AMERICAN MARKETING ASSOCIATION

Article

Journal of Marketing
2018, Vol. 82(6) 109-131

DOI: 10.1177/0022242918805468
journals.sagepub.com/home/jmx
SAGE

Swayed by the Numbers: The Consequences of Displaying Product Review Attributes

Jared Watson, Anastasiya Pocheptsova Ghosh, and Michael Trusov

Abstract

Prior research has shown the independent effects of average product ratings and number of reviews for online purchases, but the relative influence of these aggregate review attributes is still debated in the literature. In this research, the authors demonstrate the conditional influences of these two attributes as a function of the valence of average product ratings and the level of review numbers in a choice set. Specifically, they argue that the diagnosticity of the number of reviews, relative to average product ratings, increases when average product ratings are negative or neutral (vs. positive) and when the level of review numbers in a choice set is low (vs. high). As a result, when consumers choose among the best options on one of the review attributes (average product ratings or the number of reviews), their preference shifts from the higher-rated option with fewer reviews toward the lower-rated option with more reviews. The authors demonstrate this preference shift in seven studies, elucidate the underlying process by which this occurs, and conclude with a discussion of the implications for retailers and brands.

Keywords

average product ratings, number of reviews, online product reviews, online retail, choice

Online supplement: https://doi.org/10.1177/0022242918805468

With the rise of Internet shopping, product reviews have gained prominence. Nearly 60% of consumers now say that the average product rating is the most important product attribute in their purchase decisions. Within consumer reviews, 54% of consumers report paying attention to average product ratings, while 46% pay attention to the number of reviews (BrightLocal 2017). Because average product ratings and the number of reviews (referred to as "aggregate reviews" hereinafter) play a significant role in consumer behavior, marketing academics have tried to understand the processes by which consumers incorporate aggregate review information into their purchase decisions (Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Clemons, Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008; Godes and Mayzlin 2004; Ho-Dac, Carson, and Moore 2013; Liu 2006; Moe and Trusov 2011; Sun 2012; Zhu and Zhang 2010).[1]

While the literature has demonstrated the positive influence of both average product ratings and the number of reviews on sales, their relative influence is still debated. Indeed, two recent meta-analyses arrived at opposing conclusions. Floyd et al. (2014) find support for the claim that average product ratings are more influential than the number of reviews, whereas You, Vadakkepatt, and Joshi (2015) argue for the greater importance of the number of reviews.[2] The goal of this article is to clarify

[1] More recently, academics have begun to investigate how the text content of reviews affects consumers' online behavior (Chen and Lurie 2013; Moore 2015; Schlosser 2011). Given that the current research is a systematic investigation of the joint influence of two numerical review attributes, average product ratings and the number of reviews, on consumer choices, the relationship between average product ratings, number of reviews, and review sentiment is outside the scope of our research. We leave this topic for future studies, where the conditions by which consumers incorporate individual review content can be more thoroughly explored.

[2] Similarly, there is no clear consensus from retailers on the importance of these attributes. Some retailers choose to present only average product ratings, whereas others present both ratings and review numbers. In a preliminary exploration of the market, we analyzed the review attribute presentations of over 300 websites and found that although 99 of the 337 websites chose not to display any review information (29%), of the remaining 238 websites, 54% chose to display the number of reviews on the search page, while the others chose instead to display the number of reviews on the product page or reviews page. This preliminary study explored the review attribute presentations for three of the highest-grossing online retail categories (apparel, small electronics, and consumer appliances; National Retail Federation 2010) from over 300 of the largest (based on Alexa.com rankings) retailers, providing a conservative

Jared Watson is Assistant Professor of Marketing, Leonard N. Stern School of Business, New York University (email: jwatson@stern.nyu.edu). Anastasiya Pocheptsova Ghosh is Assistant Professor of Marketing, Eller College of Management, University of Arizona (email: anastasiya@email.arizona.edu). Michael Trusov is Associate Professor of Marketing, Robert H. Smith School of Business, University of Maryland (email: mtrusov@rhsmith.umd.edu).

our understanding of the relative influence of average product ratings and number of reviews by specifying the conditions by which the interactive effect of these two attributes on consumer preference takes place.

Some previous research has also explored the interactive relationship between various aspects of aggregate review information. Chintagunta, Gopinath, and Venkataraman (2010) investigate the impact of aggregate movie reviews on box office sales (considering only average product ratings but not the review numbers of competing choice options) and find no interactive effects of average product ratings and number of reviews. Chen, Dhanasobhon, and Smith (2008) investigate the impact of reviews on book sales and find that individual reviews that have received high proportions of "helpful" votes by other consumers are more impactful on sales relative to other reviews, and that this effect is stronger for less popular (vs. more popular) books. Finally, Khare, Labrecque, and Asare (2011) find that ratings dispersion (i.e., the distribution of individual ratings) has a differential impact on sales of negatively and positively rated products, but only when the products have a large number of reviews. A wide ratings dispersion increases evaluations of negatively rated products, whereas a narrow dispersion increases evaluations of positively rated products. Thus, previous literature provides some indication that people might evaluate average product ratings differently when choice options have a different number of reviews (and vice versa), yet the precise nature of when such an interactive effect occurs has not been explored.

To answer this question, we examine consumer choice between multiple products that vary on average product ratings and number of reviews but whose other attributes (e.g., price points, brand recognition, functionality) are similar. This enables us to investigate the relative diagnosticity of these two attributes in consumer decisions and the conditions by which these relative diagnosticities change. Furthermore, whereas prior research has largely focused on preference of individual choice options (for a notable exception, see Chintagunta, Gopinath, and Venkataraman 2010), in this research, we examine the influence of average product ratings and number of reviews on preferences and choices within a choice set.

Increasingly, retailers provide consumers with product options within a choice set rather than individual options (e.g., product search pages, "recommended for you" lists; for examples, see Web Appendix W1); thus, consumers encounter aggregate review information for multiple choice options simultaneously. Within these choice sets, we specifically examine the conditions by which consumers prefer higher-rated choice options with fewer reviews over lower-rated choice options with more reviews as a function of different levels of average product ratings and review numbers.

