

Is the Whole Equal to the Sum of its Parts? Exploring the Impact of Inconsistency on Perceived Helpfulness of a Set of Reviews

Dezhi Yin, Triparna de Vreede, Logan Steele, Gert-Jan de Vreede
University of South Florida
{dezhiyin, tdevreede, lmsteele, gdevreede}@usf.edu

Abstract

Despite a growing literature on the helpfulness of individual reviews, scant attention has been paid to the helpfulness of a set of reviews. Moreover, it is generally assumed that the helpfulness of a review set can be derived from averaging the individual reviews' helpfulness evaluations. Drawing on bounded rationality theories, we hypothesize that this assumption may not always be valid, and that consumers' helpfulness perception of a review set is also determined by inter-review consistency. A carefully designed experiment revealed that the whole differs from the sum of its parts when the reviews are consistent with each other or are contradicting each other, but does not differ when the reviews contain mixed, non-contradicting opinions. These findings deepen our understanding of inconsistency in online reviews and have a potential to change how reviews should be presented to the consumers.

1. Introduction and Motivation

Despite the increasing popularity and importance of online reviews for consumer decision making, the sheer amount of available reviews requires consumers to be selective and pay attention to only the most helpful ones [1]. To meet this challenge, almost all review platforms allow review readers to vote on the helpfulness of an individual review, display the helpfulness score, and bring the most helpful reviews to the forefront. A growing literature has examined diverse factors that influence consumers' helpfulness perception of an individual review, including rating, reviewer characteristics, and review content [e.g., 2, 3, 4].

However, a single review is rarely sufficient for consumers to make a purchase decision, and they routinely read multiple reviews in a short period of time. Despite its relevance, we are not aware of any study that systematically investigates the antecedents of consumers' evaluation of the helpfulness of a *set* of

reviews (which we call a *review set*). An implicit assumption in both academic research and practice is that the perceived helpfulness of a review set could be easily derived from averaging the helpfulness of individual reviews. For example, Amazon prominently displays three to six most helpful reviews of a product based solely on the helpfulness score of individual reviews, with the assumption that the whole is equal to the sum of its parts. However, the whole may have an independent existence from its parts [5]. In the case of reviews, when consumers form an overall helpfulness assessment of a review set, their assessment may also be influenced by the relationships among the individual reviews. If this is indeed the case, the current emphasis of nearly all review platforms on identifying and promoting the reviews that are individually very helpful may be misguided and ineffective in actually helping consumers with their purchase decisions.

Moreover, the role of inconsistent opinions in shaping helpfulness evaluations has only been studied in light of a single review. Yet, inconsistency is a much richer concept when examined at the level of a review set. Prior research has shown that, within a single review, the presence of mixed opinions or two-sided arguments generally enhances review credibility and helpfulness [6, 7]. However, it is unclear whether consumers prefer opposing opinions among multiple reviews. For instance, inconsistency among different reviews could manifest as a direct contradiction about the same attribute of a product, which rarely occurs within a single review. Despite the prevalence of direct contradiction among real-world reviews, to our knowledge, no research has looked into its implications. As a result, little guidance exists in terms of how review platforms and product manufacturers should deal with a set of prominently displayed reviews that are often mixed in their opinions or even contradictory.

In this study, we explore *how and why inconsistency among reviews in a review set influences consumers' helpfulness perception of the review set*. Drawing on bounded rationality theories, we hypothesize that greater inconsistency among a set of reviews reduces consumers' helpfulness perception of the review set, and that this negative effect can be explained through decreased credibility, increased

confusion, and a lower sense of information being complete. We conducted a controlled experiment in which we manipulated inconsistency of a review set at three levels: no inconsistency, inconsistency across product attributes, and inconsistency regarding the same attribute (i.e., direct contradiction). The results provided support for our predictions. Interestingly, we found that overall helpfulness evaluation of a review set is higher than the average of individual review helpfulness when the reviews are consistent, lower than the average when the reviews are contradicting each other, and not significantly different from the average when the reviews include inconsistent but non-contradicting opinions. These findings challenge the implicit assumption of the whole being equal to the sum of its parts, deepen our understanding of the role and impact of inconsistency in online reviews, and offer important implications for review platforms, product manufacturers, and reviewers.

