votes out of the total votes a review received (i.e., the number of people who voted "Yes" divided by the total number of people who cast a vote). Therefore, the helpfulness score ranged from 0 to 1, with a higher percentage indicating a more helpful review. The average helpfulness of the analyzed reviews was 0.59, indicating that most reviews in the final set were considered relatively helpful. Tables 1 and 2 present summary statistics and correlations for the variables in our analysis (described below).

Table 1: Descriptive Statistics for Final Review Pool (N = 418,415)

Variable		Mean	Std. Dev.	Min	Max
1	*Review helpfulness*	0.59	0.42	0	1
2	*Rating*	3.45	1.68	1	5
3	*Reading difficulty*	8.70	54.60	-16.1	25428.7
4	*Review specificity*	41.63	48.96	1	1134
5	*Review surprise*	1.29	0.88	0	3.96

Table 2: Variable Correlations for Final Review Pool (N = 418,415)

Variable		1	2	3	4	5
1	*Review helpfulness*	1				
2	*Review rating*	0.361	1			
3	*Reading difficulty*	-0.002	0.004	1		
4	*Review specificity*	0.135	0.027	-0.005	1	
5	*Review surprise*	-0.281	-0.494	-0.003	-0.064	1

The other variables in our empirical analysis were operationalized as follows: (1) The primary independent variable in our model is *Review rating*. The review rating refers to the star rating of a review; the more stars a review received, the more positive the review is. Ratings ranged from 1 star to 5 stars, and the average rating for the reviews in the data set was 3.45. (2) *Review specificity* is operationalized through the length of the review (the number of words in a review). Reviews in our data set had on average 41.63 words. (3) *Review surprise* is operationalized by the absolute difference between the rating of a review and the average rating of that app. The average rating of the app embodies the general expectations about the app, and review surprise captures deviations from expectations. (4) We used a price dummy (*Paid*) to capture the price of the product, which is equal to 1 if an app was paid for by the consumer, and 0 otherwise.

Following prior literature that examines review helpfulness scores (Korfiatis et al. 2008; Mudambi and Schuff 2010), our analysis controlled for a series of relevant variables. To control for the difficulty levels of text reviews, we calculated the Coleman–Liau Index as a proxy for *Reading difficulty*, which is an estimate of the U.S. grade level that a student would need to have achieved in order to read and understand the text (Coleman and Liau 1975). On average, the reviews in our data set were written at a 9th grade level. We also controlled for the effects of app characteristics, including an app's *Category*, *Average rating* and the *Number of ratings* available for the app. Each app belongs to one of the twenty categories, and we added nineteen dummies to control for cross-category heterogeneity. Average rating captures the overall quality of an app, while the count of all ratings for an app captures its popularity. The operationalization of all variables is summarized in Table 3.

Method

Following the approach in Mudambi and Schuff (2010), we used Tobit regressions in our data analysis. We deemed this approach appropriate for the following reasons. First, the dependent variable is censored in nature: because it was constructed as a ratio, its value is bounded in range. Second, there exists a potential selection bias because not every review reader casts a helpfulness vote. Therefore, a sample containing only voted reviews might be non-random, and the least-squares estimation of this sample would produce biased estimates (Greene and Zhang 2003).

Table 3: Variable Definitions

Variable Type	Variable Level	#	Variable	Operationalization	Notes
DV	Individual Review	1	Review Helpfulness	# helpful_votes / # total_votes	Range: [0, 1]
IV	Individual Review	2	Rating	# of stars	Range: [1, 5]
Mediators	Individual Review	3	Review Specificity	review length	
		4	Review Surprise	absolute difference between review rating and app's average rating	
Moderator	App	5	Paid	=1 if the app is paid; 0 otherwise	
Control	Individual Review	6	Reading Difficulty	Coleman-Liau Index	U.S. grade level necessary to comprehend the text
	App	7	Quality	average rating	Range: [1, 5]
		8	Popularity	# of ratings in total	
		9	Category Dummies	=1 if the app belongs to that category	The category (20 in total) an app belongs to