For example, consider the following scenario. You are searching online for a new blender and see two comparable choice options that meet your specifications. While one choice option has a higher rating but fewer reviews (e.g., 3.5 out of 5.0 based on 8 reviews), the other choice option has a lower rating but more reviews (e.g., 3.2 out of 5.0 based on 64 reviews). What is the relative diagnosticity of average product ratings and number of reviews as a signal of product quality? Would these diagnosticities—and ultimately, your decision—change if the review numbers were instead 408 and 464, respectively? What if the average product ratings were 4.5 and 4.2, respectively?

Managerially, this is an important investigation, as the choice sets in which consumers make trade-offs between product ratings and number of reviews are commonplace. For example, imagine two products released in January and June, respectively. The product released in January has had six additional months to accrue reviews, resulting in a higher number of reviews but technology that is six months older, thus potentially yielding a lower-quality product relative to the product released in June. Alternatively, consider a low-quality brand that ran a brief steep price promotion. Because of the promotion, this low-quality option could have more reviews than its high-quality competitors, which did not engage in a major price promotion. As such, it is quite likely that consumers are faced with a choice set, as described previously.

To demonstrate the prevalence of this trade-off scenario, we analyzed over 2.5 million products across 24 product categories and their corresponding choice sets using data collected from Amazon (McAuley, Pandey, and Leskovec 2015). On average, 79% of the product choice sets in our sample featured at least one other product that was superior in one of the review attributes but inferior in the other (for details of the data set and our analyses, see Web Appendix W2). Thus, we argue that studying these types of consumer choices not only is interesting from a theoretical standpoint but also has direct practical relevance, as these are the decisions consumers face on a regular basis.

In the following sections, we develop our conceptual framework and the hypotheses to test it. We then test our hypotheses in seven studies before concluding with the managerial and theoretical implications of our findings and directions for future research.

Conceptual Background

Diagnosticity of Attributes as Signals of Product Quality

Numerous studies have examined how consumers infer product quality from multiple product attributes when making choices (Kirmani and Rao 2000; Rao and Monroe 1988, 1989; Richardson, Dick, and Jain 1994; Slovic 1966; Slovic and Lichtenstein 1971). Slovic and Lichtenstein (1971) propose the concept of attribute diagnosticity in the utilization of attributes in making choices. They argue that the perceived diagnosticity of any

estimate of review display, as smaller retailers are less likely to have the functionality for review acquisition. For the retailers in our sample that chose to display review information, the average number of reviews was 390 (SD = 1,241), and the median was 15 reviews for their most popular products. For comparison, the average number of reviews of all products on Amazon is 88 (SD = 64), and the median is 2.

attribute is a function of the degree to which it separates the available choice options on perceived quality. More diagnostic attributes take precedence over less diagnostic ones as inputs into judgments. The accessibility–diagnosticity framework further argues that the interpretation of attributes for choice and judgments is context-dependent rather than fixed (Feldman and Lynch 1988; Lynch, Marmorstein, and Weigold 1988). It suggests that the same attribute can have a different perceived diagnosticity depending on the context. For example, Lynch, Marmorstein, and Weigold (1988) demonstrate that when attributes are easily recalled, they are used as inputs for choice, but when the same attributes are difficult to recall, consumers rely on overall evaluations of the choice rather than individual attributes. Thus, it is not the inherent diagnosticity of an attribute that dictates its use in decisions but the perceived diagnosticity of that attribute at the moment of choice.

Importantly, such perceived diagnosticity of one attribute can be a function of the valence of the other attributes. For example, Purohit and Srivastava (2001) demonstrate that manufacturer reputation is considered a highly diagnostic cue, whereas warranty is not. As a result, whether warranty is used in product evaluations depends on the valence of reputation: when reputation is positive, a longer warranty improves quality judgments; when reputation is negative, the length of warranty does not affect judgments. Thus, the level of one attribute affects the perceived diagnosticity of the other, leading to joint effects of these attributes on consumer preferences.

Related work on attribute evaluability has further demonstrated that the levels of the same attribute of competing choice options can similarly change the perceived diagnosticity of the attribute and affect preferences (González-Vallejo and Moran 2001; Hsee 1996, 2000; Hsee et al. 1999; Hsee and Zhang 2010). Hsee (1996) asks participants to evaluate two dictionaries: one dictionary has 10,000 entries and is in perfect condition, while the other has 20,000 entries and a torn cover. When evaluated independently, the former dictionary is preferred, but when evaluated jointly, preference for the latter dictionary increases. In other words, when evaluated independently, the condition of the dictionary was more diagnostic, but when evaluated jointly, the number of entries was more diagnostic.

Taken together, this literature provides evidence that the diagnosticity of attributes, as signals of product quality and their influence on preference, is not fixed. Rather, consumers often evaluate attributes depending on the values and availability of other attributes in the choice set. Next, we apply these frameworks to consumers' use of average product ratings and the number of reviews in decision making.

Diagnosticity of Product Review Attributes

Consumer reviews on most sites include aggregate review information (average product ratings and number of reviews) and individual reviews that can include individual product ratings and (often) the textual content of each review. A large body of work has focused on examining the influence of aggregate review information and has demonstrated considerable impact of average product ratings and number of reviews on consumers' online behavior (Chevalier and Mayzlin 2006; Chintagunta, Gopinath, and Venkataraman 2010; Duan, Gu, and Whinston 2008; Liu 2006). Yet, as previously mentioned, the relative effects of these attributes are still debated (Floyd et al. 2014; You, Vadakkepatt, and Joshi 2015). This ambiguity has encouraged recent academic interest in how disaggregated review information, such as the sentiment of individual reviews, can improve our understanding of online consumer behavior (Ludwig et al. 2013; Villarroel Ordenes et al. 2017). This research demonstrates, for instance, that affective content of individual reviews influences conversion rates above and beyond the effects of aggregate review information; in addition, it demonstrates that aggregate review information (e.g., changes in the number of reviews) still has an impact on conversion rates (Ludwig et al. 2013).

