Available online at www.sciencedirect.com
ScienceDirect
Journal of Interactive Marketing 39 (2017) 1 – 14
JOURNAL OF INTERACTIVE MARKETING
www.elsevier.com/locate/intmar

The Presentation Format of Review Score Information Influences Consumer Preferences Through the Attribution of Outlier Reviews

Adrian R. Camilleri

RMIT University, 445 Swanston Street, Melbourne, Victoria 3000, Australia

Abstract

Review score information can be presented in different formats. In three online experiments, we examined consumers' behavior in the context of review scores presented in a *disaggregated* format (individual review scores observed sequentially and individually), an *aggregated* format (review scores summarized into a frequency distribution chart), or both together. Participants tended to attribute outlier review scores to reviewer rather than product reasons. This tendency was more prevalent when reviews were presented in disaggregated format. Moreover, reviews attributed to reviewer reasons tended to be perceived with low credibility. When presented with a choice between two products with equal average review scores but different variances, participants chose as if outlier review scores were discounted when scores were presented in the disaggregated format. This tendency emerged even when disaggregated and aggregated formats were presented together. The number of review scores moderated the effect of format on choice. We argue that disaggregated information allows consumers to better track the number of outliers and, when the number of outliers is small, prompts them to attribute these outliers to reviewer reasons, and subsequently discount them.

Keywords: Word of mouth; Online consumer reviews; Attribution theory; Decisions under uncertainty

Introduction

"Word of mouth" (WoM) refers to informal person-to-person communications regarding a brand, product, or service (Arndt 1967; Westbrook 1987). Many have argued that WoM is one of the most powerful forces shaping consumer behavior (Trusov, Bucklin, and Pauwels 2009; Whyte 1954). Correspondingly, a large body of work has focused on WoM antecedents, consequences, and management (Berger 2014; De Matos and Rossi 2008; Lang and Hyde 2013). We focus here on product review scores because prior research has established them as an important source of consumer information that influences purchasing decisions (Chatterjee 2001; Chevalier and Mayzlin 2006; Senecal and Nantel 2004). Much work has examined how consumer responses to reviews are influenced by stimuli (such as volume), communicators (such as expertise), and contextual factors (such as platform) (see Cheung and Thadani 2012, for a review). The focus of the current paper is the influence of review score format on consumer's product preference.

Preliminary research suggests that consumers may form different product preferences depending on whether associated product review scores are presented as an aggregated review score distribution or as disaggregated individual review scores (Wulff, Hills, and Hertwig 2014). This finding is important because managers have the ability to design online platforms that direct consumer attention to different formats of review score information, which may subsequently influence consumer behavior. We contribute by extending this finding to contexts in which both aggregated review score information and disaggregated review score information are presented. We also contribute by explaining this behavior in the light of attribution theory: outlier review scores are attributed to reviewer (vs. product) reasons, discredited, then discounted, and this attribution is more likely when review scores are presented in a disaggregated format.

Different Information Presentation Formats

Review score information is commonly presented in two basic formats. "Aggregated" information includes formats in which

E-mail address: adrian.camilleri@rmit.edu.au.

http://dx.doi.org/10.1016/j.intmar.2017.02.002
1094-9968/

multiple pieces of information are obtained and processed collectively and simultaneously (for example, review scores that have been summarized into a frequency histogram at Amazon.com). "Disaggregated" information includes formats in which multiple pieces of information are obtained and processed individually, sequentially, and over time (such as a list of individual reviews left at Amazon.com).

Most online websites give consumers access to *both* disaggregated and aggregated review score information. However, consumers naturally have an option of which format of information to focus on, and this focus is likely to vary (Purnawirawan, De Pelsmacker, and Dens 2012). To test this assumption, we conducted a survey of 104 American participants recruited from Amazon's Mechanical Turk regarding their behavior when looking at online review aggregator websites. The participant's behavior fell into four categories such that on 38% of occasions the average review score, the distribution of review scores, and the individual review scores were all considered, on 36% of occasions the individual review scores were ignored, and on 11% of occasions both the average review score and the distribution of review scores were ignored. Furthermore, some websites directly limit the format in which reviews are presented. For example, Zomato.com, one of the fastest-growing restaurant search websites, does not provide users with a frequency histogram summarizing review scores. Other websites, such as HealthGrades.com, focus on presenting aggregated data to specifically reviewed service characteristics (e.g., trustworthiness, helpfulness). In sum, consumers will often be exposed to review score information that is in aggregated format, disaggregated format, or both, prior to making a purchase decision.

Wulff, Hills, and Hertwig (2014) compared choices between two alternative products for which review score information was presented in either aggregated or disaggregated format. Each choice pair comprised of a low-variance product for which the review scores clustered tightly around the average score, and a high-variance product for which the review scores clustered widely around the average score. In this study, and in those that follow, the high variance product was associated with an outlier review; that is, a review score left by only a minority of reviewers. Half the participants saw the review scores summarized in a single frequency histogram (an aggregated information format) whereas the others observed the review scores individually and sequentially (a disaggregated information format). Two procedural features are worth noting. First, the disaggregated review scores were not accompanied by a textual elaboration of the score. Second, participants in the latter group were unconstrained in terms of the number of reviews they could sample.

The researchers found significant differences in choices depending on presentation format; namely, those presented with disaggregated reviews tended to make choices as if discounting outlier reviews. The authors attributed their observations primarily to sampling error: those who were sampling individual review scores often failed to sample a sufficient number to ever observe the outlier scores. Indeed, in the aggregated format, each product was associated with 100 review scores (a total of 200 scores per choice). By contrast, in the disaggregated format, on average, participants considered just 21 reviews per choice, meaning that the two conditions differed considerably in terms of the number of product reviews participants observed. Therefore, it appears that many participants in this study chose *as if* discounting outlier reviews when presented in a disaggregated format because outliers were never observed in the first place.

We believe, despite the confound in information provided, that the general conclusion from Wulff, Hills, and Hertwig (2014) – that there are format-dependent differences in choice – may be accurate. To test this hypothesis directly, in our studies we ensured that participants in different groups were always presented with equivalent information. By ruling out information differences between groups, we were able to directly test the impact of different review score information formats, and eliminate sampling error as an explanation for any observed format dependent differences. We also hypothesized a different underlying mechanism: outlier review attribution.

Outlier Review Attributions

We argue that consumers form attributions about reviews – particularly outlier review scores – that vary depending on the review score presentation format. Kelley's (1967) covariation model of attribution theory explains how people make causal inferences to understand why communicators advocate certain positions. According to this theory, two possible attributions relate to whether conveyed opinions are based on external (product) reasons or internal (reviewer) reasons, and one factor that helps determine an attribution is consensus: the extent to which other people behave in the same way in a similar situation (Kelley 1973). A low level of consensus tends to be associated with internal attributions. For instance, if ten reviewers have ten different opinions about a product, then one might conclude that the ten opinions stem more from internal reviewer reasons (e.g., personality) than from external product reasons (e.g., quality). Conversely, if ten reviewers have ten similar opinions about a product, then one might conclude that the ten opinions stem more from product reasons than from reviewer reasons.

Research suggests that consumers are more likely to discount an individual review after reading the reviewer's comments if they attribute that review to reviewer reasons, which can subsequently affect brand evaluation (Laczniak, DeCarlo, and Ramaswami 2001). Research has also shown that consumers are more likely to attribute a review to reviewer (vs. product) reasons for experience (e.g., cosmetics) versus search (e.g., MP3 player) goods, which can subsequently affect product attitude (Park and Han 2008). Consumers are also more likely to attribute product reasons to negative reviews about utilitarian products but more likely to attribute reviewer reasons to negative reviews about hedonic products, which can subsequently affect attitude towards the review (Sen and Lerman 2007).

To our knowledge, only one previous study has examined the combined effect of an aggregated rating and individual reviews from an attribution perspective. Qiu, Pang, and Lim (2012) presented participants with a single review that was of either negative or positive valence. Half the participants were also shown the average product rating based on a total of 96 reviews,

which was arranged to be incongruent with the individual review. The researchers found that attribution of the individual review to product reasons was reduced when presented with the incongruent aggregated rating, but only when the review was positive (and hence, the average rating was low). In contrast, when the review was negative (and hence, the average rating was high), attribution of the review to product reasons was no different when presented together with the incongruent aggregated rating. The presence of incongruent aggregated rating information reduced perceived review credibility and diagnosticity of the review via the review attribution. These results suggest that consumers can attribute a review to reviewer reasons when they learn that it is an outlier – that is, when the review conflicts with the consensus opinion – and subsequently discount it. An important but unanswered question emerging from this study is whether such attribution varies depending on whether the degree of reviewer consensus is learned about from aggregated or disaggregated review scores, and when based purely on the distribution of the review scores (i.e., without reading accompanying reviewer comments).