Table 4: Review Ratings, Review Specificity and Review Surprise

	Model 1	Model 2
	Review Specificity	*Review Surprise*
Review rating	-0.033***	-0.451***
	(0.001)	(0.001)
Paid	0.254***	-0.029***
	(0.002)	(0.001)
Average rating	-0.011***	-0.032***
	(0.001)	(0.001)
Number of ratings	-0.064***	0.013***
	(0.001)	(0.001)
Reading difficulty	-0.006***	0.000
	(0.001)	(0.001)
Category Dummies	included	included
Constant	-0.045***	0.077***
	(0.009)	(0.008)
Log Likelihood	-2253835.41	-2095823.11
R-square	0.060	0.226

Notes: Standard errors in parentheses.*** p<0.01; ** p< 0.05; * p< 0.1

Results

Mediation Analysis

First, we examined the relationship between review rating and the two mediators in the empirical model (Figure 1 Panel B). All continuous variables (except *Review helpfulness*) were standardized to ease interpretation of the results in the following analyses. *Review specificity* and *Review surprise* were entered as the dependent variables (see Table 4). As predicted, *Review rating* is negatively related to both *Review specificity* ($\beta = -0.033$, $p < 0.01$) and *Review surprise* ($\beta = -0.451$, $p < 0.01$).

Next, we tested for mediation at the review level (see Table 5). In the baseline model (Model 1), the coefficient of *Review rating* is negative ($\beta = -0.131$, $p < 0.01$), confirming a negativity bias as predicted in H1. When review length is entered in Model 2, it is positively related to *Review helpfulness* ($\beta = 0.242$, $p < 0.01$) as expected. However, the effect of rating decreased only slightly ($\beta = -0.124$, $p < 0.01$), suggesting that *Review specificity* partially mediates the relationship between *Review rating* and *Review helpfulness*. In Model 3, when *Review surprise* is entered in the regression, it is positively associated with *Review helpfulness* ($\beta = 0.314$, $p < 0.01$). Moreover, the negativity bias is greatly reduced and the sign of the *Review rating* variable turns positive ($\beta = 0.016$, $p < 0.01$), indicating a full mediating effect of *Review surprise*. The coefficient for the *Review rating* variable remains positive when both mediators are included in the model (Model 4). In summary, these results provide empirical support for the first three hypotheses, and confirm that *Review surprise* and *Review specificity* together fully mediate the effect of *Review rating* on *Review helpfulness* (negativity bias).

Moderation Analysis

Finally, to test Hypothesis H4, we included an interaction term in the analysis. When the interaction term (*Review rating * Paid*) is entered in the regression (Model 5), the coefficient for the interaction term is negative and significant ($\beta = -0.031$, $p < 0.01$). In line with H4, the negativity bias is stronger for paid apps and weaker for free apps.

Summary of Results

Our results can be summarized as follows:

(a) We find that *Review rating* is negatively associated with *Review helpfulness* after controlling for relevant review and app characteristics, confirming the negativity bias in earlier literature.

(b) We find that *Review specificity* partially mediates the negative association between *Review rating* and *Review helpfulness* (negativity bias).

(c) We find that *Review surprise* fully mediates the negative association between *Review rating* and *Review helpfulness* (negativity bias).

(d) We find that the negative association between *Review rating* and *Review helpfulness* (negativity bias) is stronger for paid than free apps.

Overall, our results support the three mechanisms underlying the negativity bias in online reviews – review specificity, review surprise and loss avoidance.

