Anxious and Angry: A Replication Investigating the Effects of Emotions on Perceptions of Online Review Helpfulness

Teresa M. Shaft

Division of Management Information Systems
University of Oklahoma
tshaft@ou.edu

Annie (Chuan) Tian

Department of Information Systems, Statistics, and Management Science
University of Alabama
atian@cba.ua.edu

Sun-young Whang

Division of Management Information Systems
University of Oklahoma
sywh@ou.edu

Abstract:

This study is an exact replication of three studies investigating how emotional content embedded in a product review influences perceptions of review helpfulness (Yin, Bond, and Zhang 2014). The replication confirms that emotional content influences perceptions of review helpfulness. Consistent with the original study, our experimental findings indicate that the relationship between review helpfulness and the emotion embedded in a review is mediated by the perception of reviewer cognitive effort. However, this experiment found that angry as well as anxious reviews were considered helpful, deviating from the original findings in which only anxious reviews were perceived as helpful. Further, examination of a large sample of reviews confirms that reviews with anxious content are rated as more helpful, while angry content does not influence helpfulness. However, we were unable to replicate the findings of the second experiment reported in the original research as neither anxious nor angry reviews were considered helpful in our second study. It may be that, at this time, reviews with high levels of emotional content are not perceived as helpful. Instead, we found that only empathy (perceptions of shared emotion with the reviewer) impacted perceptions of review helpfulness. Therefore, the influence of empathy on review helpfulness may be worth investigating in future studies. That our findings deviate slightly from those of the original study provides further support for the importance of replications.

Keywords: Laboratory Experiments, Psychological, Electronic Commerce

1   Introduction

Online reviews have been an important factor for product sales for some time (Dellarocas 2003) and remain important. In recent studies, 93% of consumers indicate that online reviews impact their purchase decisions (Podium 2017) and 62% consider online reviews very helpful (Clement, 2019). Hence, understanding what qualities of a review make it helpful to other consumers remains of interest to online retailers and consumers. The emotional content of reviews influences consumers' perceptions of review helpfulness (Kuan et al., 2015). Yin et al. (2014) investigated the influence of distinct emotions on perceptions of review helpfulness and how perceptions of reviewers' cognitive effort mediate the relationship between a review's emotional content and perceptions of helpfulness. Given the sustained importance of product reviews and review helpfulness, we chose to replicate their study. Figure 1 shows the theoretical framework from the original study. We investigated the same hypotheses as the original study:

Hypothesis 1: Anxiety-embedded reviews are perceived to be more helpful than anger-embedded reviews.

Hypothesis 2: Perceived cognitive effort mediates the differential impact of anxiety and anger on the perceived helpfulness of reviews.
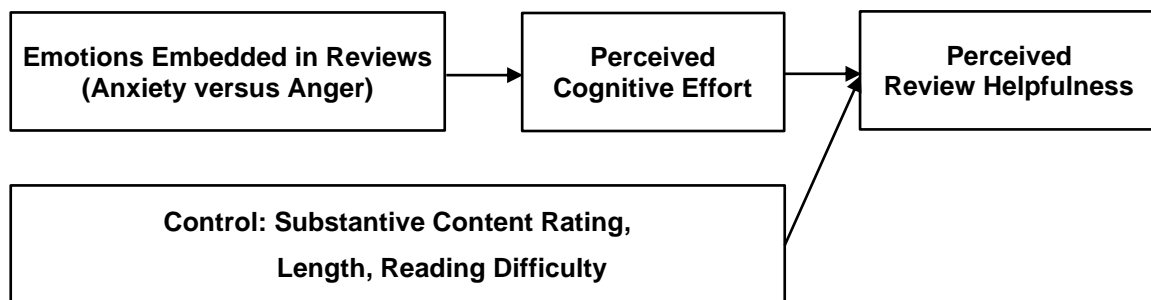

Figure 1. Theoretical Framework

Yin et al. (2014) conducted three studies. Study 1 was an experiment that manipulated conditions of anxiety and anger while controlling for differences in objective review content. Study 2, also an experiment, used a different manipulation to investigate alternative explanations. Yin et al.'s (2014) findings from these studies confirmed both Hypotheses. Study 3 examines seller reviews from an online platform and considers only the first hypothesis which was supported. Yin et al. (2014) relied upon reviews from YahooShopping.com. Instead, we used reviews from SiteJabber.com. Using data from a different seller site creates a stronger replication than if we use the same reviews as Yin et al. (2014). Further, some of the data we required is no longer available on YahooShopping.com.

2   Research Methodology

2.1   Study 1: Experiment

As did Yin et al. (2014), we conducted a laboratory experiment to manipulate anxiety and anger in seller reviews via a repeated measures design that controlled for potential differences in substantial content. The review set included nine reviews that were negatively valenced but modified to express distinct emotions. Each participant read and evaluated six reviews, and provided their perception of review helpfulness and the cognitive effort of the reviewer. We used different reviews than the original study, but otherwise all items (Appendix A) and manipulations are the same as those employed by Yin et al. (2014).

2.1.1   Study 1: Stimulus Materials

We followed the same two steps as Yin et al. (2014) to prepare the stimuli for Study 1: 1) identify text reviews that were negatively valenced but relatively non-emotional (as described in more detail below); 2) added emotional content to represent the manipulations. To complete the first step, we collected reviews from SiteJabber.com merchants in the "electronics" category. We collected all reviews for each merchant that has at least one review on the platform. Our set contained 61,308 total reviews covering 513 merchants compared to Yin et al.'s sample of 154,834 reviews from 167 merchants. To obtain negatively valenced but relatively unemotional reviews for use in our study we followed Yin et al.'s (2014) approach. After obtaining reviews for merchants with at least one review, we next identified reviews with negative valence by applying Yin et al.'s (2014) criteria: reviews with one star out of five possible, a total of 2,352 reviews. As in the original study, we dropped reviews that were extremely short or long and revised those remaining by removing any sentences that directly indicated reviewer emotions, thus creating an initial set of 41 reviews (Yin et al. 2014 used 37 reviews for this step). Then, we selected 14 reviews with content that could have been written by an anxious or angry customer. We conducted a pretest using this set of reviews with 28 participants who rated the perceived anxiety and anger of each review. Following Yin et al.'s (2014) approach, we wished to select three reviews where participants' ratings of the reviews' perceived anxiety and perceived anger were not statistically significantly different. However, our first pre-test failed as all reviews demonstrated a statistically significant difference with regard to perceived anger and perceived anxiety. We selected the 14 reviews for the pre-test based on similarities between the Linguistic Inquiry and Word Count (LIWC) scores for anxiety and anger. In hindsight, it was inappropriate to consider LIWC scores comparable across categories. That is, comparing LIWC scores within a semantic category is reasonable. However, comparing scores across categories implies that a score of 10 on anxiety and of 10 also on anger (for instance) yield an equivalent emotional response from a person. This was not the case in our sample. Therefore, we re-examined our set of reviews and selected 19 reviews and conducted a second pre-test with 41 participants. Based on this analysis, we identified three reviews where perceived anxiety and perceived anger did not differ significantly ($p > .78$). Note that all baseline reviews had LIWC scores of zero for anger and anxiety (Table 1).