This article complements these lines of research by aiming to clarify when and how review attributes jointly affect consumers' decisions. We argue that the relative diagnosticity of the average product ratings and the number of reviews in consumer choices is not fixed and may depend on the value of each attribute. We focus on the diagnosticity of the number of reviews and average product ratings relative to each other (and not relative to other attributes, such as textual content of reviews) for several reasons. First, the accessibility–diagnosticity framework posits that attributes that are more easily accessible to consumers are more likely to be inputs into judgments (Feldman and Lynch 1988). When evaluating products online, consumers are likely to first examine the most accessible information. Individual reviews are less accessible than aggregate review information, as the former are often placed at the bottom of a product's page or on a separate page from other product attributes. Second, attributes that are easier to evaluate because they are comparable with each other are more likely to be inputs into judgment (Hsee 1996; Markman and Medin 1995; Zhang and Markman 1998).

Prior work has demonstrated that comparisons between different levels of alignable attributes (which would be the case of two numerical attributes, such as average product ratings and number of reviews) are easier to make than comparisons of nonalignable attributes (Cho, Khan, and Dhar 2013; Nowlis and Simonson 1997; Slovic and MacPhillamy 1974; Yeung and Soman 2005; Zhang and Markman 2001). Relatedly, comparisons of quantitative attributes are easier to make than comparisons of qualitative ones (Nowlis, Dhar, and Simonson 2010), suggesting that aggregate review information expressed in a numerical way (average product ratings and number of reviews) is more likely to influence consumer judgments.

Relative Diagnosticity of Average Product Ratings and the Number of Reviews

We propose that, for consumers, average product ratings act as more evaluable signals of quality than the number of reviews.

This happens because average product ratings are generally bound by a scale with two endpoints (e.g., 1–5 stars), so consumers can easily compare the average product rating with the scale endpoints to infer the level of absolute quality. By contrast, the number of reviews is presented on an unbound scale; furthermore, while the minimum number of reviews a product can possess is zero, the potential maximum is infinite. We argue that having the number of reviews unbound, at least on one end of the scale, can make the absolute number of the reviews more difficult to interpret as a signal of quality and can lower the perceived diagnosticity of this attribute in consumers' judgments. Consistent with our proposition, recent work has demonstrated that consumers believe that average product rating is the strongest indicator of a product's objective quality, more so than other quality cues, such as product price (De Langhe, Fernbach, and Lichtenstein 2015).

We build this argument drawing on numerical cognition literature, which has investigated how changing the features of numeric scales of a single attribute influences product evaluations (Bagchi and Li 2010; Monga and Bagchi 2011; Pandelaere, Briers, and Lembregts 2011; Schley, Lembregts, and Peters 2017). This literature has found that numbers and calculations that are easier to process positively improve brand evaluations and product promotions (King and Janiszewski 2011). This finding suggests that attribute values presented on easier-to-evaluate bound scales would lead to greater influence on consumer judgments (although this has not been tested directly in prior work). In one recent study consistent with this view, Chandon and Ordabayeva (2017) find that consumers make better (i.e., more accurate) judgments when they are estimating decreasing quantities of food (on a scale bound by two endpoints: zero and a maximum possible quantity) as opposed to increasing quantities of food (on a scale bound by only one endpoint, minimum possible quantity, and unbound by infinity on the other). Prior work on evaluability discussed previously (Hsee 1996) can also be viewed through the lens of bound and unbound scales. The number of words in a dictionary is more difficult to evaluate because it does not have a well-defined endpoint, whereas the dictionary's condition has implicit endpoints (e.g., "perfect" vs. "completely destroyed"). Thus, the findings that the dictionary's condition is easier to evaluate and does not require an additional anchor (in the form of joint evaluation), as compared with the number of words in the dictionary, supports our view that attributes expressed on bound scales are easier to evaluate and would be judged as more diagnostic than attributes expressed on unbound scales.

Building on these literature streams, we propose that the diagnosticity of average product ratings as a signal of product quality is higher than that of the number of reviews. Furthermore, similar to prior work on attribute diagnosticity (Purohit and Srivastava 2001), we argue that whether the number of reviews is incorporated into decisions depends on the valence of a more diagnostic cue (i.e., average product ratings). When making a buying decision, consumers are motivated to avoid choosing an inferior option to avoid postdecisional regret (Tsiros and Mittal 2000; Zeelenberg and Pieters 2007). This anticipated regret from making an incorrect (i.e., suboptimal) decision often drives consumers to engage in a more comprehensive assessment of the choice options (Bockenholt et al. 1991). Similarly, research on the negativity bias (Baumeister et al. 2001; Ito et al. 1998; Rozin and Royzman 2001) suggests that consumers attend to and elaborate more on information about a judgment in the presence of negative information. Thus, we argue that the diagnosticity of the number of reviews is likely to increase when average product ratings contain negative information (e.g., negative or neutral average product ratings) relative to when they contain positive information (e.g., positive average product ratings).

How might consumers use the number of reviews in their judgments under these conditions? We believe that the difference in the number of reviews between two choice options would appear more diagnostic when the level of review numbers in a choice set is low relative to high. Holding the absolute difference constant in the number of reviews between choice options, the difference between the number of reviews would appear relatively larger in choice sets with fewer (vs. more) reviews. Prior work has demonstrated that people attend more to relative differences than absolute differences in values (Thaler 1980; Tversky and Kahneman 1985). For example, people are willing to exert more effort to save $5 on a $15 purchase than on a $125 purchase. This happens because of a steeper slope for smaller values and a shallower slope for larger values of the utility function, as argued by the prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1985). Thus, the same absolute difference in numbers (e.g., 10) would loom larger when the numbers being compared are low (e.g., 20 vs. 30) versus high (e.g., 200 vs. 210). Applying this sensitivity in relative differences to the context of the number of reviews, we would expect that, holding the absolute difference in the number of reviews between choice options constant, having a choice set with a low (vs. high) level of reviews would increase diagnosticity of the review numbers attribute.