Format-dependent Differences in Attribution and Preference

Our literature review suggests that consumers can discount individual reviews because of reviewer attributions based on factors such as type of product (search vs. experience), the valence of the review (negative vs. positive), and whether the review conflicts with the average rating. Here, we propose a new factor that may lead consumers to discount reviews through review attribution: the presentation format of the review score. Our conceptualization is summarized in Fig. 1 and expanded below.

Information Format and Perceptions of Number of Reviews (and Outliers)

According to our conceptualization, the two different presentation formats influence how information is encoded, which can be identified by people's perceptions of the number of reviews.

When people are presented with disaggregated information, they tend to automatically encode into memory the frequency of events (Hasher and Zacks 1984). This tendency, often called the "frequency heuristic" (Alba and Marmorstein 1987), is the basis of many exemplar-, or instance-, based decision models such as the ACT-R model (Anderson and Lebiere 2014). In general, people are good at identifying the frequency of disaggregated numbers when later probed (Hau et al. 2008). Thus, we argue that people will have a relatively accurate perception of how many reviews there were after being presented with review scores in disaggregated format. Given that a pilot study ($N = 96$) showed that there was a very strong positive correlation between accurate perception of total number of reviews and accurate perception of outlier reviews, for our studies we assumed accurate perception of the former implied accurate perception of the latter.

In contrast, when people are presented with aggregated information, they tend to overlook the frequency of events (Griffin and Tversky 1992). For example, Obrecht, Chapman, and Gelman (2007) asked participants to evaluate products associated with review scores that varied in terms of the average product rating, the number of reviews, and the variance of the ratings. They found that only about half of the participants used the number of reviews information — the others simply ignored it. The likely reason for this oversight is that consumers are cognitive misers who often rely on mental short cuts to make assessments and decisions (Fiske and Taylor 2013; Gigerenzer and Selten 2002; Payne, Bettman, and Johnson 1993). In the context of review scores, it is cognitively less demanding to make an evaluation based on the average review score and distribution of reviews than it is to additionally evaluate this information in conjunction with the number of reviews. Therefore, we expect consumers presented with aggregated review score information to have a relatively inaccurate perception of the number of reviews and, correspondingly, a relatively inaccurate perception of the number of outlier reviews.

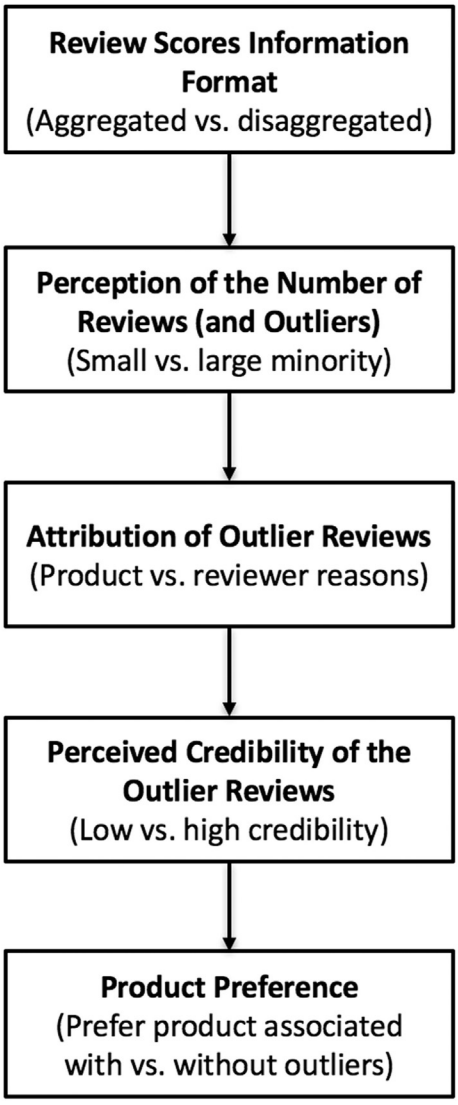


Fig. 1. Conceptual depiction of how review score information presentation format influences preference.

H1. People will be better able to correctly identify the number of reviews when those scores are presented in a disaggregated (vs. aggregated) format.

Although, people presented with disaggregated information tend to have a good sense of the relative frequency of events – that is, judged frequency tends to increase with increases in actual frequency – absolute accuracy of frequency judgments can suffer from distortions (Zacks and Hasher 2002). In particular, frequency tracking becomes less accurate as the number of events increases with a tendency to overestimate the frequency of rare events and underestimate the frequency of common events (Lejarraga 2010; Ungemach, Chater, and Stewart 2009). This phenomenon is not surprising given the limited capacity of working memory (Baddeley 1992) and tendency to focus on small and more recent samples of experience (Hertwig and Pleskac 2010; Plonsky, Teodorescu, and Erev 2015) when manipulating information necessary for complex tasks, such as the evaluation of a product based on reviews.

H2. The effect of format on ability to correctly identify the number of reviews (i.e., H1) will be moderated by the number of reviews: when there are few reviews, identification will be better when information is disaggregated; when there are many reviews, identification will be no different between groups.

Number of Outliers and Attribution

According to our conceptualization, format-dependent differences in perceptions of the number of reviews are related to whether outlier reviews are attributed to reviewer or product reasons. We argue that an outlier (e.g., 10% of reviews are 1 star) is more likely to be attributed to reviewer reasons when that outlier is based on a small absolute number of reviews (e.g., 1) compared to a large absolute number of reviews (e.g., 10). This argument is consistent with models of social influence, which argue that greater numbers of sources of opinion are associated with stronger influence (Latané and Wolf 1981; Cialdini 2009). According to these models, a minority opinion of one is more likely to be seen as reflecting an internal attribution such as an idiosyncratic or biased perspective (Moscovici 1985).

We suggest that when review scores are presented in disaggregated format, people will relatively accurately assess the absolute number of reviews forming the outlier opinion. If there is a small absolute number, the outlier will likely be attributed to reviewer reasons whereas if there is a large absolute number, the outlier will likely be attributed to product reasons. In contrast, when review scores are presented in aggregated format, people will relatively inaccurately assess the absolute number of reviews forming the outlier opinion. Therefore, people will be insensitive to the absolute number of reviews and follow the default attribution. Given that previous research demonstrates that people tend to attribute reviews to product reasons unless given reason not to (Qiu, Pang, and Lim 2012), we assume this is the default attribution. Therefore, when review score information is aggregated, we expect that outlier review scores will likely be attributed to product reasons, regardless of the number of reviews.

H3. People will be more likely to attribute outlier scores to reviewer (vs. product) reasons when review score information is presented in a disaggregated (vs. aggregated) format.

Attribution and Perceived Credibility

According to our conceptualization, attributing an outlier review score to reviewer rather than product reasons will have a number of downstream consequences for how that review is treated. One of these consequences relates to credibility, which refers to the extent to which a person perceives a review as believable, true, or factual (Cheung et al. 2009). Consistent with previous research (Qiu, Pang, and Lim 2012), we expect that attributing a review to reviewer reasons reduces the perceived credibility of that review. This is because, in general, a review based on reviewer reasons is likely to be less informative to a consumer than a review based on product reasons.

H4. People will perceive outlier review scores to be less credible when review score information is presented in a disaggregated (vs. aggregated) format.

H5. The effect of format on perceived credibility (i.e., H4) will be mediated by attribution of the outlier review score.

Credibility and Preference

People tend to rely on information that they perceive as credible (Cheung et al. 2009; Petty and Cacioppo 1986). Indeed, prior research has shown that perceived credibility is one of the most important antecedents of eWOM adoption (Chih et al. 2013; Fan et al. 2013; Wathen and Burkell 2002). Therefore, the effect of format on credibility should have flow-on effects for how people use review score information to form relative preferences between alternatives in a consideration set that vary in terms of the distribution of review scores.

H6. People will be more likely to form product preferences that discount outlier review scores when those scores are presented in a disaggregated (vs. aggregated) format. Manifestation of this discounting will depend on the valence of the outlier review (e.g., a discounted positively-valenced review will reduce product preference whereas a discounted negatively-valenced review will increase product preference).

H7. The effect of format on preferences (i.e., H6) will be mediated by attribution of the outlier review.

As described earlier, outlier attribution for those presented with disaggregated information will depend on the absolute number of reviews forming the outlier opinion.

H8. The effect of format on preferences (i.e., H6) will be moderated by the number of reviews: when there are few reviews, preferences will be consistent with discounting outlier review scores when information is disaggregated; when there are many reviews, preferences will not differ between groups.