	Table 1. Baseline Review Stimuli Used in Study 1		
	Review Content – Replication	Anxiety	Anger
1	Placed an order on 2/15. And I have still not received my order. I didn't see all these poor reviews before placing my order.	6.0	5.8
2	No shipment of product. No response to my inquiries. No service department. But, they have my credit card number in their hands.	7.0	6.9
3	Purchased a battery for my laptop on Sunday. By Tuesday I still hadn't received any confirmation of the order. Both customer service numbers had a recorded message saying they were busy and that if you need to contact them please do it via their website.	5.6	5.5
	Review Content – Yin et al. (2014)		
1	I purchased a camera on February 27 for two-day delivery and on March 23 I am still waiting for it, plus they billed me for it on February 27.	6.7	6.8
2	Ordered a laptop battery (12 cell) and RAM. I received a 6 cell battery and the incorrect RAM. I returned the products to this merchant three weeks ago (and they were received), but still have not received my refund.	6.8	6.8
3	I placed an order on December 14 using standard shipping because it said if I ordered by the 19th it would be delivered by Christmas. I just received an E-mail saying they shipped it today (December 23) and estimated arrival date is December 30.	6.5	6.2

Next, to manipulate emotional expression, we varied the sentence appearing at the beginning of the review using the same phrasing as Yin et al. (2014). For the anger condition, the review began with: "I was very angry after everything that happened." For the anxiety condition: "My experience with this has caused a lot of anxiety." A baseline (control) review contained no up-front sentence. This process was applied to the three reviews yielding nine reviews.

2.1.2   Study 1: Procedure

Yin et al. (2014) conducted a separate study to assess the effectiveness of the emotional manipulations of the reviews prior to Study 1. As this is a replication, we would not alter the phrasing used to create the experimental conditions, therefore we included the manipulation check items in the main study.

Participants were 292 (166 male) undergraduate students enrolled in an introductory Information Systems (IS) course. Our sample is considerably larger than Yin et al.'s (2014) sample of 78. The introductory courses at our university are quite large and all enrolled students needed to be provided an equal opportunity to earn extra credit by completing the experiment. Our sample included 88% from the United States (the same as Yin et al. (2014)), 2% were freshmen, 72% were sophomores, 21% juniors, 6% seniors (compared to 80% juniors). The average age was 20 (compared to 21), and on average they had 12 years' experience using the internet (the same as the original study). Hence our sample was generally consistent with that of the original study.

As did Yin et al. (2014), we administered this study via a Qualtrics survey. Participants were introduced to a fictitious website 'OnlineConsumerReview.com.' Yin et al. (2014) generously provided their materials, hence we relied upon their descriptions for this purpose (see Appendix B). Participants read and evaluated six text reviews, one at a time. As did Yin et al. (2014) three filler reviews were presented in positions 1, 3, and 5 of the sequence of reviews that respondents examined. Our filler reviews are displayed in Table 2. Note that Yin et al. (2014) only provided the contents of one filler review. Therefore, it was necessary to select three filler reviews for our study. We applied the criteria described in the original study: filler reviews were one or two sentences in length and positive overall. We identified 10 five-star reviews from SiteJabber.com. The three authors jointly examined these reviews and selected three reviews that met the criteria. Typographical errors in the original reviews were not corrected to add genuineness. All of the filler reviews (including that of the original study) had LIWC scores of zero for anxiety and anger.

Table 2. Filler Reviews	
Review Content – Replication	**Review Content – Yin et al. (2014)**
1 "One of the best and easiest ordering processes I come across. They keep you updated on the progress of your order."	"I liked their web site – lots of items with a decent description of each. Received exactly what I ordered in a timely manner…"
2 "I love the product and I can't believe I got it in 2 business day tops. And it was well package as well."	
3 "Very easy to order and very helpful and very good products. Shipped them very promptly, had no issues and very flexible payment plan."	

After reading a review, participants reported their perceptions of 1) the helpfulness of the review, 2) the cognitive effort expended by the reviewer, and 3) the manipulation check items. We used the same nine-point semantic differential items as Yin et al. (2014) (see Appendix A).

2.1.3   Study 1: Analysis and Results

We first examined the manipulation check items via repeated measures ANCOVA controlling for the review order and contrasted each review condition (anxious and angry). The reviews in the anxiety condition were more related to anxiety than anger ($M = 8.05$ v. 6.96, $F(1, 286) = 70.31$, $p < .0001$). The reviews in the angry condition were more related to anger than anxiety ($M = 8.32$ v. 5.77, $F(1, 286) = 340.51$, $p < .0001$). The reviews in the baseline condition were more aligned with the angry than the anxious condition ($M = 7.26$ v. 6.07), $F(1, 286) = 78.69$, $p < .0001$ – the difference between the means is 1.19). This result was surprising for two reasons. First, recall that our pre-test subjects rated the reviews as essentially equally angry and anxious. Second, Yin et al.'s (2014) participants rated the baseline review to have similar levels of anxiety and anger ($M = 6.87$ v. 7.17 – a difference of .3). The magnitude of the difference between the two studies is substantial (.3 v. 1.19). Therefore, our statistically significant difference is unlikely to be due to the larger sample in our replication. To investigate this further, we conducted an additional follow up analysis. The baseline review was rated as less angry than the angry review ($M = 8.32$ v. 7.26, $F(5,286) = 19.33$, $p < .0001$) and less anxious than the anxious review ($M = 8.05$ v. 6.07, $F(5,286) = 10.48$, $p < .0001$). This additional analysis provides further evidence that the manipulation of emotional content was successful. Despite this, respondents in our first study perceived the baseline reviews to be more angry than anxious, which is inconsistent with the pre-test participants and the original study. We considered if the elapsed time between the original study and replication could explain the inconsistency. However, recall that our pre-test participants did not rate the reviews as possessing different levels of angry or anxious content. Hence, it seems unlikely that the inconsistency can be explained by the difference in time frames of this and the original study. Therefore, it appears that this inconsistency is attributable to a sampling difference, which reinforces the need to replicate important studies. This difference does not invalidate the Study 1 results as the experimental manipulation was found to be effective. Since our subsequent analyses focus on the angry and anxious reviews and that manipulation was effective, we proceeded with our analysis.