The proposed change in perceived diagnosticity of the number of reviews (relative to average product ratings) has direct implications for consumer preferences. In a choice set in which one option has a higher rating but fewer reviews and another option has a lower rating but more reviews, an increase in the perceived diagnosticity of the number of reviews would lead to a weaker preference for the higher-rated option. This occurs because when the level of review numbers is low (vs. high), the diagnosticity of the number of reviews increases, leading to joint influence of average product ratings and the number of reviews. As such, when the higher-rated choice option has fewer reviews, it is perceived as superior on one quality attribute and inferior on another quality attribute, thus weakening preference for it. However, this effect of a low level of review numbers on preferences would be attenuated when the average product ratings of both choice options are high, as most consumers will engage in less elaborative decision-making and will be less likely to incorporate the number of reviews into their decision.

Formally, we propose that, given a choice set in which consumers face a trade-off between average product ratings and the number of reviews,

H_1: Preference for a higher-rated choice option with fewer reviews is weaker when the level of review numbers is low (vs. high).
H_2: The influence of the number of reviews on preference for a higher-rated choice option with fewer reviews is attenuated when average product rating is positive (vs. neutral or negative).
H_3: Attribute diagnosticity of the number of reviews (relative to average product ratings) is the highest when the level of review numbers is low (vs. high), and this increase in diagnosticity influences preference between choice options.

Overview of Studies

We test our predictions in a series of seven studies. Study 1 demonstrates the systematic shift in preference between choice options as a function of the level of review numbers, providing support for H_1. Study 2 generalizes this finding by using an expanded choice set while also demonstrating the impact of the number of reviews on choice and choice deferral. Study 3 then tests H_2 by examining the interaction of the level of the average product ratings' valence and the level of review numbers on preference between options. Study 4 demonstrates that a large ratings difference between choice options only increases diagnosticity of the average product ratings when the level of review numbers is high versus low. Study 5 investigates how the diagnosticity of average product ratings is increased when one product rating rests at the scale boundary (e.g., 1.0, 5.0). Finally, Studies 6 and 7 test H_3's assertion that the difference in the diagnosticity of average product ratings and number of reviews drives the effect of the level of review numbers on preference, using an evaluative measure of attribute diagnosticity (self-reported attribute weights in Study 6) and an attention measure of diagnosticity (eye tracking in Study 7).

Studies' Paradigm

In every study, participants were asked to imagine that they were considering the purchase of a new product and had narrowed their choice set to two comparable choice options (four choice options in Study 2). Participants then saw both choice options side by side, with information about the brand name, price, average product rating, and the number of reviews for each choice option presented beneath the product images (for an example, see Web Appendix W3). In each choice set (except for Study 2), choice option A always had a higher average product rating with fewer reviews, and choice option B had a lower average product rating with more reviews. Other product attributes were not significantly different.

Specific values of average product ratings and the numbers of reviews varied between studies and between choice sets within the studies to extend the generalizability of our results (see Table 1). We chose the numbers of reviews by selecting values at the lower and upper limits of the perceived average number of reviews based on a pretest (N = 182) in which we had participants classify various numbers of reviews along a continuum from 1 = "Far fewer than average," and 7 = "Far more than average." In general, consumers found 74 reviews to be about average, independent of any other information (e.g., product category, website).

After viewing each choice set, participants were asked to indicate their relative preference between choice options on a seven-point scale (1 = "Strongly prefer option A," and 7 = "Strongly prefer option B"; except for Studies 2 and 6, in which we used choice as a dependent measure). This measure anchored preference for the higher-rated choice option with fewer reviews at 1 and the lower-rated choice option with more reviews at 7. As such, a higher number on this measure would indicate a weaker preference for the higher-rated choice option with fewer reviews. For a summary of results across studies, see Table 2.

Study 1: Effect of the Level of Review Numbers on Preference Between Choice Options

The purpose of this study was to test the systematic shift in preference as a function of the level of review numbers outlined in H_1. We argue that preference for the higher-rated options with fewer reviews is weaker when the level of review numbers in a choice set is low relative to high or absent. To test this assertion, we examine four increasing levels of review numbers while keeping average product ratings constant across conditions. We also include a fifth condition in which the number of reviews is absent from the information provided to participants. This provides a condition in which preference is based only on average product ratings; furthermore, by comparing this condition with the others, we can conduct an initial test of the dynamic diagnosticity of the review attributes (H_3). Finally, to demonstrate the robustness of this effect, we replicate it across five product categories in which the brands, prices, average product ratings, and numbers of reviews all slightly vary for each product to avoid any demand effects from specific values.

Participants and Design

We recruited 250 participants (M_{age} = 31.35 years; 31% female) from Amazon Mechanical Turk (MTurk) in exchange for $.50 and randomly assigned them to one of five levels of review numbers: low (8 vs. 64 reviews), moderate (72 vs. 128 reviews), moderately high (201 vs. 257 reviews), high (456 vs. 512 reviews), or control (i.e., the review numbers were absent), in a between-subjects design. Within-subject, each participant viewed five product choice sets.

Table 1. Design and Measures Summary.