The remainder of this paper is structured as follows. The next section presents the development of our research model and hypotheses. The design of our laboratory experiment and the results are presented in the subsequent two sections. We conclude the paper with a discussion of our findings, their practical and theoretical implications, and the study's limitations and future research directions.

2. Model Development and Hypotheses

Theories that examine the role of information content and context on how messages are perceived have been valuable in explaining and predicting consumer behavior. Specifically, theories that assume a bounded nature to rationality, like the prospect theory, have been effective in the field of persuasive messaging and subsequent behaviors [8, 9].

Prospect theory [10] is an extension of the theory of bounded rationality. It states that people have systematically different preferences depending on how their options are framed in terms of gains and losses. Specifically, people use cognitive shortcuts (i.e., heuristics) to make decisions and interpret outcomes in terms of gains and losses relative to a reference point, and they give more weight to losses than to equivalent potential gains. Further, according to this theory, since prospects are coded as gains or losses from a zero-point, people tend to be more risk averse when faced with potential gains and more risk seeking when faced with potential losses.

Prospect theory identifies two distinct phases of a person's choice process: framing and valuation [11]. In the framing phase, the individual creates scenarios which include various possibilities, actions, and outcomes that are relevant to the decision. Then, in the

valuation phase, the person assesses the value of each of these scenarios and chooses accordingly. However, these choices are limited by the bounded nature of rationality and heuristics ingrained in people.

We use prospect theory to inform our review set assessment model which predicts the helpfulness of a review set in making purchase decisions. We contend that since consumers have limited time and possess limited information about the product that they want to purchase, they often prefer to apply heuristics to determine the viability of their purchase option. We argue that the review sets that reduce cognitive load and allow the use of heuristics will be perceived to be *most* helpful. In contrast, review sets that confuse potential buyers with conflicting information (and preclude the use of heuristics) will be perceived to be *least* helpful.

When faced with consistent information across a set of reviews, consumers are less likely to experience cognitive dissonance while trying to reach a satisficing outcome, reducing the potential for confusion. We define consumers' confusion as their failure in developing a correct interpretation of various aspects of a product during information processing [12]. Consistent opinions across reviews leave very little room for consumers to get confused. Additionally, higher information consistency in a review set allows potential consumers to utilize cognitive shortcuts, such as inferring overall quality of the product based simply on the consensus. In contrast, reviews with inconsistent opinions do not allow for the use of heuristics to make an easy assessment of the product. This increases the likelihood of confusion in the mind of the decision maker regarding the value of the review set. Prior studies also found evidence for this inverse relationship between consistency and confusion [13, 14]. Therefore, we hypothesize that *the inconsistency of information provided in a review set is positively related to perceived confusion (H1)*.

Based on prospect theory, we further argue that inconsistency in the review set also influences its credibility. When the information provided is consistent across reviews, it is easier for consumers to use heuristics to evaluate these reviews [6]. However, using heuristics makes it less likely that consumers would employ critical processing to evaluate the merits of the arguments presented, and thus are more likely to accept the information as credible without questioning it. In contrast, a greater level of inconsistency among reviews may prompt consumers to doubt the credibility of the review set. Consequently, we hypothesize that *the inconsistency of information provided in a review set is negatively related to perceived information credibility (H2)*.

A deeper exploration of review inconsistency and its influence on consumers, however, suggests a more layered relationship. The confusion elicited by review inconsistencies is not an end state by itself; rather, it triggers the consumer to assess the usefulness of the review set in making the right purchase decision. Based on prospect theory, when consumers are confused regarding the value of the product, they become more risk averse because of the greater uncertainty and risk closely associated with confusion [15]. This reduces their confidence in purchasing the product, thus lowering their perceived helpfulness of the review set. In contrast, when the potential consumers are faced with consistent reviews, it elicits clarity in their mind about the value of the product. Any information that reduces cognitive load and facilitates the use of decision heuristics is deemed to be helpful to the decision maker. Therefore, we hypothesize that *perceived confusion is negatively related to perceived review set helpfulness (H3)*.

Similarly, we argue that the evaluation of the review set's credibility is not the final cognitive outcome. Instead, it prompts the consumer to evaluate the utility of the review set in making the right purchase decision. Credible information reduces cognitive load and enables the consumer to apply heuristics to make the purchasing decision. As stated earlier, any information that permits the use of heuristics is deemed to be helpful to the decision maker. Therefore, we hypothesize that *perceived information credibility is positively related to review set helpfulness (H4)*.