There are many marketplace situations in which consumers have access to both aggregated and disaggregated review score information. Although in such contexts disaggregated review score information is informationally redundant, our conceptualization suggests that in practice it may not be, and may actually dominate if consumers choose to consider it. First, only disaggregated information allows people to accurately track the frequency of outliers, which can impact upon perceptions of small and large minorities (Hasher and Zacks 1984). Second, such tracking occurs automatically – with little directed intention or effort (Zacks and Hasher 2002) – and is thus consistent with people acting as cognitive misers. Therefore, we argue that those presented with both information formats will tend to rely on disaggregated information. Subsequently, and similar to those presented with disaggregated information alone, small minority outliers should be attributed to reviewer reasons, and then discounted.

H9. When people are presented with both disaggregated and aggregated information they will tend to behave more like they do when presented with disaggregated information alone than aggregated information alone.

Experiments

In order to test these nine hypotheses, we carried out three online experiments. In each experiment we manipulated the format of review score information and the distribution of review scores. The first experiment, which tested H3, H4, H5, and H9, presented participants with a single product together with review score information for 10 reviews. Participants were asked to reflect on an outlier review score, attribute it to reviewer or product reasons, and then judge its credibility. The second experiment, which tested H1, H3, H6, H7 and H9, presented participants with two products each with review score information for 10 reviews. Participants were asked to indicate a preferred product, attribute an outlier review to reviewer or product reasons, and also identify the total number of reviews. The third experiment, which tested H1, H2, H3, H6, and H8, was similar to the second experiment but also presented the two products with review score information for 50 reviews each. As described below, we found support for all hypotheses.

Experiment 1

In Experiment 1, we examined whether people's attribution of an outlier review to reviewer (vs. product) reasons varied depending on the format that the outlier review was presented in, and whether such a difference impacted perceived credibility of the review. For the purpose of generalizability, we examined four review score distributions that varied the average review score and valence of the outlier review to be extremely positive or extremely negative. We expected the effect of format on attribution to hold regardless of the distribution. In order to equate the information across formats and also remain consistent with the procedure of Wulff, Hills, and Hertwig (2014), we removed review comments and only presented review scores.

Methods

The participants were 602 American adults (336 females; M_{age} = 32.7, SD_{age} = 10.7) recruited online from Amazon's Mechanical Turk (AMT).

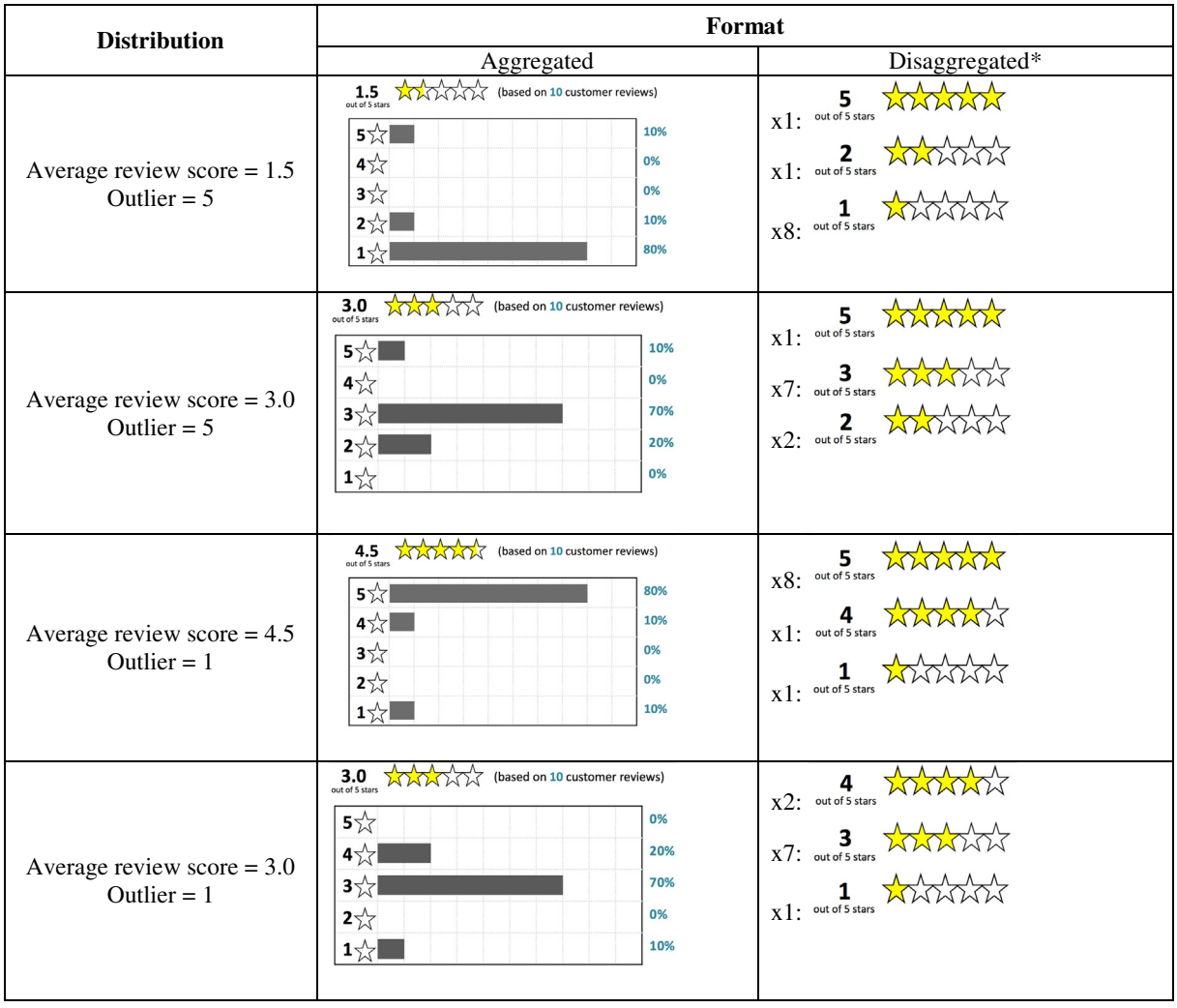
The study was conducted online. Participants were asked to evaluate an audiobook given its title, description, and product review score information. The audiobook was *Last Bus to Wisdom* by Ivan Doig. The accompanying audiobook summary was six sentences long. The audiobook was associated with 10 reviews. The specific review score information presented to participants varied depending on which of twelve groups the participant had been randomly allocated to. Formally, the experiment used a 3 (Format: Aggregated [A] vs. Disaggregated [D] vs. A + D) × 4 (Distribution [of the review scores]) between-subjects design.

Participants in the aggregated group were presented with a frequency chart outlining the average review score out of 5, a visual depiction of this average in yellow-colored stars, the number of reviews, and the percentage of each star category (see Fig. 2, middle panel). Those in the disaggregated group observed each of the review scores in numerical and visual form, individually and sequentially over pages (see Fig. 2, right panel). The order of the review scores was randomized. Those in the A + D group were first presented with the frequency chart and then the individual review scores were presented individually and sequentially over pages.

Participants were presented with review scores in one of four distributions (see Fig. 2). The first distribution had an average review score of 1.5 and an outlier 5 score. The second distribution had an average review score of 3.0 and an outlier 5 score. The third distribution had an average review score of 3.0 and an outlier 1 score. The fourth distribution had an average review score of 4.5 and an outlier 1 score.

Once participants had viewed the review scores they were asked to think about the outlier review score and indicate whether they attributed the review to product reasons or to reviewer reasons. In the question, definitions and examples of "product reasons" and "reviewer reasons" were provided. This question, which was developed from a number of previous studies (Laczniak, DeCarlo, and Ramaswami 2001; Park and Han 2008; Qiu, Pang, and Lim 2012), was answered on a 7-point scale anchored at one end with, "This reviewer's rating was based entirely on reasons associated with the true quality of the audiobook" and at the other end with, "This reviewer's rating was based entirely on reasons that have nothing to do with the true quality of the audiobook". Next, participants were asked to indicate the credibility of the outlier review score. Credibility was measured using three 7-point scales measuring trustworthiness, reliability, and credibility (Cronbach's alpha = 0.95).

On the next page participants answered a manipulation check question and some attention check questions. The manipulation check question asked, "To what extent do you feel that the audiobooks were disliked or liked by the reviewers?" on a 10-point scale ranging from "Very much disliked" to "Very much



*Review scores were presented in random order.