Next, we examined the reliability and validity of the constructs used in the study. Cronbach's alphas for review helpfulness were between .88 and .95 (compared to .93-.95 in the original study). For perceived cognitive effort, the values were between .86 and .95 (compared to .86-.94). As did Yin et al. (2014), we conducted an exploratory factor analysis (EFA). Our results are consistent with Yin et al.'s in that the EFA for each review provided two factors with loadings higher than .7, and higher than loadings on the other factor (< .36). Further, the average variances extracted (AVEs) were above .5 (actually greater than .7). Finally, the square roots of the AVEs were greater than the correlations between them. Together the results demonstrate adequate internal consistency, convergent validity, and discriminant validity.

To assess if the perceived helpfulness of reviews varied across the angry and anxious conditions, we conducted a repeated measures ANCOVA. The pattern of the means for perceived helpfulness in our study and Yin et al.'s (2014) is displayed in Figure 2 and is generally consistent across the two studies. The emotional condition was a within-subject factor, and counterbalancing of the treatment reviews was entered as a covariate. We used specific contrasts to compare the differences in the emotional conditions. The difference in perceived helpfulness between the anxiety and anger conditions was statistically significant ($M = 7.35$ v. 7.02, $F(5, 286) = 10.97$, $p < .0001$) which is consistent with Yin et al.'s (2014) findings ($M = 7.57$

v. 7.23, $t(77) = 2.59$, $p < .05$)[1]. Our smaller p-value is most likely explained by our larger sample size. We also compared the helpfulness of the emotional reviews with the baseline review. The anxious reviews were considered significantly more helpful than the baseline review ($M = 7.35$ v. 6.59, $F(5,286) = 4.46$, $p < .0001$), also consistent with Yin et al.'s (2014) results ($M = 7.57$ v. 7.00, $t(77) = 3.97$, $p < .001$). The angry reviews were also considered more helpful than the baseline ($M = 7.02$ v. 6.59, $F(5,286) = 10.12$, $p < .001$)[2]. This finding is inconsistent with Yin et al. (2014) as their angry reviews were not considered more helpful than baseline reviews ($M = 7.23$, v. 7.00, $t(77) = 1.42$, $p = .16$). Note that the difference between the means of helpfulness of angry and baseline reviews in our study was 0.43, nearly double the difference reported in the original study. This is particularly interesting given that the baseline reviews were rated as more angry than anxious, yet reviews manipulated to be angry were perceived as more helpful than the baselines. Hence, unlike the original research, in our Study 1 both angry and anxious negative reviews were considered more helpful than the baseline reviews. We return to this discrepancy in our discussion.
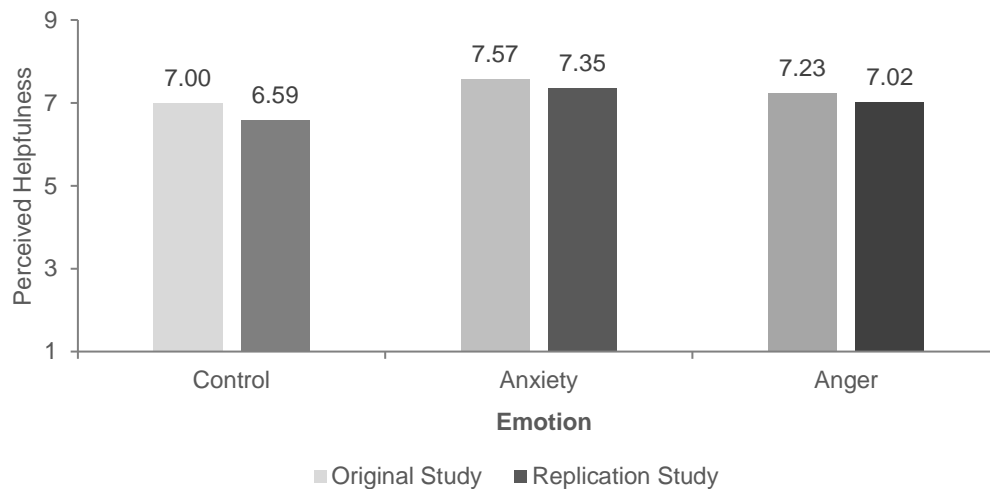

Figure 2. Perceived Helpfulness of Seller Reviews Across Emotion Conditions in Study 1

As did Yin et al., we next conducted a mediation analysis employing Judd et al.'s (2001) procedure for within-subjects designs. First, we confirmed that the independent variable (i.e., discrete emotion) was significantly related to the dependent variable (perceived helpfulness) and the mediator (perceived effort). The above analysis confirms that perceived helpfulness was greater for the anxiety condition than the angry condition. A repeated measures ANCOVA confirmed that the difference was also obtained for perceived effort ($M = 6.41$ v. 5.92, $F(5,286) = 39.96$, $p < .0001$). The means of perceived effort from this study and the original are quite consistent (see Figure 3). Next, we confirmed that the proposed mediator (perceived effort) was statistically dependent on both levels of the independent variable. Consistent with Yin et al. (2014), a greater perceived effort was associated with greater perceived helpfulness for both the anxiety and angry conditions ($\beta = 0.58$ and 0.60, $t = 12.30$ and 12.80, $p < .0001$). Finally, we regressed the difference in perceived helpfulness across the anxiety and anger conditions on three terms: 1) the difference in perceived effort across anxiety and anger conditions; 2) the sum of perceived effort across anxiety and anger conditions (mean-centered), and 3) an intercept term. The analysis confirmed the presence of mediation: differences in perceived effort predicted differences in perceived helpfulness ($\beta = 0.52$, $t = 1.41$, $p < .0001$). The coefficient for the intercept was not significant ($\beta = -0.20$, $t = -0.42$, $p = 0.62$), indicative of full mediation. Thus, our findings are consistent with the original study and support the hypothesis that the impact of anxiety and anger on perceptions of review helpfulness is mediated by cognitive effort.