		Average Product Ratings					Review Numbers					Reported Measures			
Product	Option	(1.x)	(2.x)	(3.x)	(4.x)	(5.x)	Low	Moderate	Moderately High	High	Absent	Relative Preference	Absolute Choice	Choice Deferral	Process Measures
Study 1															
Over-the-ear headphones	A	—	—	.5	—	—	8	72	201	456	N.A.	x	—	—	—
	B	—	—	.3	—	—	64	128	257	512	N.A.				
Coffee makers	A	—	—	.4	—	—	6	64	180	412	N.A.	x	—	—	—
	B	—	—	.0	—	—	58	116	232	464	N.A.				
Microwaves	A	—	—	.5	—	—	9	71	195	443	N.A.	x	—	—	—
	B	—	—	.2	—	—	62	124	248	496	N.A.				
Speaker systems	A	—	—	.4	—	—	12	86	234	530	N.A.	x	—	—	—
	B	—	—	.1	—	—	74	148	296	592	N.A.				
Lounge chairs	A	—	—	.7	—	—	5	103	299	691	N.A.	x	—	—	—
	B	—	—	.4	—	—	98	196	392	784	N.A.				
Study 2															
Camping lamps	A	—	—	.2	—	—	61	—	—	361	N.A.	—	x	x	—
	B	—	—	.6	—	—	22	—	—	322	N.A.				
	C	—	—	.8	—	—	5	—	—	305	N.A.				
	D	—	—	.4	—	—	43	—	—	343	N.A.				
Study 3															
Blenders	A	—	.4	.4	.4	—	8	—	—	408	—	x	—	—	—
	B	—	.1	.1	.1	—	64	—	—	464	—				
Study 4															
Earbud headphones	A	—	—	.8, .6, .4, .2	.8, .6, .4, .2	—	9	—	—	409	—	x	—	—	—
	B	—	—	.6, .4, .2, .0	.6, .4, .2, .0	—	57	—	—	457	—				
Study 5															
Hand mixers	A	.3 (.4)	—	—	(.9)	.0	13	—	—	313	—	x	—	—	—
	B	.0 (.1)	—	—	.7, (.6)	—	34	—	—	334	—				
Study 6															
Blenders	A	—	—	.4	—	—	8	—	—	408	—	—	x	x	x
	B	—	—	.1	—	—	64	—	—	464	—				
Study 7															
Microwaves	A	—	—	.5	—	—	9	—	—	443	—	x	—	—	x
	B	—	—	.2	—	—	62	—	—	496	—				

Notes: For Study 4, all pairwise comparisons were used for average product ratings where A was greater than B, resulting in ten comparisons. For Study 5, numbers in parentheses were paired together in their respective negative/positive conditions.

Table 2. Means Summary Across Studies.

	Average Product Rating		Level of Review Numbers				
	Valence	Difference	Low	Moderately Low	Moderately High	High	Absent
Study 1 (n = 250)	Neutral	—	3.96	2.93	2.83	2.63	2.81
Study 2 (n = 144)	Neutral	—	49	—	—	78	76
Study 3 (n = 433)	Negative	—	4.15	—	—	3.32	—
	Neutral	—	4.38	—	—	3.38	—
	Positive	—	3.04	—	—	2.97	—
Study 4 (n = 705)	Neutral	—	3.98	—	—	2.77	—
	Positive	—	3.32	—	—	2.66	—
	—	Small	3.62	—	—	2.98	—
	—	Large	3.68	—	—	2.45	—
Study 5 (n = 410)	Negative	—	3.58	—	—	3.67	—
	Negative (extreme)	—	3.84	—	—	3.44	—
	Positive	—	4.08	—	—	2.71	—
	Positive (extreme)	—	3.48	—	—	2.96	—
Study 6 (n = 143)	Neutral	—	36	—	—	61	61
Study 7 (n = 92)	Neutral	—	4.09	—	—	3.68	—

Notes: Studies 1, 3, 4, 5, and 7 use a relative preference measure (1 = "Strongly prefer higher-rated, fewer reviews option," and 7 = "Strongly prefer lower-rated, more reviews option"). Studies 2 and 6 use an absolute choice measure in which the number reported is participants' percentage choosing the higher-rated option with fewer reviews.

Choice Sets

For each of the five choice sets (headphones, microwaves, coffee makers, speaker systems, and lounge chairs), participants viewed two products that were nearly identical, with the exception of their average product ratings and number of reviews (see Table 1).

Measure

After viewing each choice set, participants were instructed to indicate their preference between the options on a seven-point scale (1 = "Strongly prefer option A" [the higher-rated option with fewer reviews], and 7 = "Strongly prefer option B" [the lower-rated option with more reviews]). Thus, a higher score on this scale indicated weaker preference for the higher-rated option with fewer reviews.

Results

A 5 (level of review numbers: low, moderate, moderately high, high, control) × 5 (product category: headphones, microwaves, coffee makers, speaker systems, lounge chairs) repeated-measures analysis of variance (ANOVA) on preference yielded significant main effects of the level of review numbers ($F_{(4, 246)} = 11.45, p < .001$) and product category[3] ($F_{(1, 246)} = 18.54, p < .001$). The interaction was not significant ($p > .10$). Because of this, we collapsed across the product category factor to simplify the reporting of results (see Figure 1), though the same directional pattern held for all products. In support of H_1, planned contrasts demonstrated that preference for the choice option with a higher rating and fewer reviews was significantly weaker in the low review numbers condition ($M_{low} = 3.96$) compared with all other conditions ($M_{moderate} = 2.93$; $t_{(246)} = -4.74$, $p < .001$; $M_{m\text{-}high} = 2.83$; $t_{(246)} = -5.12$, $p < .001$; $M_{high} = 2.63$; $t_{(246)} = -6.15$, $p < .001$; $M_{control} = 2.81$; $t_{(246)} = -5.27$, $p < .001$). Importantly, not displaying the number of reviews led to no significant difference in preferences relative to when the level of review numbers was high ($M_{high} = 2.63$, $M_{control} = 2.81$; $p > .10$). This is consistent with our prediction that the number of reviews is less diagnostic and, therefore, less likely to influence preferences relative to average product ratings when the level of review numbers is high (H_3).

Discussion

Study 1 demonstrated that consumers' preference for the option with higher ratings and fewer reviews is weaker when the level of review numbers is low versus high or absent, consistent with H_1. We replicated the effect in a follow-up consequential

[3] We conducted paired sample t-tests to contrast the product categories. Regardless of the level of review numbers, preference for the higher-rated choice option with fewer reviews was marginally weaker for headphones (M = 3.25) than that for coffee makers (M = 2.96; $t_{(250)} = 1.93$, $p = .055$) and speaker systems (M = 3.00; $t_{(250)} = 1.71$, $p = .088$) and significantly weaker than for chairs (M = 2.56; $t_{(250)} = 4.93$, $p < .001$). Preference for the higher-rated choice option with fewer reviews in the coffee makers choice set was significantly stronger than that for microwaves (M = 3.25; $t_{(250)} = -2.25$, $p = .025$) and chairs ($t_{(250)} = 3.16$, $p = .002$). Preference for the higher-rated choice option with fewer reviews in the microwaves choice set was significantly stronger than that for speaker systems ($t_{(250)} = 2.00$, $p = .047$) and chairs ($t_{(250)} = 5.11$, $p < .001$). Preference for the higher-rated choice option with fewer reviews in the speaker systems choice set was significantly stronger than that for chairs ($t_{(250)} = 3.65$, $p < .001$). Although there were minor categorical differences between products, we urge the reader to not read too far into this as it could merely be a function of the stimuli. The important takeaway is that the same systematic shift in preference occurred across the various product categories.