Fig. 2. Summary of the stimuli used in Experiment 1. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

liked". The attention checks were basic questions about the type of product presented in the scenario, and the score of the outlier review. Given that the vast majority of participants were able to accurately answer the attention check questions, we included all participants in the data analysis for all studies. The experiment concluded with a series of standard demographic questions.

Results and Discussion

A manipulation check confirmed that participants felt that the audiobook was liked more when the average review score was 4.5 (M = 8.5, SD = 1.7) than 3.0 (M = 5.8, SD = 1.6), $t(449) = 16.56$, $p < .0001$. In addition, participants felt that the audiobook was liked more when the average review score was 3.0 than 1.5 (M = 2.3, SD = 1.9), $t(451) = 20.17$, $p < .0001$.

The average attribution across groups is presented in the upper section of Table 1. To investigate H3, we conducted an ANOVA with format, distribution, and their interaction entered as independent variables, and attribution entered as the dependent variable. The analysis revealed a significant main effect of format, $F(2, 590) = 5.29$, $p = .005$, no main effect of distribution, $F(3, 590) = 2.11$, $p = .10$, and no interaction, $F(6, 590) = 0.34$, $p = .91$. In support of H3, follow-up contrasts revealed that those in the aggregated group were less likely to attribute the outlier review score to reviewer (vs. product) reasons compared to those in the A + D group, $F(1, 590) = 4.48$, $p = .03$, and those in the disaggregated group, $F(1, 590) = 10.23$, $p = .001$. In support of H9, there was no difference in attribution between those in the A + D and disaggregated groups, $F(1, 590) = 1.17$, $p = .28$.

The average perceived credibility across groups is presented in the lower section of Table 1. To investigate H4, we conducted a similar ANOVA this time with credibility entered as the dependent variable. The analysis revealed a significant main effect of format, $F(2, 590) = 6.03$, $p = .003$, a significant main effect of distribution, $F(3, 590) = 10.53$, $p < .0001$, and no interaction, $F(6, 590) = 0.83$, $p = .55$. In support of H4, follow-up contrasts revealed that those in the aggregated group were more likely to perceive the outlier review score as credible compared to those in the A + D group, $F(1, 590) = 11.42$, $p = .0008$, and also those in the disaggregated group, $F(1, 590) = 5.67$, $p = .02$. In support of H9, there was no difference in perceived credibility between those in the A + D and disaggregated groups, $F(1, 590) = 0.99$, $p = .32$.

Table 1
A summary of attribution and perceived credibility responses across groups in Experiment 1.

Measure	Distribution		Format		
			Aggregated (A)	A + D	Disaggregated (D)
Attribution	Average score = 1.5	*M*	4.45	4.63	5.12
	Outlier = 5	*SD*	1.93	2.02	1.62
	Average score = 3.0	*M*	4.14	4.57	4.64
	Outlier = 5	*SD*	1.56	1.74	1.71
	Average score = 3.0	*M*	4.35	4.80	5.14
	Outlier = 1	*SD*	1.94	1.79	1.58
	Average score = 4.5	*M*	4.71	5.16	5.04
	Outlier = 1	*SD*	1.87	1.81	1.94
Credibility	Average score = 1.5	*M*	3.25	2.71	3.20
	Outlier = 5	*SD*	1.22	1.45	1.52
	Average score = 3.0	*M*	4.18	3.70	3.81
	Outlier = 5	*SD*	1.42	1.17	1.54
	Average score = 3.0	*M*	3.61	2.89	3.20
	Outlier = 1	*SD*	1.44	1.39	1.58
	Average score = 4.5	*M*	3.40	3.16	2.84
	Outlier = 1	*SD*	1.66	1.49	1.52

In order to investigate H5, we conducted a mediation analysis using Hayes' (2012) bootstrapping PROCESS model. We simplified the analysis by excluding the group of participants presented with both disaggregated and aggregated information and collapsing the distribution variable. Therefore, our mediation analysis (Model 4, 5,000 bootstrap samples) tested the effect of format (0 = aggregated, 1 = disaggregated) on perceived credibility of the outlier review (1 = very low to 7 = very high) via attribution (1 = entirely product reasons to 7 = entirely reviewer reasons). Supporting H5, the analysis revealed a significant indirect effect (95% confidence interval: −.17, −.04), which suggests that the effect of format on perceived credibility was partially mediated by attribution such that those who received disaggregated information were more likely to attribute outlier reviews to reviewer reasons, and those who attributed outlier reviews to reviewer reasons were less likely to perceive the outlier review as credible.

In summary, Experiment 1 provides strong evidence that outlier reviews can be attributed to reviewer (vs. product) reasons, and that this attribution is more likely when the outlier review is learned from disaggregated (vs. aggregated) information formats. This format-dependent attribution appears to occur for low, moderate, and high rated products, and regardless of whether the outlier review is extremely low or extremely high. Moreover, those presented with both disaggregated and aggregated information tended to give responses similar to those presented with disaggregated information alone even though in the former case this information is redundant after having already been presented with the aggregated information.

We also showed that differences in review score attribution had a flow-on effect to the perceived credibility of the review score: reviews attributed to reviewer reasons were perceived as less credible. We expect that the relatively low credibility of the outlier score will lead to this review being discounted, which should have consequences for consumer preference and choice. We tested this expectation in Experiments 2 and 3.

Experiment 2

In Experiment 2, we examined whether people's preference for a product was influenced by the format in which the product's review scores were presented. We modeled our experiment on research conducted by Wulff, Hills, and Hertwig (2014), but ensured that participants were presented with equivalent information. Based on our conceptualization of how participants would process outlier reviews across the different information formats (Fig. 1), we expected that participants would be more likely to form product preferences that discounted outlier review scores when those scores were presented in a disaggregated (vs. aggregated) format. This expectation produced different predictions depending on the average review score of products in the consideration set. Therefore, in this experiment we presented participants with products that either had very low average reviews (i.e., 1.5 out of 5) or very high average reviews (i.e., 4.5 out of 5). When the average review scores of products in the consideration set were all relatively low and one product had a high-scoring outlier review, we expected that people would be *less* likely to prefer the high-variance (vs. low variance) product when review scores were presented in a disaggregated (vs. aggregated) format. In contrast, when the average review scores of products in the consideration set were all relatively high and one product had a low-scoring outlier review, we expected that people would be *more* likely to prefer the high-variance (vs. low variance) product when scores were presented in a disaggregated (vs. aggregated) format.

Methods

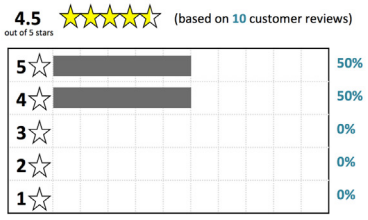
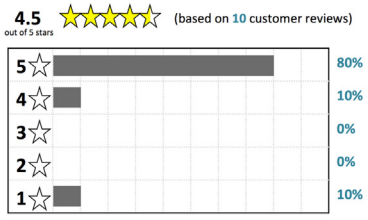
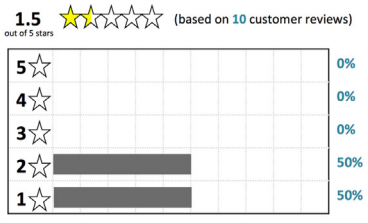
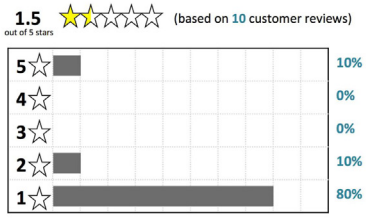
The participants were 302 American adults (169 females; M_{age} = 31.2, SD_{age} = 10.1) recruited online from AMT.

The study was conducted online. Participants were asked to choose between two audiobooks, "Book A" and "Book B," for which the only information available was product reviews. Participants were separately and sequentially presented with the

review score information for the two audiobooks (or in reverse order, depending on the counterbalancing order). Book A was associated with 10 review scores that clustered narrowly around the average (low variance). Book B was associated with 10 review scores that clustered widely around the mean (high variance). The specific information presented on these pages varied depending on which of six groups the participant had been randomly allocated to. Formally, the experiment used a 3 (Format: Aggregated [A] vs. Disaggregated [D] vs. A + D) × 2 (Average Review Score [out of 5]: 1.5 vs. 4.5) between-subjects design. The operationalization of format was the same as in Experiment 1. The average review score of the 10 reviews was set at either 1.5 out of 5 or 4.5 out of 5 for both products (see Fig. 3).

Once participants had viewed the review scores for both audiobooks they were asked to choose their preferred one in a binary choice question, and then explain their choice using a free response text box. Note that a tendency to discount outliers is indicated by a preference for the low-variance option in the context of low average review scores and a preference for the high-variance option in the context of high average review scores.