---

[1] Yin et al. (2014) conducted t-tests. We were able to specify contrasts, hence we report F-values.
[2] A paired t-test is also statistically significant ($t = 3.96$, $df = 292$, $p < 0.0001$)
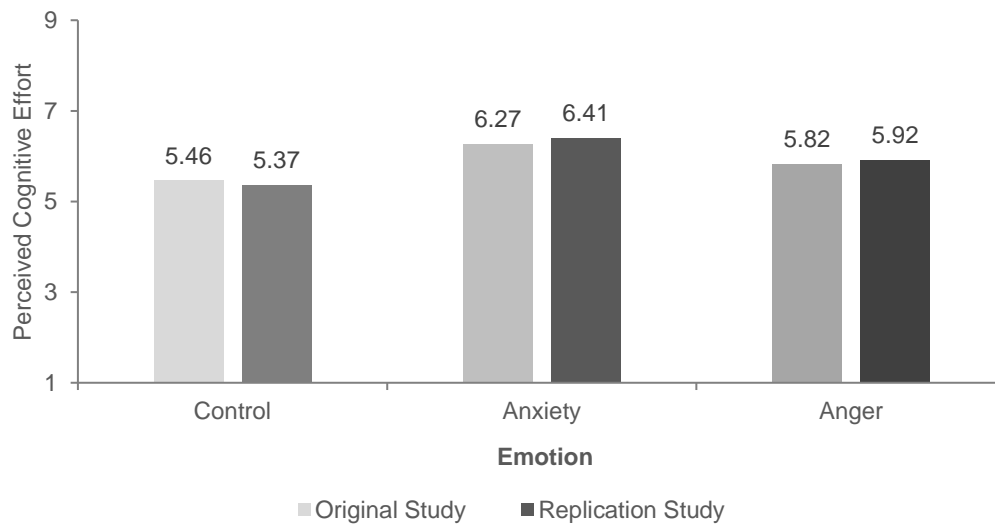
Figure 3. Perceived Cognitive Effort of Seller Reviews Across Emotion Conditions in Study 1

2.2 Study 2: Experiment

The purpose of the second experiment is to investigate possible alternative explanations for the results of Yin et al.'s (2014) Study 1. We employ the same between-subject design as Yin et al. (2014). This design holds the review contents constant and appends emotional content to the beginning and end of each review creating a stronger manipulation of emotion than in Study 1. Participants in this study also provided assessments of valence, arousal, attribution, and empathy. These additional measures allow us to consider the impact of these additional issues on perceptions of anger and anxiety.

2.2.1 Study 2: Stimulus Materials

This study used the same cover story and a similar procedure as Study 1. However, it required only one review from Study 1 (review 3, see Table 1). The emotional manipulation was strengthened by appending sentences to both the beginning and end of each review: "I feel so worried (mad) as I'm writing this!" to the beginning and "Let me tell you: I'm very nervous (irritated)." to the end. These additional sentences are identical to those employed in the original study. The worried/nervous combination created the anxious condition; the mad/irritated combination created the angry condition. We used the same measures as Study 1, but also asked participants to evaluate the valence, arousal, attribution and empathy of the reviewers using the same items as Yin et al. (2014) (see Appendix A).

2.2.2 Study 2: Procedure

We recruited 86 undergraduate participants (compared to 73), 52% male (compared to 47%). Our participant's average age was 20 (compared to 21) with 12 years of internet experience (compared to 11) and 82% were in their sophomore or junior years (compared to 64%). Participants were randomly assigned to either the *anxiety* or *anger* condition. Participants evaluated one review and responded to the items regarding helpfulness, perceived cognitive effort of the reviewer, valence, arousal, and empathy as well as the items used in the emotion manipulation check (see Appendix A).

2.2.3 Study 2: Analysis and Results

We investigated the manipulation check items first. The review in the anxiety condition was considered more closely related to anxiety than anger ($M = 7.60$ v. 5.00, $t(41) = 7.98$, $p < .0.0001$). The review in the angry condition was more closely related to anger than anxiety ($M = 7.89$ v. 5.20, $t(43) = 7.50$, $p < .0001$). These results confirm the effectiveness of the manipulations, mirroring the original study.

Following Yin et al.'s (2014) process, we conducted an ANCOVA to examine the perceived helpfulness of the anxiety-embedded and anger-embedded reviews controlling for valence and arousal. However, this model was not statistically significant ($p = .32$). Examination of the means for helpfulness in the anxious and angry conditions revealed that they were nearly identical ($M = 5.95$ v. 5.93). For this sample, the reviews in

both emotional conditions were rated as similarly helpful. This result is inconsistent with Yin et al.'s (2014) Study 2, which supported the findings of Study 1 – that anxious reviews were perceived as more helpful than angry reviews (M = 7.33 v. 6.26, $F(1,69)$ = 5.67, p < .05). Although the mediation test that Yin et al. (2014) conducted would no longer be appropriate, we investigated if the emotional conditions were predictive of cognitive effort. However, the ANCOVA for cognitive effort (controlling for valence and arousal) also did not achieve statistical significance (p = .37) and the means for cognitive effort for the anxious and angry reviews are quite similar (M = 4.80 v. 5.02, t = 84, p = .48). Again, our findings for Study 2 are inconsistent with those of Yin et al.'s (2014) who found that perceptions of cognitive effort were higher for reviews in the anxious condition compared to the angry condition (M = 4.84 v. 3.83, $F(1,69)$ = 5.23, p < .05)

We were curious if the other perceptual measures (empathy and attribution) could provide additional explanatory power. An ANCOVA with helpfulness as the dependent variable, emotional content as the between-subjects factor and valence, arousal, attribution, and empathy included as covariates did yield a significant model (p = .01). However, empathy was the only statistically significant predictor (p = .001). The emotional condition of the review (anxious or angry) remained insignificant (p = .64). In our study, how much empathy (shared emotion) the participant felt with the reviewer was the only predictor of helpfulness regardless of the emotional content of the review.