In addition to having a direct influence, we argue that confusion also exerts an indirect influence on perceived review set helpfulness. According to prospect theory, buyers have asymmetric responses towards better- and worse-than-expected purchase decisions and this response is influenced by the valence of the message: people prefer avoiding a regrettable decision to missing a comparable opportunity [16]. When consumers are confused, they experience a cognitive dissonance as they attempt to identify a reference point to guide their decisions (e.g. product aspects that matter the most). Without enough evidence regarding which option (buying or not buying) would be a regrettable decision, consumers are more likely to find the available information incomplete. When reviews are perceived more complete, they are more likely to lead to a satisficing decision, thereby increasing the perceived helpfulness of the review set. Accordingly, we hypothesize that *perceived confusion is negatively related to perceived information completeness (H5)* and *perceived information completeness is positively related to review set helpfulness (H6)*.

As discussed earlier, heuristics enable faster decision making with a reasonable certainty of a desired outcome. However, people can only utilize heuristics if the information presented is considered credible and complete by the receiver. That is, if the buyers perceive the information in a review set to be sufficiently credible, they will likely utilize heuristics and thus find the set more helpful. However, if consumers question the credibility of the information, they tend to seek additional evidence in order to replace less credible information with more credible information. Accordingly, we hypothesize that *perceived information credibility is positively related to perceived information completeness (H7)*.

The complete model and hypotheses pathways are depicted in Figure 1.

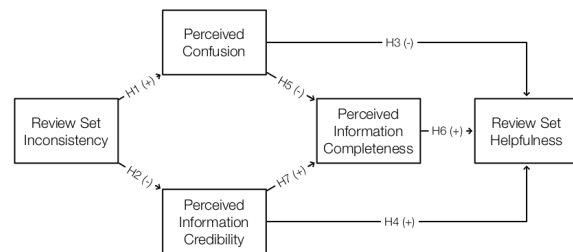


Figure 1. Review Set Assessment Model

3. Method

3.1. Sample

Participants ($n = 167$, 82 female) were recruited from an upper-level management class at a large, public university in the southern United States. Participants' age ranged from 19 to 54 ($M = 20.8$, $SD = 2.73$).

3.2. Stimulus

All participants were asked to evaluate a digital camera available for purchase on Amazon.com, which we referred to as "Model X." As a basis for evaluating the camera, participants were then shown a set of four product reviews that were ostensibly the most recently posted. Each review set consisted of four individual reviews that are adapted from prior research [e.g., 17] and based on real reviews posted on Amazon.com (see Table 1).

Table 1. Content of Individual Reviews

	Positive Reviews (+)	Negative Reviews (-)
1A	The autofocus function is very good. It is very fast and accurate, even in low light situations. I have been able to shoot decent pictures of my children even in low light using the autofocus function. This kind of pictures usually do not come out well with other cameras – but with this one it does.	The autofocus function is very bad. It is very slow and inaccurate, especially in low light situations. I have been unable to shoot decent pictures of my children in low light using the autofocus function. This kind of picture usually comes out well with other cameras – but with this one it does not.
2A	The auto focus of this camera is very useful, and I am impressed. It never struggles to find the focus in darker places. Because it's so quick, the picture is rarely out of focus. Just yesterday, I had to take several pictures of my friends at night and those pictures came out very sharp and clear.	The auto focus of this camera is almost useless, and I am unimpressed. It often struggles to find the focus in darker places. Because, it's so slow, the picture is often out of focus. Just yesterday, I had to take several pictures of my friends at night, and those pictures came out completely blurry and unclear.
1B	This camera is very user-friendly. After just a few days of use, I find it really straightforward and intuitive. For example, it only took me a few seconds to figure out how to turn off the flash. It's very easy to figure out the various settings/functions even without the instruction manual.	This camera is not user-friendly at all. I have been using it for a few days and I still find it really complicated and unintuitive. For example, it took me over an hour to figure out how to turn off the flash. It's impossible to figure out any of the settings/functions without the instruction manual.
2B	I find the camera quite easy to operate. Finding the right buttons to perform various functions is no problem, even without referring to	I find the camera quite difficult to operate. Finding the right buttons to perform various functions is a big problem, even after referring to the

the manual. All buttons and controls are well designed. Switching between different settings (e.g., portrait mode, landscape mode) is also super quick and easy.	manual. All buttons and controls are poorly designed. Switching between different settings (e.g., portrait mode, landscape mode) is also super slow and hard.
--	---

Note. "1" and "2" = different reviews about the same feature; "A" = reviews about autofocus feature; "B" reviews about user-friendliness features.