Table 5: Tobit Regressions for Mediation and Moderation Analysis

Dependent variable: *Review helpfulness*

	Model 1	Model 2	Model 3	Model 4	Model 5
Rating	-0.131***	-0.124***	0.016***	0.026***	-0.116***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Paid	0.603***	0.534***	0.611***	0.540***	0.601***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Average rating	-0.068***	-0.068***	-0.064***	-0.064***	-0.069***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of ratings	-0.629***	-0.607***	-0.638***	-0.616***	-0.628***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Reading difficulty	0.003	0.005**	0.003*	0.005**	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Category Dummies	included	included	included	included	included
Review specificity		0.242***		0.248***	
		(0.002)		(0.002)	
Review surprise			0.314***	0.319***	
			(0.002)	(0.002)	
Review rating x Paid					-0.031***
					(0.004)
Constant	-2.165***	-2.144***	-2.192***	-2.171***	-2.164***
	(0.022)	(0.021)	(0.021)	(0.021)	(0.022)
Log Likelihood	-1383798.19	-1375268.75	-1373910.12	-1364855.10	-1383768.50
Pseudo R-square	0.062***	0.067***	0.068***	0.074***	0.062***

Notes: Standard errors in parentheses. *** p<0.01; ** p< 0.05; p< 0.1

Robustness Checks

We conducted the following analyses to check the robustness of our results. First, we employed an alternative measure of review specificity: the number of sentences in a review (Cao et al. 2011; Ghose and Ipeirotis 2011) (see Table 6). As shown in Model 1, the length of a review in sentences is negatively associated with *Review rating* ($\beta = -0.021$, $p < 0.01$). Furthermore, when this alternative measure of review length is entered in Model 2, it is positively related to *Review helpfulness* ($\beta = 0.209$, $p < 0.01$), and the effect of rating decreased slightly ($\beta = -0.127$, $p < 0.01$). The prior results concerning the mediating effect of review specificity still hold when the number of sentences is used to quantify this factor.

Second, we entered an additional control variable - *Rating extremity* - to the Tobit regressions in Table 5. *Rating extremity* is measured by the absolute difference between a rating and the middle point of the rating scale (3). It controls for the possibility that extreme reviews may be perceived to be more helpful than moderate reviews (Forman et al. 2008; Ghose and Ipeirotis 2011). As depicted in Table 7, all results reported earlier still hold.

Table 6: Robustness Check with Alternative Measure of Review Specificity

	Model 1 (DV: Specificity)	**Model 2** (DV: helpfulness)	**Model 3** (DV: helpfulness)
Rating	-0.021***	-0.127***	0.020***
	(0.001)	(0.002)	(0.002)
Paid	0.226***	0.550***	0.558***
	(0.002)	(0.004)	(0.004)
Average rating	-0.006***	-0.069***	-0.065***
	(0.001)	(0.002)	(0.002)
Number of ratings	-0.055***	-0.613***	-0.622***
	(0.001)	(0.003)	(0.003)
Reading difficulty	-0.003**	0.004*	0.004**
	(0.001)	(0.002)	(0.002)
Category Dummies	included	included	included
Review Specificity		0.209***	0.211***
		(0.000)	(0.000)
Review surprise			0.315***
			(0.002)
Constant	-0.015*	-2.155***	-2.181***
	(0.001)	(0.021)	(0.021)
Log Likelihood	-2265813	-1377484	-1367422
Pseudo R^2	0.046***	0.066***	0.073***

Table 7: Robustness Check with the Rating Extremity Variable

Dependent variable: *Review helpfulness*

	Model 1	Model 2	Model 3	Model 4	Model 5
Rating	-0.149***	-0.145***	0.017***	0.015***	-0.128***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Rating extremity	0.124***	0.153***	-0.003	0.029***	0.125***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Paid	0.603***	0.530***	0.611***	0.539***	0.600***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Average rating	-0.065***	-0.065***	-0.064***	-0.064***	-0.066***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Number of ratings	-0.630***	-0.607***	-0.638***	-0.616***	-0.629***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Reading difficulty	0.003	0.004**	0.003*	0.005**	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Category Dummies	included	included	included	included	included
Review Specificity		0.256***		0.250***	
		(0.002)		(0.002)	
Review surprise			0.316***	0.306***	
			(0.002)	(0.002)	
Review Rating x Paid					-0.043***
					(0.004)
Constant	-2.148***	-2.122***	-2.192***	-2.166***	-2.146***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Log Likelihood	-1382035.40	-1372557.18	-1373909.16	-1364775.77	-1381978.24
Pseudo R-square	0.063***	0.069***	0.068***	0.074***	0.063***