Consistent with the analyses presented in the original study, next we tested for alternative explanations for the findings of Study 1. First, we considered the possibility of negativity bias in the reviews. Consistent with Yin et al. (2014), the review in the anxious condition was considered less negative than the review in the angry condition (M = 3.23 v. 2.10, $t(84)$ = 3.01, p = .004), consistent with Yin et al. (2014): M = 2.38 v. 1.41, $t(71)$ = 4.20, p < .001). Both sets of results refute a general negativity bias as an explanation for the findings of Study 1.

Examination of the arousal measure revealed that arousal in the anxious condition was less than in the angry condition (M = 5.44 v. 7.21, $t(84)$ = -5.14, p < .0001), consistent with the original study (M = 6.65 v. 8.32, $t(71)$ = -6.47, p < .001). Yin et al. (2014) argued that their finding suggested that emotional arousal could explain the differential impact of anxiety and anger. However, since our Study 2 did not detect a difference between the two conditions, emotional arousal may operate independently of the specific emotional conditions.

Analysis of the attribution measures revealed that dispositional attributions were lower in the anxious condition (M = 4.0 v. 5.0, $t(84)$ = 2.50, p = .01). The effect of dispositional attributions was stronger in this study than in the original where the difference was marginally significant (M = 5.03 v. 3.97, $t(71)$ = 1.81, p = .07). Finally, consistent with the original study, the measure of empathy did not differ across the two conditions (M = 4.22 v. 4.38, $t(84)$ = 0.40, p = .69), compared to $t(71)$ = 0.67, p > .05) .

2.3   Study 3: SiteJabber.com Merchant Reviews

The primary goal of Study 3 is testing Hypothesis 1, the effects of discrete emotions on review helpfulness in a real world setting. As noted above, we based this analysis on the reviews scraped from the "electronics" section of SiteJabber.com. SiteJabber.com shares similarities with YahooShopping.com relied upon by Yin et al. (2014) in that they both provide user ratings and text reviews to rate and describe the overall quality of online merchants. The use of different review websites forms a stronger replication, as these reviews have similar characteristics as the original study but were obtained from a different source at a different time.

SiteJabber.com has accumulated data on ratings and reviews on businesses since 2008. SiteJabber.com was developed in part with a grant from the National Science Foundation, (https://nsf.gov/awardsearch/showAward?AWD_ID=1127567).   Consistent   with   YahooShopping.com, SiteJabber.com also allows a merchant's customers to evaluate their experience using a five-star rating. These ratings are accompanied by optional text reviews with details on their experience with the merchant, namely, their service orientation and the quality of products they sell.

2.3.1   Study 3: Data Collection

Individual reviews serve as the unit of analysis and we conducted our initial scraping on April 30, 2019. This process created an initial sample consisting of 61,308 reviews. Following Yin et al. (2014), we collected 1) ratings, 2) the contents of text reviews, 3) helpful votes, and 4) total votes for each review. We also collected store-level information: 5) overall average ratings and 6) count of all ratings for each store (popularity). For noise reduction purposes, we adopted the same criteria as Yin et al. (2014). Specifically, we removed

reviews that; included non-ASCII characters, had no textual content, contained only EOM ("End of Message"), or contained only symbols or dates. From this group, we selected reviews that received at least one helpfulness vote for our data analysis. Out of the 61,308 reviews, only 3,519 reviews had received helpfulness votes, accounting for 5.74% of the original sample. This is slightly higher than the 4.7% reported by Yin et al (2014). We conduct the subsequent analysis based on this sample of 3,519 reviews.

2.3.2   Study 3: Variables

Review helpfulness was operationalized as follows. For each review SiteJabber.com presents the question "Helpful?" adjacently aligned with a "Yes" button that can be clicked only once, followed by a numbered bubble that indicates the total number of helpfulness votes the review has earned. Yin et al. (2014) measured helpfulness using the proportion of helpful votes out of the total votes a review received. They accomplished this by dividing the number of helpful votes by the total number of votes regarding a particular review. However, SiteJabber.com records only the number of helpfulness votes that a review receives. Therefore, unlike YahooShopping.com, consumers cannot vote to indicate that a review is explicitly not helpful[3]. Therefore, we needed to use a different measure for helpfulness than that used in the original study. The original study focused on a particular review as the unit of analysis. Hence, we needed a similar focus and for our dataset the most appropriate parallel measure is the count of the number of helpfulness votes each received[4]. As such, our measure of helpfulness is a count variable rather than a percentage. Since there are no limits to the maximum number of helpful votes a review can earn, reviews with a greater number of helpfulness votes are considered more helpful. Figure 4 provides a screenshot of a merchant review in SiteJabber.com. Note that helpfulness is the sole measure that is operationalized differently than the original study.
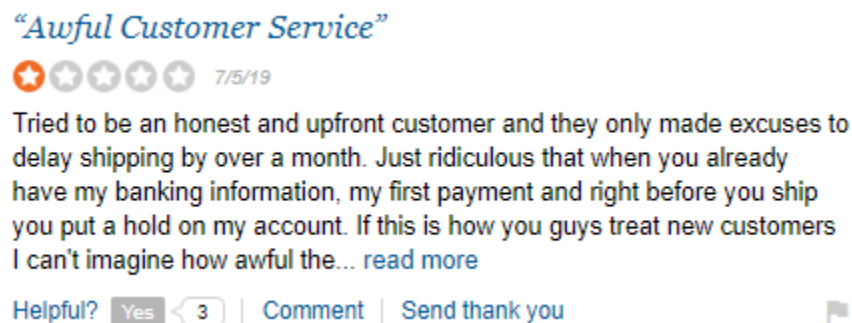