Each of the individual reviews focused on one of two features of digital cameras — that is, the autofocus function (Feature A) or its user-friendliness (Feature B). Two positive reviews were first written for each feature (1A+, 2A+, 1B+, 2B+). Reviews 1A+ and 2A+ were worded very differently to give the impression they were written by different reviewers. The same applies to Reviews 1B+ and 2B+. Then, negative versions of each review were written (1A-, 2A-, 1B-, 2B-). Substantive content was made to be as similar as possible between the positive and negative versions of individual reviews in order to remove possible confounds.

Using these eight reviews, we created four experimental conditions that vary in the level of inconsistency. In the first two conditions, there was *no inconsistency* in the review sets. The reviews were either all positive (1A+, 2A+, 1B+, 2B+) or all negative (1A-, 2A-, 1B-, 2B-). In the review set with *low inconsistency*, two reviews were negative and two were positive, and the camera feature on which the negative reviews focused differed from the focal feature of the positive reviews (i.e., 1A-, 2A-, 1B+, 2B+, or 1A+, 2A+, 1B-, 2B-). In the *high inconsistency* condition, two reviews were negative and two were positive, and the camera feature on which all four reviews focused was the same — that is, the reviews directly contradicted each other (i.e., 1A+, 2A-, 1B+, 2B-, or 1A-, 2A+, 1B-, 2B+).

3.3. Procedure

Participants were randomly assigned to read and evaluate one of the four aforementioned review sets. In the low inconsistency condition, a half of the participants were randomly assigned to view negative reviews on the autofocus feature and positive reviews on the user-friendliness (1A-, 2A-, 1B+, 2B+), while the other half viewed positive reviews on the autofocus feature and negative reviews on the user-friendliness (1A+, 2A+, 1B-, 2B-). This counterbalancing allowed us to account for preferences our population may have had for one

feature over another. A similar counterbalancing procedure was used for participants in the high inconsistency condition. After reading the review set, participants evaluated the set in terms of helpfulness [3], completeness of information [18], credibility of information [6], and breadth of information [19]. They also reported how confused they felt by the reviews [20]. Finally, each individual review in the review set was evaluated in terms of its helpfulness [3]. All measures exhibited high internal consistency (Cronbach's alphas > .84).

4. Results

Before examining the internal validity of our experiment and testing our hypotheses, we collapsed the 'no inconsistency' conditions into a single group. This was done because the valence of the reviews was not of our interest in the present study. To ensure this aggregation was appropriate, we used a series of *t*-tests to assess whether the two groups differed significantly across any of our dependent variables. No significant differences were observed. Thus, from this point forward, we will refer to *three* experimental conditions, rather than four.

The internal validity of the experiment was evaluated with two analyses. First, we conducted a manipulation check using perceived consistency as the dependent variable [6]. Participants in the low inconsistency condition reported the highest consistency ($M = 5.80, SD = 1.02$), followed by the moderate inconsistency condition ($M = 4.07, SD = 1.25$), and high inconsistency condition ($M = 2.93, SD = 1.80$). A one-way ANOVA showed these means were significantly different, $F(2,164) = 72.22, p < .001$. These results support that our manipulation worked as intended.

Second, we examined the possibility of an alternative explanation of the observed effects—namely, information breadth: Because the high inconsistency condition had two reviews that directly contradicted the two other reviews, readers may perceive this review set to have limited information breadth. Since reviews with more information breadth are seen as more helpful, perceptions of limited breadth could explain changes in helpfulness [21], rather than the inconsistency itself. A test of this alternative explanation using participants' ratings of information breadth showed non-significant differences across conditions, $F(2, 164) = .41, p = .66$.