Participants were then instructed to think about the outlier review score and indicate whether they attributed the review to product reasons or reviewer reasons in a binary choice question. The two options were, "There is something unusual about the reviewer (e.g., they had unrealistic expectations, they are lying, they are an idiot, etc.)" and "There is at least one person who really liked this audiobook". Participants next answered the same manipulation check question as in Experiment 1, some attention check questions, and a question designed to measure the participant's ability to identify the total number of reviews.

Average Review Score	Format	
	Aggregated	Disaggregated*
4.5	Low variance option:  High variance option: 	Low variance option:  High variance option: 
1.5	Low variance option:  High variance option: 	Low variance option:  High variance option: 

*Review scores were presented in random order.

Fig. 3. Summary of the stimuli used in Experiment 2.

The attention checks were basic questions about the type of product presented in the scenario, and the number of options. The identification question asked, "Approximately how many reviews were there for each audiobook?" with the response options 10, 50, 100, and 150. The experiment concluded with a series of standard demographic questions.

Results and Discussion

A manipulation check confirmed that participants felt that the audiobooks were liked more when the average review score was 4.5 (M = 8.2, SD = 1.4) than when it was 1.5 (M = 2.5, SD = 1.4), $t(300) = 34.04$, $p < .0001$.

To test H1, we conducted a logistical regression analysis with format, average review score, and their interaction entered as independent variables, and ability to correctly identify the total number of reviews entered as the dependent variable. We also entered the order in which products were presented as a covariate. In support of H1, we found a significant main effect for format, χ^2 (1, N = 302) = 13.14, p = .001, indicating that participants were better able to identify the number of reviews when review score information was presented in disaggregated format (96.1%) or combined disaggregated and aggregated format (96.0%) compared to aggregated format (83.8%). There were no significant effects of order, χ^2 (1, N = 302) = 0.005, p = .94, average review score, χ^2 (1, N = 302) = 0.66, p = .42, or the interaction between format and average review score, χ^2 (1, N = 302) = 0.81, p = .67.

To test H3, we conducted a logistical regression analysis with format, average review score, and their interaction entered as independent variables, and response to the binary attribution question entered as the dependent variable. We also entered the order in which products were presented as a covariate. In support of H3, we found a significant main effect for format, χ^2 (1, N = 302) = 11.64, $p < .0001$, indicating that participants were more likely to attribute outlier scores to reviewer reasons when review score information was presented in a disaggregated format (46.6%) or combined disaggregated and aggregated format (51.0%), compared to an aggregated format (29.3%). There were no significant effects of order, χ^2 (1, N = 302) = 3.26, p =.07, average review score, χ^2 (1, N = 302) = 3.35, p = .07, or the interaction between format and average review score, χ^2 (1, N = 302) = 1.34, p = .51.

To test H6, we conducted the same logistical regression analysis but this time with choice entered as the dependent variable. In support of H6, we found a significant interaction effect between format and average review score, χ^2 (1, N = 302) = 18.02, $p < .0001$, indicating that participants made different product choices depending on format, and that this effect was moderated by average review score. There were no significant effects of order, χ^2 (1, N = 302) = 0.08, p = .77, average review score, χ^2 (1, N = 302) = 2.96, p = .09, or format, χ^2 (1, N = 302) = 0.05, p = .97.

As shown in Fig. 4, when the average review scores were 1.5 (left panel), participants were *less* likely to prefer the high-variance product when scores were presented in disaggregated format. In contrast, when the average review scores were 4.5, participants were *more* likely to prefer the high-variance product when scores were presented in disaggregated format. These results are consistent with the idea that people are more likely to make choices as if discounting outlier review scores when presented with information in disaggregated format. It is noteworthy that, unlike previous research findings (i.e., Wulff, Hills, and Hertwig 2014), these choice differences were observed despite participants being presented with equivalent information. Thus, our observations are consistent with our conceptualization that different psychological processes are engaged by different information formats.

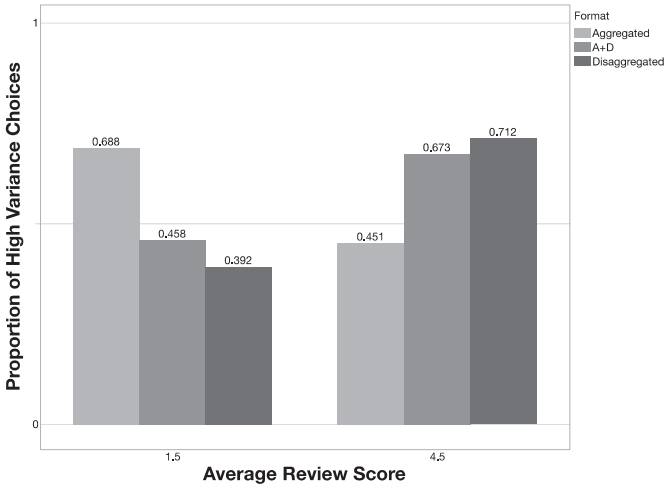


Fig. 4. Proportion of high variance choices across groups in Experiment 2.

To test H7, we conducted a simple mediation analysis focusing on the effect of format (0 = aggregated, 1 = disaggregated) on choice (0 = less weight on outlier, 1 = more weight on outlier) via attribution (0 = reviewer reason, 1 = product reason). There are three points to note about this analysis. First, as in Experiment 1, we excluded participants presented with both disaggregated and aggregated information. Second, given that the choice effects were moderated by average review score, we recalculated a new dependent variable that indicated whether the participant's choice was consistent with putting relatively less or more decision weight on the outlier review. In practice, this simply meant recoding choices for the "low variance" option to a "more weight on outlier" choice when the average review score was 4.5, and recoding choices for the "low variance" option to a "less weight on outlier" choice when the average review score was 1.5. Third, given that all variables in the analysis were binary, we were unable to use Hayes' (2012) bootstrapping PROCESS model. Therefore, we used the solution proposed by MacKinnon and Dwyer (1993), which relies on the traditional Sobel test. A summary of the mediation analysis is depicted in Fig. 5. Supporting H7, the test statistic for the Sobel test was significant, −2.28, p = .02. Therefore, the effect of format on choice was partially mediated by attribution such that those who received disaggregated information were more likely to attribute outlier reviews to reviewer reasons, and those who attributed outlier reviews to reviewer reasons were more likely to discount outlier reviews when making a choice.

To test H9, we conducted a series of follow-up contrasts examining the effect of different formats on choice. In support

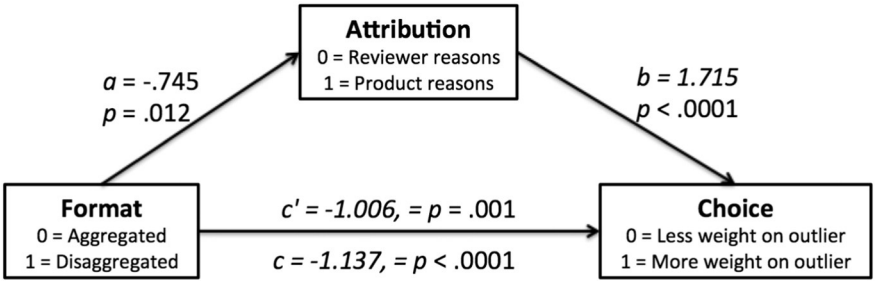


Fig. 5. Mediation analysis in Experiment 2.

of H9, there was no significant difference in the choice made between those presented with disaggregated information and those presented with both disaggregated and aggregated information when the average review score was 1.5, χ^2 (1, N = 99) = 0.44, p = .51, or 4.5, χ^2 (1, N = 104) = 0.18, p = .67.

In summary, Experiment 2 replicated the finding that outlier reviews can be attributed to reviewer (vs. product) reasons, and that this attribution is more likely when the outlier review is learned from disaggregated (vs. aggregated) information formats. We also showed that this attribution had a flow-on effect to product preferences: reviews attributed to reviewer reasons were discounted, which made high variance products more preferable when the average review score was high, and low variance products more preferable when the average review score was low. Also replicating the pattern from Experiment 1, when participants were presented with both disaggregated and aggregated information they tended to behave as if they had been presented with disaggregated information alone.

Consistent with our conceptualization in Fig. 1, participant's ability to correctly identify the number of reviews was higher when information was presented in the disaggregated (vs. aggregated) format. According to our theory, the difference between formats should close when many review scores are considered. This moderation was predicted because those presented with disaggregated information should recognize that the outliers represent a large minority, which should increase the tendency to attribute these reviews to product (vs. reviewer) reasons similar to those in the aggregated format group. We tested this expectation in Experiment 3.