Figure 4. Screenshot of a SiteJabber Merchant Review

Consistent with Yin et al. (2014), we used LIWC to obtain measures of anxiety and anger in each text review. In our sample, the highest values for anxiety and anger are 17.14 and 10.53 respectively, and the average values for anxiety and anger are 0.1501 and 0.1505 respectively. Out of the 3,519 reviews, 300 (8.52%) contained at least one word defined by LIWC as anxiety-embedded, and 354 (10.06%) included words defined by LIWC as anger-embedded. 76 reviews contained both anxious and angry words, which accounted for 2.16% of all reviews. The descriptive statistics and bivariate correlations for the dataset are presented in Tables 3 and 4, respectively. Note that both are quite consistent with Yin et al. (2014) except for the variable *review helpfulness*. Recall, we operationalized review helpfulness as the count of the helpful votes (ranging from 1 to 59). In the original study it is operationalized as the number of helpful votes divided by the number of the total votes ranging between 0 and 1. The mean and standard deviation for rating are 3.63 and 1.71 (compared to 3.29 and 1.81 in original study), for length are 68.9 and 77.36 (compared to

---

[3] It is worth noting that we distinguish the actual helpful votes per review against the "# of helpful votes" that appears below the review writer's ID because that merely indicates the number of helpful votes that ID holder earned since joining the site. Clicking the "Yes" button does not require reviews to hold an ID for SiteJabber.com.

[4] Note that measures that would be normalized by, for instance, the number of reviews for a particular store would evaluate a review with the same number of helpfulness scores quite differently depending upon the number of review as the range for the number of review for stores ranges from 1 to 6,250.

69.82 and 70.76), for reading difficulty are 8.18 and 3.25 (compared to 10.32 and 4.25), etc. We provide examples of anxiety-embedded and anger-embedded reviews in Table 5.

	Table 3. Descriptive Statistics (N=3,519) in Study 3					
	Variable	Observations	Mean	Std. Dev.	Min	Max
1	Review Helpfulness	3,519	2.53	4.01	1	59
2	Rating	3,519	3.63	1.71	1	5
3	Length	3,519	68.90	77.36	2	1,191
4	Reading difficulty	3,519	8.18	3.25	-15.9	45.5
5	Anxiety	3,519	0.15	0.72	0	17.14
6	Anger	3,519	0.15	0.64	0	10.53

	Table 4. Variable Correlations (N=3,519) in Study 3					
	Review Helpfulness	Rating	Length	Reading Difficulty	Anxiety	Anger
Review Helpfulness	1					
Rating	-0.2693*	1				
Length	0.2577*	-0.4067*	1			
Reading Difficulty	-0.0563*	0.1388*	-0.1121*	1		
Anxiety	0.0496*	-0.1147*	0.0083	0.0251	1	
Anger	0.0513*	-0.2266*	0.0907*	-0.0472*	0.0416*	1

	Table 5. Examples of Emotion-embedded Reviews at SiteJabber.com	
	Anxiety-embedded Reviews	Anger-embedded Reviews
1	I was told that my package would be shipped on Monday and i follow up with UPS and they said that they never got the package!	I wasn't made aware of everything I needed before the process started and the shipping company lost my laptop. Found the sales rep very rude on the phone
2	Every time we talked on the phone, I felt I was being misled. It took me over 10 phone calls. I don't know how many emails every I called I felt like I was getting nowhere with you. If I didn't know better, I would think you we're trying to steal my identity.	They charge a ridiculous fee for shipping thru a company named "Same Day Shipping" then tell you it's going to be 5-10 business days for delivery. Gouging at its finest.
3	After I placed the order, I got an email stating that my stuff out of stock.! and I need to wait 30 days till they full fill it. But the same item was continuously selling at the website no problem! Customer service doesn't know what's going on, they like from the other planet. Now I'm worry that somebody will use my credit card as some people say.	Company SOLD me TRASH, Scuffed and Marred. I advised that I wanted to RETURN. Company CHARGED me RE-STOCK Fee of $8, and it Cost $10 to ship back. Advised that I was a POLICE OFFICER....BUSINESS AS USUAL! RIPOFF!!

Following Yin et al. (2014), we conducted the analysis controlling for the same set of variables: the review rating, rating squared, review length, message reading difficulty, and the store characteristics including store reputation and store popularity. The average review ratings (number of stars) in our sample was 3.63 (compared to 3.29 in Yin et al., 2014). The average review length in our dataset was 68.90 (compared to 69.82). We also computed the Coleman-Liau Index and the average reading difficulty of our dataset was 8.18, approximately 2 grades lower than the original study. As did Yin et al. (2014), our analysis also controlled for the potential effect of store characteristics, the average store rating, and store popularity (the total count of all reviews as the store rating). There are 514 stores in our dataset, and the average store reputation is 3.68 out of 5, and the average popularity is 1229.3.

2.3.3 Study 3: Analysis and Results

We conducted a Poisson regression to analyze the effect of discrete emotions on review helpfulness. This model is more appropriate for our data than the Tobit regression reported by Yin et al. (2014) for two reasons. First, recall that Yin et al.'s (2014) dependent variable, helpfulness, was a percentage variable ranging from 0 to 1; and our indicator of helpfulness is measured by the total number of helpfulness votes each review received. Hence, it is a count variable, ranging from 1 to infinity (theoretically). Second, Poisson regression is appropriate when analyzing count data. Poisson regression presumes the mean and variance of the outcome variable to be similar (Gardner et al., 1995; Harris et al., 2012). We computed the mean and variance of the helpfulness variables in our dataset (mean = 2.53, and variance = 2), and found the distribution characteristics indicated that the distribution was slightly under dispersed. Thus, Poisson regression was the appropriate analytical approach (Gardner et al., 1995).

Table 6 presents the outcome of the Poisson regression. As shown below the table, the overall outcome demonstrates a fairly good fit with a highly significant likelihood ratio ($p < 0.001$) and the pseudo R^2 value of 0.118. Further, the Poisson regression outcomes regarding the control variables demonstrate a strong alignment with Yin et al (2014). The review rating (coefficient = -0.558, $p < 0.001$ v. -1.925 and $p < 0.001$ in the original study) and its squared term (coefficient=0.065, $p < 0.001$ v. 0.246 and $p < 0.001$) are significant and in the anticipated direction. Specifically, a lower review rating is perceived to be more helpful, a greater reading length is considered to be more helpful (coefficient = 0.002, $p < 0.001$), and a higher level of reading difficulty reduces helpfulness. Further, store popularity is negatively related to helpfulness, consistent with Yin et al. (2014). Store reputation was not related to helpfulness, which diverges from the original study's finding that it reduced perceived helpfulness.