Our hypotheses concerning the effects of review set inconsistency were first tested using one-way ANOVAs, all of which were statistically significant (see Table 2). First, as predicted, perceived helpfulness of the review set was lowest when

inconsistency was high ($M = 4.62, SD = 1.79$) and highest when inconsistency was absent ($M = 5.96, SD = 1.02$). Second, the pattern of mediating effects also trended in the predicted directions. Perceived confusion, for example, was highest ($M = 4.37, SD = 1.79$) when inconsistency was high, and lowest ($M = 2.69, SD = 1.39$) when inconsistency was absent. In contrast, when inconsistency was absent, perceived credibility ($M = 5.48, SD = .98$) and perceived information completeness ($M = 5.26, SD = 1.16$) were at their highest, and when inconsistency was high, perceived credibility ($M = 4.65, SD = 1.33$) and perceived information completeness ($M = 4.21, SD = 1.50$) were at their lowest.

Table 2. ANOVAs and Descriptive Statistics of Dependent Variables across Experimental Conditions

Inconsistency	Credibility	Confusion	Completeness	Helpfulness
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
None (<i>n</i> = 84)	5.48 (0.98) ^a	2.69 (1.39) ^a	5.26 (1.16) ^a	5.96 (1.02) ^a
Low (<i>n</i> = 42)	5.09 (0.89) ^a	3.37 (1.31) ^b	4.80 (1.01) ^b	5.55 (0.87) ^a
High (<i>n</i> = 41)	4.65 (1.33) ^b	4.37 (1.79) ^c	4.21 (1.50) ^c	4.62 (1.79) ^b
<i>F</i> -ratio	8.73**	17.92**	10.28**	16.65**
<i>R</i> ²	.10	.18	.11	.17

Note. Means with different superscripts are significantly ($p < .05$) different from each other.

* = $p < .05$. ** = $p < .01$.

Our full mediated model was tested using structural equation modeling in Mplus version 6.12 [22]. Because all the data were collected from a single source, we conducted the Harman single-factor test, where all items were set to load on one latent variable [23]. This model fit poorly, $\chi^2(104) = 877.26, p < .05$, RMSEA = .21, CFI = .66, TLI = .61, SRMR = .11, indicating that common source bias is not likely to be problematic.

The hypothesized model fit the observed data fairly well: $\chi^2(97) = 219.77, p < .05$, RMSEA = .09, CFI = .95, TLI = .93, SRMR = .05. Model parameter estimates are shown in Figure 2. All path estimates were statistically significant, providing support for our hypotheses. The statistical significance of the indirect effects was estimated using bootstrapped standard errors [24]. As shown in Table 3, review set inconsistency had a significant negative effect on review helpfulness through reducing perceived information credibility ($\beta = -.18, p < .01$) and increasing confusion ($\beta = -.14, p < .05$). The effect of review set inconsistency was fully mediated by perceived information credibility, confusion, and information completeness. That is, after accounting

for these mediating variables, the effect of review set inconsistency on review set helpfulness was not statistically significant ($p = .13$).

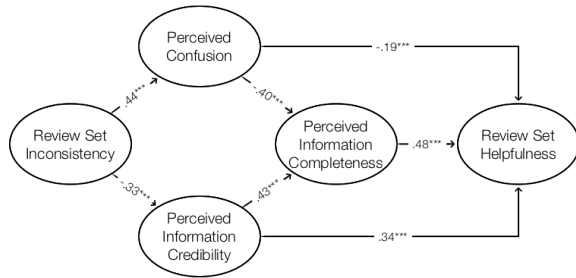


Figure 2. SEM Analysis Results

Note. All parameter estimates shown are standardized and statistically significant at the $p < .01$ level. Perceived Confusion and Perceived Information Credibility were allowed to covary, but this is not shown here. For clarity, individual item indicators and their factor loadings are also omitted.

Table 3. Tests of Indirect Effects

Path	Indirect Effect	95% CI
Inconsistency → Credibility → Completeness → Helpfulness	-.11*	[-.20, -.02]
Inconsistency → Credibility → Helpfulness	-.18**	[-.31, -.05]
Inconsistency → Confusion → Completeness → Helpfulness	-.14*	[-.24, -.03]
Inconsistency → Confusion → Helpfulness	-.14*	[-.25, -.02]

Note. * = $p < .05$. ** = $p < .01$.

We also tested the assumption that the helpfulness of a review set is equal to the average helpfulness of the individual reviews that make up the set. A 2×3 linear mixed effects model was conducted where the within-subjects factor was perceptions of review helpfulness (review set helpfulness and the average of individual review helpfulness), and the between-subjects factor was the three conditions described previously. As shown in Figure 3, review set helpfulness was significantly higher than the average of individual review helpfulness in the no inconsistency condition ($M_{\text{Difference}} = .35, SE = .12, p = .005$) but significantly lower in the high inconsistency condition ($M_{\text{Difference}} = -.58, SE = .17, p = .001$). However, in the low inconsistency condition, no significant difference was observed ($M_{\text{Difference}} = .14, SE = .17, p = .41$).