Experiment 3

In Experiment 3, we tested a key assumption of our conceptualization by observing whether the effect of format on choice observed in Experiment 2 was attenuated by the presentation of many review scores and, hence, several outlier review scores. In this experiment, we focused on the two most extreme formats – disaggregated and aggregated information – to test the theory.

Methods

The participants were 403 American adults (254 females; M_{age} = 33.0, SD_{age} = 11.0) recruited online from AMT.

The design, materials, and procedure were very similar to those employed in Experiment 2 with three main exceptions. First, the combined disaggregated and aggregated group was removed to reduce the complexity of the experimental design. Second, the number of review scores associated with each audiobook was manipulated such that half of the participants observed 10 review scores and the other half observed 50 review scores. Formally, the experiment used a 2 (Format: Aggregated vs. Disaggregated) × 2 (Average Review Score [out of 5]: 1.5 vs. 4.5) × 2 (Number of Reviews: 10 vs. 50) between-subjects design. Third, the binary attribution question used in Experiment 2 was replaced with an interval scale question that asked participants, "To what extent do you think that the scores from this minority of reviewers had to do with their own personal idiosyncratic preferences as opposed to a true reflection of the audiobook's value?". The question was answered on a 10-point scale anchored at one end with, "Their scores were definitely *not* influenced by personal reasons," and at the other end with, "Their scores were definitely influenced by personal reasons".

Results and Discussion

A manipulation check confirmed that participants felt that the audiobooks were liked more when the average review score was 4.5 (M = 8.1, SD = 1.6) than when it was 1.5 (M = 3.1, SD = 1.9), t(401) = 28.36, p < .0001.

To test H1 and H2, we conducted a logistical regression analysis with format, average review score, number of reviews, and all their interactions entered as independent variables, and ability to correctly identify the total number of reviews entered as the dependent variable. We also entered the order in which products were presented as a covariate. In support of H1, we found a significant main effect for format, χ^2 (1, N = 403) = 14.21, p = .0002. In support of H2, this main effect was qualified by a significant interaction between format and number of reviews, χ^2 (1, N = 403) = 27.78, p < .0001, indicating that participants' ability to identify the number of reviews differed depending on format, and that this effect was moderated by the number of reviews. When the number of reviews was 10, participants had better accuracy when review score information was presented in disaggregated format (96.1%) than aggregated format (67.0%). In contrast, when the number of reviews was 50, there was no difference in accuracy as a function of whether review score information was presented in disaggregated (64.6%) or aggregated format (72.3%). There were no significant effects of order, χ^2 (1, N = 403) = 0.36, p = .54, average review score, χ^2 (1, N = 403) = 0.43, p = .51, nor the interaction between format and average review score, χ^2 (1, N = 403) = 0.83, p = .36. However, there was a significant main effect for number of reviews, χ^2 (1,

$N = 403) = 16.51$, $p < .0001$, that was qualified by a significant three-way interaction, χ^2 (1, $N = 403$) = 10.69, $p = .001$.

To test H3, we conducted an ANOVA with format, average review score, number of reviews, and all their interactions entered as independent variables, and response to the attribution question entered as the dependent variable. We also entered the order in which products were presented as a covariate. Unlike Experiments 1 and 2, the analysis failed to support H3 in that we did not find a significant main effect for format, $F(1, 394) = 2.89$, $p = .09$. However, the direction of the difference was consistent with our prediction in that participants were more likely to attribute outlier scores to reviewer reasons when review score information was presented in disaggregated format ($M = 7.58$, $SD = 1.98$) compared to aggregated format ($M = 7.25$, $SD = 1.99$). We also note that we received some feedback from participants indicating that the wording of this question was confusing, which could have reduced our ability to detect a difference. There were no significant effects of order, $F(1, 394) = 0.87$, $p = .35$, average review score, $F(1, 394) = 1.99$, $p = .16$, number of reviews, $F(1, 394) = 0.14$, $p = .70$, nor any of the interactions (all ps > .14).

To test H6 and H8, we conducted another logistical regression analysis this time with choice entered as the dependent variable. In support of H8, we found a significant three-way interaction between format, average review score, and number of reviews, χ^2 (1, $N = 302$) = 5.63, $p = .02$, indicating that the moderated effect of format on choice identified in Experiment 2 changed as a function of number of review scores. As shown in Fig. 6, and supporting H6, when the number of reviews was 10, we replicated the pattern of results observed in Experiment 2 such that participants were more likely to choose as if discounting outlier review scores when presented with information in disaggregated format. In contrast, when the number of reviews was 50, this pattern of results was eliminated (and actually reversed when the average review score was 4.5). In addition, there was a significant effect of order, χ^2 (1, $N = 403$) = 4.66, $p = .03$, reflecting that the proportion of high variance choices was higher when the high variance option was presented second. There were no significant main effects of format, χ^2 (1, $N = 403$) = 2.05, $p = .15$, average review score, χ^2 (1, $N = 403$) = 1.32, $p = .25$, number of reviews, χ^2 (1, $N = 403$) = 2.92, $p = .09$, nor any of the other two-way interactions (ps > .29) except between number of reviews and average review score, χ^2 (1, $N = 403$) = 6.05, $p = .01$.

In summary, Experiment 3 replicated the finding that outlier reviews tend to be discounted when presented in disaggregated (vs. aggregated) format. Consistent with our theory, this tendency changed when many review scores were presented. Specifically, when the average review scores of the products under consideration were both low, the difference between formats disappeared, and when the average review scores of the products under consideration were both high, the difference between formats actually reversed. Also consistent with our theory, participant's ability to correctly identify the total number of reviews was similarly moderated by the number of reviews. However, unlike the first two experiments, we found only directional support for our hypothesis regarding outlier review attribution. We suspect

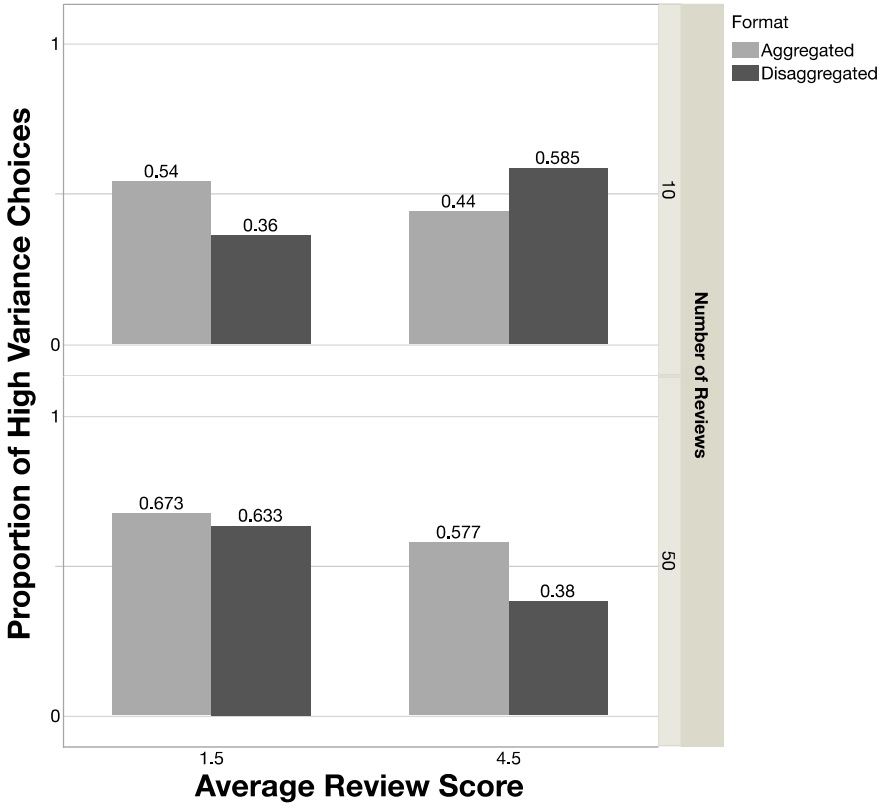


Fig. 6. Proportion of high variance choices across groups in Experiment 3.

that this variable was associated with considerable noise given that participants had difficulty interpreting the question.

General Discussion

The present work adds to the growing literature on consumers' use of online review scores (Cheung and Thadani 2012), and specifically the research of Wulff, Hills, and Hertwig (2014), by showing that people can make different product choices depending on whether review score information is aggregated or disaggregated. Specifically, when choosing between a low- and high-variance product, people were more likely to form preferences as if outlier review scores were discounted when those scores were presented in disaggregated format. Thus, for example, when review scores of both products were generally positive (i.e., 4.5 out of 5 stars), more people preferred the high-variance option when they learned about reviews by sequentially seeing individual review scores than when the same information was summarized in a frequency histogram.