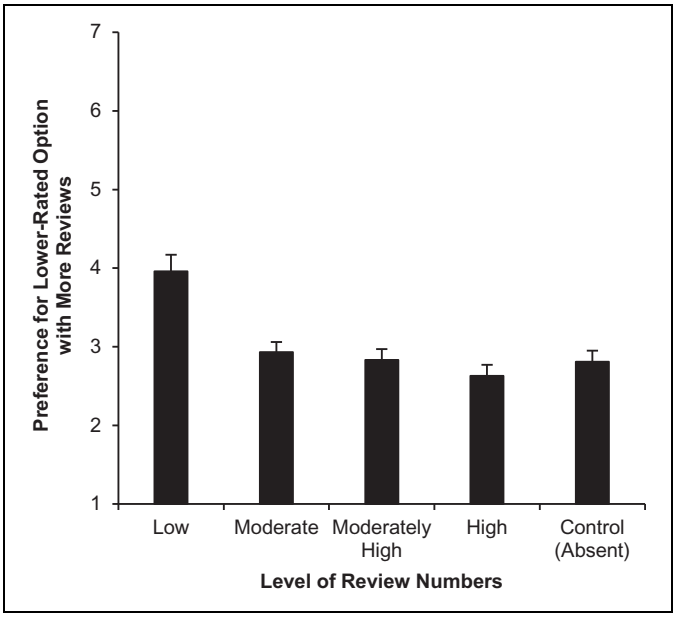

Figure 1. Study 1 results: Preference across five levels of review numbers.

choice study, in which participants were entered into a raffle to receive their preferred product option (a blender). Using the low and high levels of review numbers from the main study, we observed the same shift in preference away from the higher-rated option with fewer reviews as the level of review numbers decreased ($M_{low} = 4.68$, $M_{high} = 3.08$; $F_{(1,104)} = 19.10$, $p < .001$), which is consistent with H_1.

To test the robustness of this effect, we also examined several potential boundary conditions to the effect of number of reviews on consumer preference (full descriptions of each study are available in the Web Appendix). A large body of work on numerosity (Burson, Larrick, and Lynch 2009; Lembregts and Pandelaere 2013; Pandelaere, Briers, and Lembregts 2011) makes a different prediction from the one we demonstrate in Study 1. This line of research shows that, in choice options, keeping relative differences between attribute values constant while expanding the scale in which the values are presented (e.g., a warranty described in months vs. years)—thus changing the *absolute difference*—causes differences between attribute values to appear exaggerated on an expanded scale (e.g., warranty described in months). By contrast, in Study 1 we keep the absolute difference between choice options constant while changing the *relative difference* between attribute values by changing the scale. We believe that both effects might exist in the context of product reviews, but it is unclear a priori which has a stronger effect on consumer judgment because prior work on numerosity has not studied its effects in the context of multiple numerical attributes, as we do. Indeed, in an additional study (Web Appendix W4), we compare the influence of an increase in *relative* difference between choice options on the review numbers attribute (as in our main studies) with an increase in an *absolute* difference between choice options on the review numbers attribute (as demonstrated by prior work). This study provides initial evidence that both effects exist at varying degrees of strength.

In another study (Web Appendix W5), we examine whether aggregate review information would be equally diagnostic in all product categories. Specifically, we examine products with primarily aesthetic or taste value. Prior literature has suggested that consumer responses to reviews differ as a function of the consumer's self-expression goals and whether a product is being evaluated on the basis of taste versus quality (He and Bond 2015; Rozenkrants, Wheeler, and Shiv 2017). Consistent with this literature, in this study, we demonstrate that aggregate review attributes are less diagnostic when consumers make preferences on the basis of taste or self-expression concerns (e.g., artwork).

Finally, in another study (Web Appendix W6), we examine the role of single- versus joint-option evaluations (i.e., viewing option A and then viewing option B vs. viewing both options simultaneously). Prior literature investigating this aspect of choice has demonstrated that joint evaluation increases the diagnosticity of difficult-to-evaluate attributes (Hsee 1996) and attenuates the effects of numerosity (Schley, Lembregts, and Peters 2017). In our research, this would suggest that diagnosticity of review numbers (a more-difficult-to evaluate attribute) could be further attenuated under single-evaluation conditions. We test this proposition in Web Appendix W6 and find that our effect holds for both joint and single evaluations of choice options. Taken together, these studies demonstrate robustness of the effect of number of reviews on preferences.

Study 2: Replication Using an Expanded Choice Set

Study 2 is designed to demonstrate that the effect of the level of review numbers on preferences is robust for larger choice sets. Importantly, it also examines whether the effect of the level of review numbers has an impact on discrete option choice and consumer likelihood to defer choice. Prior research has shown that large choice sets increase the use of noncompensatory decision strategies (Johnson and Meyer 1984; Payne 1976), such that consumers are more likely to choose options that are superior on one of the most important or easiest-to-differentiate attributes rather than incorporating multiple attributes. In the context of this research, we argue that when the level of review numbers is low (vs. high), the diagnosticities of average product ratings and number of reviews are more similar. As such, although consumers are more likely to use a noncompensatory strategy in the multiple–choice option context, the likelihood of consumers using the number of reviews as a diagnostic cue increases. When the level of review numbers is high, most consumers will use average product ratings as their primary diagnostic cue, leading them to choose the highest-rated choice option. In contrast, when the level of review numbers is low, both attributes appear diagnostic, and consumers are less likely to use average product ratings as their primary diagnostic cue, leading to weaker preference for the highest-rated choice

option. Thus, we expect a similar pattern of preferences to emerge as we observed in Study 1, in which participants will be less likely to choose the highest-rated option with the fewest reviews when the level of review numbers is low relative to high or absent, even when choosing from a choice set with several options.