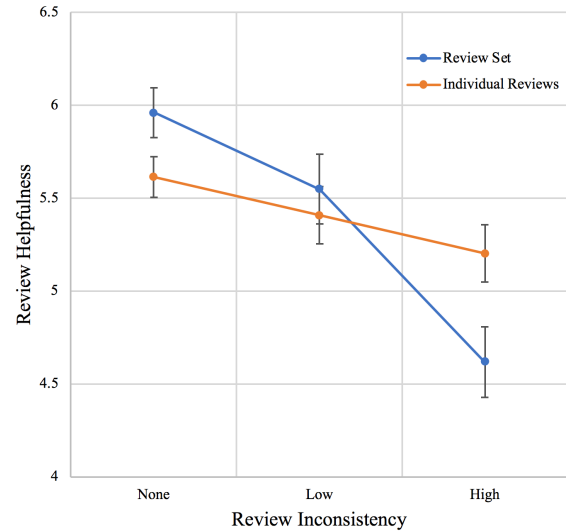


Figure 3. Review Set Helpfulness Compared to Average of Individual Review Helpfulness

Note. Standard error bars are shown.

5. Discussion

In this study, we explored how and why different levels of inconsistency among a set of reviews influence consumers' helpfulness perception of the review set. Our carefully designed experiment provides compelling evidence that greater inconsistency reduces perceived helpfulness of the review set, and that this overall effect is fully mediated by decreased perceived information credibility and perceived information completeness, as well as increased perceived confusion. Moreover, we find evidence that “the whole is not equal to the sum of its parts” when the information provided in the individual reviews is consistent or contradicting.

5.1. Theoretical Implications

This study has a number of theoretical contributions. First, our research challenges the implicit assumption of the whole being equal to the sum of its parts in the online reviews context. Both researchers and practitioners have placed disproportionate attention to studying and capturing the helpfulness of an individual review. In reality, however, consumers rarely read a single review. They typically consult multiple reviews before making their purchase decisions. If the helpfulness of a set of reviews cannot be derived from averaging the helpfulness of individual reviews, then a huge gap exists regarding our understanding of how consumers form helpfulness judgment of a review set. Our study

fills this gap by revealing conditions under which this implicit assumption holds and conditions under which it does not hold. Our findings further illustrate the importance of exploring the whole beyond effects of its individual parts in online reviews, and they also open up exciting opportunities for future research to explore how consumers integrate multiple and oftentimes conflicting pieces of information in decision-making.

Second, this study deepens our understanding of inconsistency and emphasizes the importance of differentiating diverse types of inconsistency in user-generated content. Inconsistency within a single review is typically manifested as mixed opinions or two-sided arguments, and prior studies found a boosting effect of two-sidedness on review helpfulness [6, 7]. In contrast, we find a dampening effect of inconsistency among reviews at the level of a review set, suggesting the different roles of inconsistency within vs. across reviews. Moreover, inconsistency in a review rarely involves direct contradiction, but inconsistency among different reviews can manifest as mixed opinions as well as direct contradiction regarding the same attribute of a product. The finding of a contradicting review set being the least helpful shows that different types of inconsistency should not be treated equally.

Finally, our study also demonstrates the critical role of confusion as another important determinant of consumers' helpfulness perception at the review set level. Prior research studying antecedents of the helpfulness evaluation of individual reviews has revealed the impact of information amount and credibility [e.g., 1, 25], and our findings suggest that they continue to influence perceived helpfulness of a review set. At the same time, consumers may also get confused when they are faced with a set of reviews that are inconsistent or contradictory, and this confusion can directly and indirectly reduce their helpfulness evaluation of the review set. As a widely experienced but under-studied emotion in online reviews, confusion of consumers may play an indispensable role when they integrate and make sense of different pieces of information from diverse sources. Our study represents one of the initial attempts to understand this critical emotion in consumer decision making.