Our work extends these previous findings in a number of ways. First, we show that format-dependent choices can be observed even when the review scores presented to participants in the different format groups is informationally equivalent. This is an important revelation because the only previous study to show such format-dependent differences in the context of review scores confounded information format with amount of information available, thereby obscuring whether there were actually format-dependent differences (Wulff, Hills, and Hertwig 2014). Second, we show that these format-dependent choices can occur in the context of a low average review score consideration set and also a high average review score consideration set. Third, we show that when people are presented with *both* aggregated and disaggregated information together, they tend to choose as if they had been presented with disaggregated information alone. Fourth, we reveal that one boundary condition to this effect is how many review scores are considered such that the effect is reduced – sometimes even reversed – when many reviews are presented. Lastly, and perhaps most importantly, we provide preliminary evidence for a psychological mechanism driving these behaviors: attribution.

We interpret our observations in the context of Kelley's (1967, 1973) covariation model of attribution. According to our account, consumers automatically attempt to attribute outlier review scores to either product or reviewer reasons and, when the latter is applied, tend to discount this information (Laczniak, DeCarlo, and Ramaswami 2001; Park and Han 2008; Qiu, Pang, and Lim 2012). The explanation is supported by mediation analyses showing that reviews attributed to reviewer reasons were discredited (Experiment 1) and discounted in choice (Experiment 2). Comments made by our participants also support this interpretation. For example one participant in Experiment 2 wrote: "I figure you have to throw out one of the highest and one of the lowest scores. There is always somebody that was unhappy because they misunderstood and purchased the wrong product, had lousy shipping, or something else largely unrelated to the quality of the actual item".

Our work therefore adds to existing findings that implicate attribution in the effect of online reviews on product evaluations (Lee and Youn 2009; Park and Han 2008; Qiu, Pang, and Lim 2012; Sen and Lerman 2007). Importantly, our study extends these findings by demonstrating that attributions can be made about individual reviews based on the distribution of other scores, and that these attributions are influenced by the presentation format. Indeed, our major theoretical contribution is that the tendency to attribute an outlier review to reviewer reasons and subsequently discount it is higher when review score information is presented in disaggregated format. We argue that this occurs because consumers are more likely to attribute an outlier review to reviewer reasons when there is only a single outlier (for example, one that makes up 10% of 10 total reviews) rather than a group of outliers (for example, five that make up 10% of 50 total reviews). This point is highlighted by a comment left by one participant from the group presented with disaggregated information and 50 reviews: "While Book B appeared to have a lot of 5 star reviews, it had too many 1 star reviews for me to ignore".

Our key insight was that people perceive outliers, what you might think of as minority opinions, differently depending on presentation format. This insight stems from previous research showing that people automatically track frequencies when information is presented in disaggregated format (Hasher and Zacks 1984) but tend to overlook frequencies when information is presented in aggregated format (Griffin and Tversky 1992; Obrecht, Chapman, and Gelman 2007). Consistent with this interpretation, we found that people had very good identification of the number of reviews when information was presented in disaggregated format compared to when the same information was presented in aggregated format. As a result, when information was disaggregated, small minorities (e.g., 1 outlier review) tended to be attributed to reviewer reasons and discounted. In contrast, when information was aggregated, small minorities tended to be attributed to product reasons and not discounted.

One result we observed that was unpredicted is shown in the lower panel of Fig. 6: when the average review score of products in the consideration set was high and there were 50 reviews, the proportion of high variance choices was significantly lower in the disaggregated compared to aggregated group. In other words, when there were 10 reviews, people in the disaggregated group chose as if putting relatively less weight on the (low-scoring) outlier review score. In contrast, when there were 50 reviews, people in the disaggregated group chose as if putting relatively more weight on the (low-scoring) outlier review score. We speculate that participants in this group may have given significant meaning to the minority group of outlier reviews by interpreting them as a systematic problem with the product. For example, one participant reported, "Audiobook A had either 4 of 5 and 5 of 5 ratings only. Whereas B had a combination of 1 of 5 or 4 of 5 and 5 of 5. This might mean that B might have some sort of problem that comes only so many books".

Managerial Implications

Our results have managerial implications regarding how product and service evaluations are influenced by review score

presentation format. To some degree, the number of reviews that people actually consider limits our findings. This is because the differences that we observed were most prominent when participants considered only 10 reviews per option. Importantly, when we asked the same group of 104 American participants mentioned in the introduction when they did read reviews how many they considered prior to making a purchase, the average response was 8.8 (*SD* = 10.8). Therefore, we believe that our results do have substantial practical implications for managers.

Most products are associated with many high scores together with a few low scores (Hu, Pavlou, and Zhang 2007). Given that most websites tend to present aggregated review score information prominently (e.g., directly under the product title), our results suggest that consumers may have an overall lower evaluation of products than may be warranted. Therefore, a manager may strategically decide to eliminate the presentation of aggregated review score information; that is, remove the review score frequency distribution as has been done at Zomato.com. Relative to its rivals, a website that contains only disaggregated review scores may lead consumers to have an overall higher evaluation of products listed on that website. The ultimate effect of overall higher versus lower evaluation of products will depend on how accurate these evaluations are. Interestingly, some have argued that online user ratings are actually poorly correlated with objective quality information (De Langhe, Fernbach, and Lichtenstein 2015).

Of course, most websites will present both aggregated and disaggregated review score information. In these cases, managers could strategically use the present findings by making it relatively easier to see the disaggregated reviews for favored products. This could be accomplished by, for example, placing some individual reviews near the top of a product's webpage, or by displaying many individual reviews simultaneously in a compact space (e.g., a word cloud of review scores). The likely effect is to increase the overall positive association of the favored products.

Limitations and Future Research

As with all laboratory experiments, in trying to achieve a balance between external and internal validity, our design was imperfect. One limitation was that accompanying reviewer comments were not presented. Previous research has shown that review comments are an important factor affecting consumer behavior (e.g., Packard and Berger 2016). However, like Wulff, Hills, and Hertwig (2014), we made this design decision to ensure that information was held constant across all groups, which would not have been possible if comments were attached to each disaggregated review. Future research could explore how the format-dependent behavior discussed in this paper is moderated by different review comments. It would be interesting to learn, for example, what comments in a written review could "save" an outlier review from being discounted.

A second limitation is that we did not allow participants to sample as many reviews as desired, which is the case in the actual marketplace. Rather, our participants sampled either 10 or 50 reviews. As discussed earlier, this design decision was made in order to equate information between groups. Importantly, the observations made here correspond to similar observations made when participants were free to select the number of reviews (Wulff, Hills, and Hertwig 2014), which provides converging evidence for our conclusions. Moreover, as mentioned in the introduction, people report on average considering around 9 reviews prior to making a purchase, suggesting that our obliged sample size of 10 was not far away from participants' average tendency. Although consumers rarely sample 50 reviews, this extreme condition allowed us to test critical predictions of our theory, which were supported.

A third limitation was that the disaggregated information was presented in a random order whereas the aggregated information was presented in highest-to-lowest order. Similarly, when aggregated information and disaggregated information were both presented, they were only presented in this order. We made these order decisions to better replicate real marketplace conditions where the order of disaggregated reviews is not systematically ordered, and usually only considered after the aggregated review score information has first been observed.

A fourth limitation was that we did not directly test the link between credibility and preference. Previous research suggests that this relationship exists (Chih et al. 2013; Fan et al. 2013; Wathen and Burkell 2002); however, it will be useful for future research to examine if and how this relationship is moderated by the format and number of reviews.

Conclusion

In summary, we have shown that a review can be attributed to reviewer reasons and discounted simply by being an outlier in the distribution of review scores. This tendency is more likely when review scores are presented in a disaggregated format, such as individual review scores read one at a time, than in an aggregated format, such as reading a summary frequency distribution chart. These findings have implications for consumer choice and the optimal design of review score aggregator websites.

Acknowledgments

The author was supported by an Australian Endeavour Fellowship received from the Australian Government grant number: 2787_2012, an Alcoa Foundation Fellowship received from the American Australian Association, and an internal grant from RMIT University College of Business'. The author thanks Rick Larrick, John Roberts, Ben Newell, Jim Bettman, Joel Huber, Elke Weber, Christoph Ungemach, the members of the Duke University MaSC lab, and the reviewers. The author also thanks Peter Camilleri and Mirra Seigerman for their valuable research assistance.

References

Alba, Joseph W. and Howard Marmorstein (1987), "The Effects of Frequency Knowledge on Consumer Decision Making," *Journal of Consumer Research*, 14, 1, 14–25.