Table 6. The Poisson Regression Analysis Result (N = 3,519, DV: Review Helpfulness)				
Helpfulness	**Coefficient**	**Standard Error**	**Z-value**	**Significance**
Constant	2.038	0.069	29.6	0.000
Rating	-0.558	0.052	-10.67	0.000
Rating2	0.065	0.009	7.54	0.000
Length	0.002	0.000	17.47	0.000
Reading Difficulty	-0.015	0.004	-3.88	0.000
Store Reputation	-0.017	0.014	-1.23	0.220
Store Popularity	-0.000	0.000	-14.38	0.000
Anxiety	0.042	0.012	3.52	0.000
Anger	-0.024	0.016	-1.5	0.135
*Log likelihood = -8263.4 **Likelihood Ratio = 2210.80 ($p = 0.000$, df =8) ***Pseudo R^2 = 0.1180				

To test the first hypothesis, we examine the coefficients and significance level associated with anxiety and anger. As shown in Table 6, anxiety is significantly associated with a positive coefficient (coefficient = 0.042, $p < 0.001$), whereas anger is associated with a negative coefficient (-0.024) but is not statistically significant ($p = 0.135$). The outcomes indicate that a more anxious review is perceived to be more helpful, while a review with a higher level of anger has no impact on perceived helpfulness.

Furthermore, we compared the nested-model fitness indices by adding anxiety and anger to the baseline model specifying only control variables. As shown in Table 7, after adding the anxiety variable (as in Model 2), the model's log-likelihood increased by 5.4 points, which indicates a significant increase in model fitness. However, when we added the anger variable to Model 2 (as in Model 3), the log-likelihood increased by only 1.167, showing an insignificant increase. Therefore, we conclude that anxiety is associated with a stronger positive effect on review helpfulness than anger, and the first hypothesis is well-supported.

Helpfulness	Model 1 (Baseline)	Model 2	Model 3
Constant	2.002	1.990	2.003
Review Rating	-0.569	-0.564	-0.567
Rating2	0.066	0.066	0.066
Length	0.002	0.002	0.002
Reading difficulty	-0.014	-0.015	-0.015
Store popularity	0.000	0.000	0.000
Anxiety		0.041	0.041
Anger			-0.024
Log Likelihood	-8270.731	-8265.334	-8264.168

Table 7. Fitness Comparison with the Baseline Model

3   Discussion

Thanks to the generosity of Yin et al. (2014), who provided us with the 'cover' story used to motivate the review contents, we were able to conduct a replication that altered the contents of reviews but otherwise was an exact replication of their study. While our findings were generally consistent with those of the original study, there were some surprising inconsistencies. The results of our replication of the first experiment confirmed the main findings of the original study (Yin et al., 2014). As in the original, anxious reviews were considered more helpful than angry reviews and the relationship between the emotional content and the perceived helpfulness of a review was fully mediated by perceptions of the cognitive effort a reviewer exerted when writing the review. Hence, our Study 1 results confirmed both hypotheses. However, both angry and anxious reviews were considered more helpful than baseline reviews, whereas in the original only anxious reviews were rated as helpful. It may be that since the time of the original study, users of product reviews have become more accepting of angry content in reviews. Despite this, anxious content was rated as more helpful than angry as in the original study.

Our replication of Study 2 obtained results that were less consistent with the original study. Despite effective manipulations, a review's emotional condition was unrelated to perceived helpfulness. Not only was there no effect for emotional condition (anxious or angry), the model was not statistically significant. Additional analyses revealed that, in our sample, empathy (perception of shared emotion) was the sole predictor of perceived helpfulness. Interestingly, general perceptions of empathy were consistent across the two emotional conditions. Hence, the embedded emotional content did not influence participants' feelings of empathy. It may be fruitful for future research to further investigate empathy's influence on review helpfulness and what type of review content leads to a feeling of empathy in those who read reviews.

When we consider the findings of Study 1 with Study 2, recall that in the first study reviews with either type of emotional content was more helpful than the baseline review. However, participants in Study 2 rated neither the anxious nor the angry reviews helpful. Study 2 utilized stronger manipulations of the emotional content (sentences appended to the front and back of each review). Hence, the results of Studies 1 and 2 together appear to support the notion of a curvilinear relationship between the anger/anxiety emotional content and helpfulness. Study 1 employed a weaker manipulation of emotional content and participants rated both anxious and angry reviews as helpful. Study 2 employed stronger manipulation and neither type of emotional review was rated as helpful. A recent study detected a curvilinear effect of emotional content (measured as expressed emotional arousal) on perceptions of review helpfulness (Yin et al., 2017). It is possible that since the time of the initial study, users have become more sophisticated in their assessment of emotional content of reviews so that some emotional content (in our study, anxious and angry content) is helpful, but when too much is included in a review it is no longer helpful. Our findings regarding Study 3 were highly consistent with those of the original study. Our investigation of a large sample of reviews scraped from SiteJabber.com found that anxiety was positively related to review helpfulness, while anger did not impact ratings of review helpfulness. The results for Study 3 diverged from the original study in one minor way. In our study, store reputation was unrelated to helpfulness while in the original study it was associated with lower levels of helpfulness. However, this shift did not unduly influence our results which support the original authors' first hypothesis – that anxiety-embedded reviews are perceived as more helpful than anger-embedded reviews. This confirmation is particularly notable as we extracted reviews from a different review site and time period than that was relied upon by Yin et al. (2014). Hence, that anxious but not angry reviews

were associated with perceptions of review helpfulness is not an artifact of a particular review site or a specific time period, but instead, appears to be a robust finding. When we consider Study 3 in the context of the findings of Studies 1 and 2, our argument that more recent users of online reviews are more accepting of angry content in reviews may still apply. SiteJabber.com contains reviews and ratings accumulated since its inception. If it is a relatively recent phenomenon to find angry reviews helpful, it is likely that would not be reflected over the entire history of SiteJabber.com.