5.2. Practical Implications

Our findings also offer useful implications for review platforms and product manufacturers. For instance, product manufacturers can use our framework to anticipate the likelihood of prospective consumers to perceive the set of prominently displayed reviews as complete and helpful. If they

expect the review set to lead to higher confusion and lower information credibility, they can prioritize their efforts to address the concerns raised by the reviewers, for example through their responses to reviews or to questions from prospective consumers. Furthermore, manufacturers can carefully select which consumer reviews, or parts of consumer reviews, they will display on their product page so that they can avoid direct contradictions while retaining the richness of the review assessments.

Review platforms could increase the helpfulness of the overall review set for consumers by trying to contain the perceived confusion that consumers may experience. This could, for example, be done by automatically extracting parts of individual reviews that address the same product aspect and creating a summary assessment. This may enable consumers to easily see the review crowd's assessment (e.g., of a camera's autofocus functionality) without having to reconcile contradictory reviews.

5.3. Limitations and Future Research

There are several limitations that have to be considered when interpreting the findings of this study. Each of these also offers exciting opportunities for future research.

First, our study focused on a specific product, a digital camera. Subjects may have different levels of experience with different types of products or services. Most subjects will be familiar with the essential functionalities of a digital camera due to the proliferation of smart phone cameras. Thus, further studies are needed to demonstrate that our findings hold across different products and services that subjects have different levels of familiarity with.

Second, we designed a collection of stylized reviews for our experiment. The purpose of our review design was to be able to control for confounds due to the product aspects being discussed and the linguistic manner in which the reviews were written. The resulting reviews only differed in terms of valence, but not in terms of absolute content or tone. Real reviews on commercial platforms typically are not similarly comparable. They demonstrate a wide variety in terms of length, breadth, content richness, tone, and language use. Future research should replicate our experiment using a collection of carefully selected realistic reviews that are modified to change their valence, but not their breadth, depth, and focus.

Third, our subjects were only presented with a collection of four reviews. On typical consumer platforms, users can see more than four reviews and have control over which reviews they want to read in more detail. Future efforts should increase the size of

the review set. This would also allow researchers to study more nuanced combinations of consistent and contradictory reviews. For example, it would be useful to determine the effect of a single contradictory review that results in a high contradiction situation, i.e. the review gives an opposing opinion about a specific product attribute. At what point does a single high contradictory review no longer raise a consumer's perception of confusion and lower their perception of information credibility? Future efforts should also allow users to have control over the number of reviews that they want to open or expand and read. Tracking their behavior will provide insight into their actual sense of information completeness before coming to a decision.

Fourth, our subjects were not necessarily interested in buying a digital camera at the time of the experiment. Thus, they may not have been as motivated to cognitively process the various reviews as real consumers that are actively searching for a camera to buy. Therefore, future research should solicit samples of users that are genuinely interested in the product for which reviews are presented. This could, for example, be done through crowdsourcing services such as Amazon's Mechanical Turk or Qualtrics Panels.

Fifth, our study primarily focused on the effects of different compositions of a review set on review set helpfulness. Future efforts should study other critical downstream consequences such as consumers' intention to seek more reviews and decisions to purchase a product or service, and further explore how they could be shaped by different levels of inconsistency among a set of reviews.

Sixth, another interesting avenue for future studies is to explore whether our findings hold across cultures. This will be especially interesting from the perspective of the comfort level in different cultures to deal with inconsistent information.

Finally, it will be worthwhile to explore the extent to which the phenomena we observed can also be found in other contexts where decision makers rely on a collection of individual (and oftentimes conflicting) opinions and recommendations, such as group decision-making, editors and program officers working with review panels for academic publications and funding proposals, and media panels that offer commentary to inform the public.

References

[1] S. M. Mudambi and D. Schuff, "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com," *MIS Quarterly*, vol. 34, no. 1, pp. 185-200, 2010.

[2] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, vol. 19, no. 3, pp. 291-313, 2008.

[3] D. Yin, S. D. Bond, and H. Zhang, "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews," *MIS Quarterly*, vol. 38, no. 2, pp. 539-560, 2014.

[4] D. Yin, S. Mitra, and H. Zhang, "When Do Consumers Value Positive Vs. Negative Reviews? An Empirical Investigation of Confirmation Bias in Online Word of Mouth," *Information Systems Research*, vol. 27, no. 1, pp. 131-144, 2016.

[5] K. Koffka, *Principles of Gestalt Psychology*. Oxford, England: Routledge, 2013.