In addition, this study tests the effect of the level of review numbers on choice deferral, to provide additional support for H_3. We argue that the relative diagnosticity of average product ratings and the review numbers is more similar when the level of review numbers is low (vs. high or absent). The need to make trade-offs between attributes of similar importance increases choice difficulty (Chatterjee and Heath 1996; Dhar and Simonson 2003), which makes choice deferral more likely (Tversky and Shafir 1992; Dhar and Nowlis 1999; Etkin and Ghosh 2018). Thus, we expect the rate of choice deferral to be the highest when the level of review numbers is low, where the trade-off between the average product ratings and the review numbers are more salient compared with the conditions where level of review numbers is high or when the review numbers are absent.

Participants and Design

We randomly assigned 144 undergraduate students (M_{age} = 20.91 years; 50% female) to one of three levels of review numbers (low, high, control), in a between-subjects design. In exchange for participating in the study, the students received course credit.

Choice Set

Participants viewed a choice set of four camping lamps. The options were nearly identical except for their average product ratings and number of reviews. While one choice option had the highest rating with the fewest reviews (3.8, 5 reviews), another choice option had the lowest rating with the most reviews (3.2, 61 reviews), and two choice options in the middle were compromise choice options that were neither the highest nor lowest on either attribute but were superior on one relative to the other compromise choice option (3.4, 43 reviews vs. 3.6, 22 reviews) (see Table 1). We manipulated the level of review numbers by either withholding the number of reviews in the control condition or adding 300 reviews to the numbers reported previously to generate the high level of review numbers.

Measures. To capture preference among the four choice options, we used a discrete choice measure of the highest-rated choice option with the fewest reviews rather than the relative preference measure used in the previous study. Next, to assess the likelihood of choice deferral, we asked participants, "Are you more likely to purchase one of the available options or defer purchase and look elsewhere?" and analyzed this as a binary measure. Finally, to assess the need for more information, participants were asked, "How would you classify the amount of information provided?" on a seven-point scale (1 = "not enough information," and 7 = "too much information"). A more difficult trade-off would require more information to help participants make a decision; thus, participants in the condition with a low level of review numbers would be expected to require more information relative to those in the other conditions.

Results

Choice of the highest-rated option with the fewest reviews. A binary logistic regression in which we dummy-coded the level of review numbers yielded an omnibus effect of the level of review numbers ($\chi^2(2) = 10.84$, $p = .004$). Consistent with H_1, when the level of review numbers was low, participants were significantly less likely to choose the highest-rated option with the fewest reviews ($P_{low} = 49\%$) relative to when the level of review numbers was high ($P_{high} = 78\%$; $\chi^2(1) = 7.09$, $p = .004$) or absent ($P_{control} = 76\%$; $\chi^2(1) = 8.24$, $p = .008$). There was no significant difference in choice in the high and control conditions ($p > .80$).

Rate of choice deferral. A binary logistic regression in which we dummy-coded the level of review numbers yielded an omnibus effect of the level of review numbers ($\chi^2(2) = 6.73$, $p = .035$). When the level of review numbers was high ($P_{high} = 53\%$; $\chi^2(1) = 4.30$, $p = .038$) or absent ($P_{control} = 49\%$; $\chi^2(1) = 6.02$, $p = .014$) participants were significantly less likely to defer choice relative to when the level of review numbers was low ($P_{low} = 73\%$). There was no significant difference between high and control conditions ($p > .70$). A higher rate of choice deferral under the low level of review numbers is consistent with prior work linking choice difficulty with choice deferral (Dhar and Nowlis 1999; Etkin and Ghosh 2018; Tversky and Shafir 1992). Consistent with our theorizing, when the level of review numbers is low (relative to high or absent), the diagnosticity of the number of reviews increases (H_3), creating a more difficult choice involving the trade-offs, ultimately increasing choice deferral.

Need for additional information. A one-way ANOVA of the level of review numbers on the need for additional information yielded a marginal effect of the level of review numbers of reviews ($F_{(2,144)} = 2.98$, $p = .054$). Planned contrasts further demonstrated that participants who encountered a high level of review numbers indicated that they had significantly more information than those who had encountered a low level of review numbers ($M_{high} = 3.37$, $M_{low} = 2.73$; $t_{(144)} = 2.43$, $p = .016$), which is consistent with our prediction. Participants who did not see the review numbers were not significantly different from those of the other groups ($M_{control} = 3.00$, $p > .15$). Although we expected the low level of review numbers to increase the need for additional information, we were surprised to find that when the level of review numbers was absent, participants felt no greater need for additional information than when the level of review numbers was high. This result

suggests that withholding the number of reviews from the list of attributes would not have a negative impact on consumers' perceptions of the amount of information they are provided during a decision.

Discussion. Study 2 provided additional evidence for the effect of the level of review numbers by demonstrating that a low level of review numbers shifts preference away from the higher-rated choice options relative to a high level of review numbers or no review numbers at all, which is consistent with H_1. We demonstrate this effect in the context of an expanded choice set while also using a discrete choice measure (rather than the relative preference measure used in Study 1). Furthermore, this study provided additional evidence in support of H_3 by demonstrating that the choice deferral rate is highest when the level of review numbers is low relative to high or absent. In addition, the need for additional information was greatest when the level of review numbers was low, suggesting that the trade-off between average product ratings and the number of reviews was most difficult because of the similarity in the diagnosticity of the two attributes in this condition. This is, again, consistent with H_3.