Notes: Standard errors in parentheses. *** p< 0.01; ** p< 0.05; * p< 0.1

Conclusions and Implications

Online review systems represent an influential means by which consumers acquire valuable information about online products. Given the large amount of information available online as well as the potential risk inherent to e-commerce, it is not surprising that a negativity bias exists when consumers process review information. In the past years, we have witnessed a proliferation of studies observing a negativity bias in e-commerce (e.g., Mudambi and Schuff (2010), Sen and Lerman (2007), and Cao et al. (2011)). More specifically, the impact of negativity bias on sales and price premiums has been established previously (Ba and Pavlou 2002; Chevalier and Mayzlin 2006; Pavlou and Dimoka 2006). However, no research to date has examined its underlying causes. Utilizing a novel data set collected from Apple's App Store, we intended to open the black box of the negativity bias by examining three probable mechanisms: review specificity, review surprise, and loss avoidance. The empirical analysis revealed that all three mechanisms (especially the latter two) contribute to the negativity bias in helpfulness perceptions of app reviews.

The theoretical contributions of this research are twofold: First, this paper represents the first attempt to systematically evaluate the underlying mechanisms of negativity bias. The conventional wisdom of negativity bias was confirmed repeatedly in various disciplines and recently in the online word-of-mouth literature, and a number of theories were proposed to account for this phenomenon (Skowronski and Carlston 1989). However, there has been no evidence documenting the existence of these reasons. This paper seeks to deepen our understanding of this universally accepted phenomenon.

Second, based on a large-scale data set, we provided empirical evidence that negativity bias may result from review specificity, review surprise and loss avoidance. The unique data set collected from Apple's App Store enabled us to quantify review specificity and review surprise at the review level, as well as the risk factor at the app level. Despite the intuitive appeal of review specificity, the 'surprise' factor and heightened risk play a more prominent role when consumers process and integrate positive and negative review information.

Given that online merchants rely heavily on online reviews to promote sales (Ghose and Ipeirotis 2011), our research has significant practical implications: Our findings can help review websites to predict the helpfulness of a review based on ratings even before votes are accumulated. Negative reviews are in general more helpful, but negativity bias is diminished and even reversed under certain conditions (e.g., when a review is not surprising or when an app is free). An understanding of underlying causes of negativity bias can help review platforms to design better algorithms to determine review helpfulness in addition to relying on helpfulness votes. Furthermore, our findings can also benefit online companies that want to increase the helpfulness of positive reviews or decrease the helpfulness of negative reviews. According to our results, for instance, online merchants may want to provide more structured guidance concerning how to write a more helpful review about a positive experience: to be more specific, to provide more distinctive information in the reviews, and to discuss the risks associated with products and sellers.

Our study has a few limitations that provide avenues for future research. Our data sample is from Apple's app market, so the generalizability of our findings may be limited to digital products. Future studies may want to sample a larger set of products to test if our results can still hold. Moreover, our measures for review specificity, review surprise, and ability to avoid purchase risk (*paid*) are quantitative proxies rather than direct measures of these variables. Laboratory experiments could be an alternative method to directly measure these variables.

References

Ahluwalia, R. 2002. "How Prevalent Is the Negativity Effect in Consumer Environments?," *Journal of Consumer Research* (29:2), Sep, pp 270-279.

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

Basuroy, S., Chatterjee, S., and Ravid, S.A. 2003. "How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets," *Journal of Marketing* (67:4), Oct, pp 103-117.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., and Vohs, K.D. 2001. "Bad Is Stronger Than Good," *Review of General Psychology* (5:4), pp 323-370.

Birnbaum, M.H. 1974. "Nonadditivity of Personality Impressions," *Journal of Experimental Psychology* (102:3), pp 543-561.

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

Chevalier, J.A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), Aug, pp 345-354.