Anderson, John R. and Christian J. Lebiere (2014), *The Atomic Components of Thought*. Psychology Press.

Arndt, Johan (1967), "Role of Product-related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 4, 3, 291–5.
Baddeley, Alan (1992), "Working Memory," *Science*, 255, 5044, 556.
Berger, Jonah (2014), "Word of Mouth and Interpersonal Communication: A Review and Directions for Future Research," *Journal of Consumer Psychology*, 24, 4, 586–607.
Chatterjee, Patrali (2001), "Online Reviews: Do Consumers Use Them?" *Advances in Consumer Research*, 28, 1, 129–33.
Cheung, Christy M.K. and Dimple R. Thadani (2012), "The Impact of Electronic Word-of-Mouth Communication: A Literature Analysis and Integrative Model," *Decision Support Systems*, 54, 1, 461–70.
Cheung, Man Yee, Chuan Luo, Choon Ling Sia, and Huaping Chen (2009), "Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations," *International Journal of Electronic Commerce*, 13, 4, 9–38.
Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43, 3, 345–54.
Chih, Wen-Hai, Kai-Yu Wang, Li-Chun Hsu, and Su-Chen Huang (2013), "Investigating Electronic Word-of-Mouth Effects on Online Discussion Forums: The Role of Perceived Positive Electronic Word-of-Mouth Review Credibility," *Cyberpsychology, Behavior and Social Networking*, 16, 9, 658–68.
Cialdini, Robert B. (2009), *Influence: Science and Practice*. Boston: Pearson Education Boston.
De Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016), "Navigating by the Stars: Investigating the Actual and Perceived Validity of 1065 Online User Ratings," *Journal of Consumer Research*, 42, 817–33.
De Matos, Celso Augusto and Carlos Alberto Vargas Rossi (2008), "Word-of-Mouth Communications in Marketing: A Meta-analytic Review of the Antecedents and Moderators," *Journal of the Academy of Marketing Science*, 36, 4, 578–96.
Fan, Yi-Wen, Yi-Feng Miao, Yu-Hsien Fang, and Ruei-Yun Lin (2013), "Establishing the Adoption of Electronic Word-of-Mouth Through Consumers' Perceived Credibility," *International Business Research*, 6, 3, 58.
Fiske, Susan T. and Shelley E. Taylor (2013), *Social Cognition: From Brains to Culture*. Sage.
Gigerenzer, Gerd and Reinhard Selten (2002), *Bounded Rationality: The Adaptive Toolbox*. MIT Press.
Griffin, Dale and Amos Tversky (1992), "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology*, 24, 3, 411–35.
Hasher, Lynn and Rose T. Zacks (1984), "Automatic Processing of Fundamental Information: The Case of Frequency of Occurrence," *American Psychologist*, 39, 12, 1372.
Hau, Robin, Timothy J. Pleskac, Jürgen Kiefer, and Ralph Hertwig (2008), "The Description–Experience Gap in Risky Choice: The Role of Sample Size and Experienced Probabilities," *Journal of Behavioral Decision Making*, 21, 5, 493–518.
Hayes, Andrew F. (2012), "Process: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling," http://www.afhayes.com/public/process2012.pdf.
Hertwig, Ralph and Timonthy J. Pleskac (2010), "Decisions from Experience: Why Small Samples?" *Cognition*, 115, 225–37.
Hu, Nan, Paul A. Pavlou, and Jie Jennifer Zhang (2007), "Why Do Online Product Reviews Have a J-shaped Distribution? Overcoming Biases in Online Word-of-Mouth Communication," *Working Paper*.
Kelley, Harold H. (1967), "Attribution Theory in Social Psychology," *Nebraska Symposium on Motivation*, 15, 192–238.
——— (1973), "The Processes of Causal Attribution," *American Psychologist*, 28, 2, 107–28.
Laczniak, Russell N., Thomas E. DeCarlo, and Sridhar N. Ramaswami (2001), "Consumers' Responses to Negative Word-of-Mouth Communication: An Attribution Theory Perspective," *Journal of Consumer Psychology*, 11, 1, 57–73.
Lang, Bodo and Kenneth F. Hyde (2013), "Word of Mouth: What We Know and What We Have yet to Learn," *Journal of Consumer Satisfaction, Dissatisfaction & Complaining Behavior*, 26, 1–18.
Latané, Bibb and Sharon Wolf (1981), "The Social Impact of Majorities and Minorities," *Psychological Review*, 88, 5, 438–53.
Lee, Mira and Seounmi Youn (2009), "Electronic Word of Mouth (Ewom) How Ewom Platforms Influence Consumer Product Judgement," *International Journal of Advertising*, 28, 3, 473–99.
Lejarraga, Tomás (2010), "When Experience Is Better Than Description: Time Delays and Complexity," *Journal of Behavioral Decision Making*, 23, 100–16.
MacKinnon, David P. and James H. Dwyer (1993), "Estimating Mediated Effects in Prevention Studies," *Evaluation Review*, 17, 2, 144–58.
Moscovici, Serge (1985), "Social Influence and Conformity," *The Handbook of Social Psychology*, G. Lindzey and E. Aronson, eds. New York: Random House, 347–412.
Obrecht, Natalie A., Gretchen B. Chapman, and Rochel Gelman (2007), "Intuitive T Tests: Lay Use of Statistical Information," *Psychonomic Bulletin & Review*, 14, 6, 1147–52.
Packard, Grant and Jonah Berger (2016), "How Language Shapes Word of Mouth's Impact," *Journal of Marketing Research*.
Park, Do-Hyung and Ingoo Han (2008), "Integrating Conflicting Reviews: Attributional Hypotheses of Consumer Response to Information Uncertainty Depending on Prior Brand Attitude," *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, IEEE. 287.
Payne, John W., James R. Bettman, and Eric J. Johnson (1993), *The Adaptive Decision Maker*. Cambridge University Press.
Petty, Richard E. and John T. Cacioppo (1986), "The Elaboration Likelihood Model of Persuasion," *Communication and Persuasion*, Springer. 1–24.
Plonsky, Ori, Kinneret Teodorescu, and Ido Erev (2015), "Reliance on Small Samples, the Wavy Recency Effect, and Similarity-based Learning," *Psychological Review*, 122, 4, 621.
Purnawirawan, Nathalia, Patrick De Pelsmacker, and Nathalie Dens (2012), "Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions," *Journal of Interactive Marketing*, 26, 4, 244–55.
Qiu, Lingyun, Jun Pang, and Kai H. Lim (2012), "Effects of Conflicting Aggregated Rating on Ewom Review Credibility and Diagnosticity: The Moderating Role of Review Valence," *Decision Support Systems*, 54, 1, 631–43.
Sen, Shahana and Dawn Lerman (2007), "Why Are You Telling Me This? An Examination into Negative Consumer Reviews on the Web," *Journal of Interactive Marketing*, 21, 4, 76–94.
Senecal, Sylvain and Jacques Nantel (2004), "The Influence of Online Product Recommendations on Consumers' Online Choices," *Journal of Retailing*, 80, 2, 159–69.
Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels (2009), "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73, 5, 90–102.
Ungemach, Christoph, Nick Chater, and Neil Stewart (2009), "Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)?" *Psychological Science*, 20, 4, 473–9.
Wathen, C. Nadine and Jacquelyn Burkell (2002), "Believe It or Not: Factors Influencing Credibility on the Web," *Journal of the American Society for Information Science and Technology*, 53, 2, 134–44.
Westbrook, Robert A. (1987), "Product/Consumption-based Affective Responses and Postpurchase Processes," *Journal of Marketing Research*, 24, 3, 258–70.
Whyte, William H. (1954), "The Web of Word of Mouth," *Fortune*, 50, 5, 140–3.
Wulff, Dirk U., Thomas T. Hills, and Ralph Hertwig (2014), "Online Product Reviews and the Description–Experience Gap," *Journal of Behavioral Decision Making*, 28, 3, 214–23.
Zacks, Rose T. and Lynn Hasher (2002), "Frequency Processing: A Twenty-five Year Perspective," in *Etc. Frequency Processing and Cognition.* Peter Sedlmeier, Tilmann Betsch, editors. Oxford: Oxford University Press. 21–36.

Dr. Adrian Camilleri is a lecturer in marketing at RMIT University. He holds a bachelor's degree in psychology, a Master's degree in organizational psychology, and a PhD in psychology, all from the University of New South Wales. He completed postdoctoral training in management and marketing at Duke University's Fuqua School of Business. Adrian is an expert in the field of behavioral decision-making and his research has been published in a range of highly reputed academic journals including *Management Science*, *Cognition*, and *Journal of Public Policy & Marketing*.