In summary, our replication found confirmation for both of the hypotheses investigated in the original study. In addition, we detected some differences: in addition to anxious reviews, angry reviews were now also considered helpful in some situations. Further, our findings point to the possibility that there exists now a curvilinear relationship between the amount of anxious or angry emotional content in a review and perceived review helpfulness.

Acknowledgments

References

Clement, J. (2019). Topic: Online reviews. *Statista.* Retrieved from https://www.statista.com/topics/4381/online-reviews/.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407-1424.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, *118*(3), 392-404.

Harris, T., Yang, Z., & Hardin, J. W. (2012). Modeling underdispersed count data with generalized Poisson regression. *The Stata Journal*, *12*(4), 736-747.

Kuan, K. K., Hui, K. L., Prasarnphanich, P., & Lai, H. Y. (2015). What makes a review voted? An empirical investigation of review voting in online review systems. *Journal of the Association for Information Systems*, *16*(1), 48-71.

Podium. (2017). 2017 State of online reviews. *Podium.* Retrieved from http://learn.podium.com/rs/841-BRM-380/images/2017-SOOR-Infographic.jpg.

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539-560.

Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, *54*(3), 447-463.

Appendix A: Variables

All use the nine-point semantic differential items (1 = 'not at all' and 9 = 'very much').

Manipulation check:

In your opinion, to what extent does each of the following words describe how the reviewer felt when he/she wrote the above review?

- anxious
- angry
- sad
- happy

Helpfulness:

Using the scales below, how would you describe the above consumer review?

- helpful
- useful
- informative

Perceived cognitive effort of reviewers:

- In your opinion, how much effort had the reviewer put into writing this review?
- In your opinion, how much thought had the reviewer given to the above review when he/she wrote it?
- In your opinion, how much time did the reviewer spend writing this review?

Valence:

Overall, how would you describe the above customer's feelings regarding the experience he/she wrote in the review above?:

- very bad/very good
- very unfavorable/very favorable
- very unpleasant/very pleasant

Arousal:

Using the scales below, how do you think the reviewer was feeling at the time he/she wrote the review above?:

- very passive/very active
- very mellow/very fired up
- very low energy/very high energy level

Attribution About the Reviewer:

There are a wide variety of reasons that customers might write a store review. Rate the extent to which you agree with the following statement:

- The cause of the review was something about the reviewer.

Empathy:

- While reading this review, to what extent did you feel like you were experiencing the same emotions as the reviewer?
- While reading this review, to what extent did you feel concerned for the reviewer?
- While reading this review, to what extent did you feel moved by the review?

Appendix B: Cover Stories

Study 1 Cover Story

#1: Instructions
Important! Please read the following first:
Because online shopping involves uncertainty and risk, there are a number of third-party review sites that provide consumer ratings and reviews of online stores.

One of these sites, OnlineConsumerReview.com provides the ability for customers to write a detailed text review about an online store they have recently dealt with. Reviewers are asked to evaluate the store based on their own purchasing experience.
Continued on the next screen.

#2: Instructions - Continued
We have been working with OnlineConsumerReview.com to help them improve their data mining algorithms. In order to achieve this goal, potential consumers' opinions about real text reviews are needed. In this task, we would like your help in evaluating an assortment of text reviews collected from real review websites. Specifically, you will be asked to evaluate various characteristics of the reviews and reviewers.

NOTE: You will be reading and evaluating the text reviews one at a time. Each review is describing a DIFFERENT online store. Please make sure to read the entire text review carefully before rendering a judgment.
(This task starts on the next screen. In total, you will see 6 text reviews.)

Study 2 Cover Story

#1: Instructions
Important! Please read the following first:
Imagine that you are shopping online for a digital camera. You have already decided on the specific model that you are interested in, and browsed a number of different online electronics retailers that offer this model. After considering various factors relevant to your decision (price, shipping, etc.), you have tentatively selected one store: DigitalOnline.

DigitalOnline has been in business for a long time, but you have not heard of this store before. In order to make a decision about whether to buy from DigitalOnline, you would like to find out what past customers have thought about their experience with this store. Therefore, you visit a third-party website, OnlineConsumerReview.com.
Continued on the next screen.

#2: Instructions - Continued
OnlineConsumerReview.com provides real, detailed consumer reviews of online stores. Reviewers are able to write a detailed text review about an online store based on their own purchasing experience.

On the following screen, you will be shown the most recent text review for DigitalOnline. Then you will be asked about your opinion of the review. Please make sure you read all of the available information on the page before making your evaluation.
You may begin when ready.

About the Authors

Teresa Shaft is an Associate Professor of Management Information Systems at the University of Oklahoma. She received her Ph.D. from the Pennsylvania State University. Her research interests include the cognitive processes used by IS professionals during system development and maintenance, the role of IS in environmental management and obtaining value from IT investments. Her work has been published in *Information Systems Research*, *Management Information Systems Quarterly*, *Journal of the Association of Information Systems*, *Journal of Management Information Systems*, and *Association for Information Systems - Transactions on Human-Computer Interaction.* She is a co-founder of IS-CORE, a special interest group of AIS. Her research has been supported by grants from the National Science Foundation.

Annie (Chuan) Tian is an assistant professor of Management Information Systems at the Culverhouse College of Business School of University of Alabama. She is currently a Ph.D. Candidate in Management Information System at the University of Oklahoma. Before the Ph.D. program, she received a master's degree in Management Information Technology from the University of Oklahoma and M.B.A. from Shanghai Jiao Tong University. Her research interests include information security, IT business strategy, IT innovation and Diffusion, and computer-mediated communication.

Sun-young Whang is a doctoral student in the Division of Management Information Systems at the University of Oklahoma. She holds a bachelor's degree in Economics and her Master of International Studies degree in International Trade and Finance from the Graduate School of International Studies, both from Yonsei University in South Korea. Before joining the Ph.D. Program, she created and operated the KUBS Worldwide Business Research Rankings at the Center for Business Research Analytics of Korea University Business School. She also has experience running KOICA government training programs on LDC's economic development at a private research and consulting institute. Her research interests include IT identity, social media anxiety, and diffusion of innovation.