Coleman, M., and Liau, T.L. 1975. "A Computer Readability Formula Designed for Machine Scoring," *Journal of Applied Psychology* (60:2), pp 283-284.

Feldman, J.M. 1999. "Four Questions About Human Social Behavior," in: *Social Psychology and Cultural Context,* J. Adamopoulos and Y. Kashima (eds.). CA: Sage Publications, pp. 43-62.

Fiske, S.T. 1980. "Attention and Weight in Person Perception - the Impact of Negative and Extreme Behavior," *Journal of Personality and Social Psychology* (38:6), pp 889-906.

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

Gartner. 2010. "Gartner Says Consumers Will Spend $6.2 Billion in Mobile Application Stores in 2010." from http://www.gartner.com/it/page.jsp?id=1282413

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

Greene, W.H., and Zhang, C. 2003. *Econometric Analysis*. Prentice hall Upper Saddle River, NJ.

Helson, H. 1964. *Adaptation-Level Theory*. New York: Harper.

Ito, T.A., Larsen, J.T., Smith, N.K., and Cacioppo, J.T. 1998. "Negative Information Weighs More Heavily on the Brain: The Negativity Bias in Evaluative Categorizations," *Journal of Personality and Social Psychology* (75:4), Oct, pp 887-900.

Kahneman, D., Knetsch, J.L., and Thaler, R.H. 1991. "Anomalies - the Endowment Effect, Loss Aversion, and Status-Quo Bias," *Journal of Economic Perspectives* (5:1), Win, pp 193-206.

Kahneman, D., and Tversky, A. 1979. "Prospect Theory - Analysis of Decision under Risk," *Econometrica* (47:2), pp 263-291.

Korfiatis, N., Rodriguez, D., and Sicilia, M.A. 2008. "The Impact of Readability on the Usefulness of Online Product Reviews: A Case Study on an Online Bookstore," in: *Emerging Technologies and Information Systems for the Knowledge Society, Proceedings,* M.D. Lytras, J.M. Carroll, E. Damiani and R.D. Tennyson (eds.). Berlin: Springer-Verlag Berlin, pp. 423-432.

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

Pavlou, P.A., and Dimoka, A. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation," *Information Systems Research* (17:4), Dec, pp 392-414.

Rozin, P., and Royzman, E.B. 2001. "Negativity Bias, Negativity Dominance, and Contagion," *Personality and Social Psychology Review* (5:4), pp 296-320.

Schupp, H.T., Ohman, A., Junghofer, M., Weike, A.I., Stockburger, J., and Hamm, A.O. 2004. "The Facilitated Processing of Threatening Faces: An Erp Analysis," *Emotion* (4:2), Jun, pp 189-200.

Sen, S., and Lerman, D. 2007. "Why Are You Telling Me This? An Examination into Negative Consumer Reviews on the Web," *Journal of Interactive Marketing* (21:4), Fal, pp 76-94.

Sherif, M., and Sherif, C.W. 1967. "Attitude as the Individual's Own Categories: The Social Judgment-Involvement Approach to Attitude and Attitude Change," in: *Attitude, Ego-Involvement, and Change,* C.W. Sherif and M. Sherif (eds.). New York: Wiley, pp. 105-139.

Skowronski, J.J., and Carlston, D.E. 1987. "Social Judgment and Social Memory - the Role of Cue Diagnosticity in Negativity, Positivity, and Extremity Biases," *Journal of Personality and Social Psychology* (52:4), Apr, pp 689-699.

Skowronski, J.J., and Carlston, D.E. 1989. "Negativity and Extremity Biases in Impression Formation: A Review of Explanations," *Psychological bulletin.* (105:1), pp 131-142.

Vaish, A., Grossmann, T., and Woodward, A. 2008. "Not All Emotions Are Created Equal: The Negativity Bias in Social-Emotional Development," *Psychological Bulletin* (134:3), May, pp 383-403.

Wyer, R.S. 1973. "Category Ratings as Subjective Expected Values - Implications for Attitude Formation and Change," *Psychological Review* (80:6), pp 446-467.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), Mar, pp 133-148.