[6] C. M.-Y. Cheung, C.-L. Sia, and K. K. Y. Kuan, "Is This Review Believable? A Study of Factors Affecting the Credibility of Online Consumer Reviews from an Elm Perspective," *Journal of the Association for Information Systems*, vol. 13, no. 8, pp. 618-635, 2012.

[7] A. E. Schlosser, "Can Including Pros and Cons Increase the Helpfulness and Persuasiveness of Online Reviews? The Interactive Effects of Ratings and Arguments," *Journal of Consumer Psychology*, vol. 21, no. 3, pp. 226-239, 2011.

[8] L. W. Jones, R. C. Sinclair, and K. S. Courneya, "The Effects of Source Credibility and Message Framing on Exercise Intentions, Behaviors, and Attitudes: An Integration of the Elaboration Likelihood Model and Prospect Theory," *Journal of Applied Social Psychology*, vol. 33, no. 1, pp. 179-196, 2003.

[9] A. J. Rothman and P. Salovey, "Shaping Perceptions to Motivate Healthy Behavior: The Role of Message Framing," *Psychological Bulletin*, vol. 121, no. 1, pp. 3-19, 1997.

[10] D. Kahneman and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, vol. 47, no. 2, pp. 263-291, 1979.

[11] A. Tversky and D. Kahneman, "Rational Choice and the Framing of Decisions," *The Journal of Business*, vol. 59, no. 4, pp. 251-278, 1986.

[12] P. W. Turnbull, S. Leek, and G. Ying, "Customer Confusion: The Mobile Phone Market," *Journal of Marketing Management*, vol. 16, no. 1-3, pp. 143-163, 2000.

[13] L. Baker and R. I. Anderson, "Effects of Inconsistent Information on Text Processing: Evidence for Comprehension Monitoring," *Reading Research Quarterly*, vol. 17, no. 2, pp. 281-294, 1982.

[14] M. J. Rosenberg, "Hedonism, Inauthenticity, and Other Goals toward Expansion of a Consistency Theory," in *Theories of Cognitive Consistency: A Sourcebook*, 1968, pp. 73-111.

[15] M. R. Solomon, D. W. Dahl, K. White, J. L. Zaichkowsky, and R. Polegato, *Consumer Behavior: Buying, Having, and Being*. Pearson, 2014.

[16] D. Kahneman and A. Tversky, "Choices, Values, and Frames," *American Psychologist*, vol. 39, no. 4, pp. 341-350, 1984.

[17] Q. B. Liu and E. Karahanna, "The Dark Side of Reviews: The Swaying Effects of Online Product Reviews on Attribute Preference Construction," *MIS Quarterly*, vol. 41, no. 2, pp. 427-448, 2017.

- [18] M. Shoham, S. Moldovan, and Y. Steinhart, "Positively Useless: Irrelevant Negative Information Enhances Positive Impressions," *Journal of Consumer Psychology*, vol. 27, no. 2, pp. 147-159, 2017.
- [19] K. Laursen and A. Salter, "Open for Innovation: The Role of Openness in Explaining Innovation Performance among U.K. Manufacturing Firms," *Strategic Management Journal*, vol. 27, no. 2, pp. 131-150, 2006.
- [20] P. J. Silvia, "Confusion and Interest: The Role of Knowledge Emotions in Aesthetic Experience," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 4, no. 2, pp. 75-80, 2010.
- [21] H. Hong, D. Xu, G. A. Wang, and W. Fan, "Understanding the Determinants of Online Review Helpfulness: A Meta-Analytic Investigation," *Decision Support Systems*, vol. 102, pp. 1-11, 2017.
- [22] L. K. Muthén and B. O. Muthén, *Mplus: Statistical Analysis with Latent Variables: User's Guide*. Muthén & Muthén Los Angeles, 2010.
- [23] P. M. Podsakoff, S. B. MacKenzie, J.-Y. Lee, and N. P. Podsakoff, "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, vol. 88, no. 5, pp. 879-903, 2003.
- [24] D. P. MacKinnon, A. J. Fairchild, and M. S. Fritz, "Mediation Analysis," *Annual review of psychology*, vol. 58, pp. 593-614, 2007.
- [25] M. Li, L. Huang, C.-H. Tan, and K.-K. Wei, "Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features," *International Journal of Electronic Commerce*, vol. 17, no. 4, pp. 101-136, 2013.