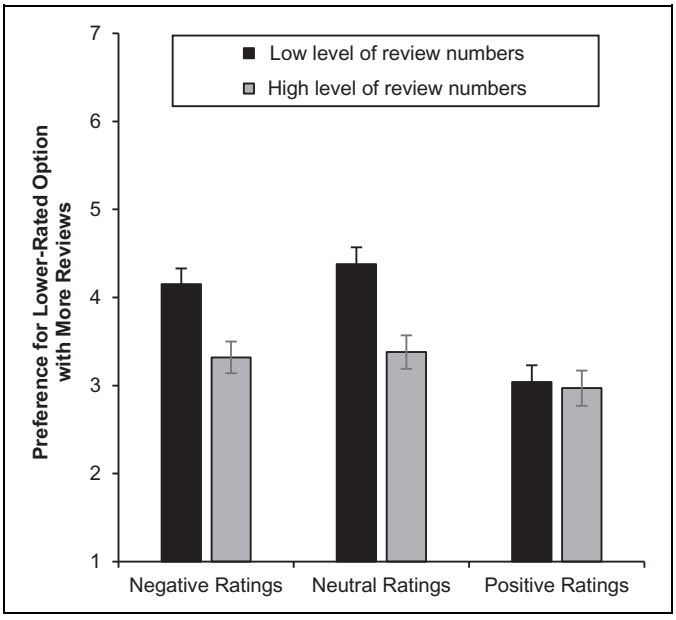

Figure 2. Study 3 results: Preference by average product ratings valence and the level of review numbers.

Study 3: Level of Review Numbers Effects at Various Valences of Average Product Ratings

Study 3's objective is to test H_2. Using relatively neutral average product ratings (e.g., 3.0–3.8), our initial studies demonstrated that consumers are more likely to use both average product ratings and the number of reviews in their choices when the level of review numbers is low. H_2 suggests that the effect of the level of review numbers on consumer preferences can be attenuated when consumers choose between positively rated products. We argue that this occurs because positively rated products can reduce consumers' need to elaborate on additional information, thereby increasing the diagnosticity of average product ratings relative to the number of reviews. To test this theory, in this study, we manipulate the valence of average product ratings while holding the ratings difference between options constant.

Participants and Design

We recruited 433 undergraduate students ($M_{age} = 20.28$ years; 46% female) to participate in this study in exchange for course credit. We then randomly assigned them to a condition in a 2 (level of review numbers: low, high) × 3 (ratings valence levels: negative, neutral, positive) between-subjects design.

Choice Set

Participants saw a choice set of two blenders. Choice options were nearly identical with the exception of their average product ratings and the number of reviews (see Table 1). In the low (high) level of review numbers condition, participants chose between 8 (408) and 64 (464) reviews, respectively. We manipulated ratings valence levels by changing the first digit of the average product ratings for both choice options. Thus, the negative condition presented consumers with 2.x choice options, the neutral condition presented 3.x choice options, and the positive condition presented 4.x choice options. After viewing the choice set, participants indicated relative preference on the same seven-point scale used in Study 1.

Results

A 2 (level of review numbers: low, high) × 3 (ratings valence levels: negative, neutral, positive) ANOVA on preference yielded main effects of the level of review numbers ($F_{(1, 427)} = 17.26$, $p < .001$) and the ratings valence level ($F_{(2, 427)} = 12.68$, $p < .001$), qualified by the predicted interaction ($F_{(2, 427)} = 3.58$, $p = .029$, see Figure 2). Replicating prior studies, in the neutral ratings valence level conditions, preference for the higher-rated option with fewer reviews was weaker when the level of review numbers was low versus high ($M_{low} = 4.38$, $M_{high} = 3.38$; $F_{(1, 427)} = 14.28$, $p < .001$). As we predicted in H_1, a similar effect was present in the negative ratings valence condition ($M_{low} = 4.15$, $M_{high} = 3.32$; $F_{(1, 427)} = 9.77$, $p = .002$). Furthermore, as we predicted in H_2, when the ratings valence level was positive, the effect of the level of review numbers on preference was attenuated ($M_{low} = 3.04$, $M_{high} = 2.97$; $F_{(1, 427)} = .07$, $p > .75$). As we expected, the effect of the level of review numbers on preference was attenuated when the more diagnostic cue (i.e., average product ratings) was positive.

Discussion

This study provided support for H_2 by demonstrating that the effect of the level of review numbers on preference between

choice options is attenuated when the choice set features only positive average product ratings. Furthermore, consistent with H_1, when consumers encountered non-positive product ratings, preference between choice options was affected by the level of review numbers, thus replicating the results of Studies 1 and 2.

Study 3 demonstrated the moderating role of average product ratings valences, but we also investigated several other potential moderators, reported fully in the Web Appendix. In one study (Web Appendix W7), we examine whether more disaggregated information (i.e., ratings distributions that display individual product ratings) moderates the effect of the level of review numbers on preference. Some literature investigating the role of a ratings distribution has demonstrated a significant effect of skew on product evaluations (Fisher et al. 2018; Khare, Labrecque, and Asare 2011). This work largely differs from ours in that it explores only single-option choice sets. In the context of multi-option choice sets, we argue that it is more difficult to compare and interpret different ratings distributions relative to the aggregate review attributes (i.e., average product ratings and review volumes). This is because the ratings distribution is a more complex-to-process attribute with several values, one for each rating. Thus, we argue and demonstrate that adding a ratings distribution to a multi-option choice sets does not change the effect of the level of review numbers on consumer preferences. We tested this effect using several different distributions (e.g., positively, negatively, and even-skewed), and our findings were robust.

In another study (Web Appendix W8), we examined whether a popularity cue can attenuate the effect of the level of review numbers on preferences. Prior research has suggested that higher review numbers can signal greater popularity, which in turn increases their trustworthiness (Chen, Wu, and Yoon 2004; Zhu and Zhang 2010). If the review numbers are perceived solely as a signal of product popularity, their effect should be attenuated in the presence of an alternative popularity cue. In this study, we test the effect of labeling one of the choice options as a "Best Seller" and demonstrate that this does not attenuate the effect of the level of review numbers.

Furthermore, in the study reported in Web Appendix W9, we tested credible reviews as another moderator. Some may argue that low review numbers contain more risk relative to high review numbers because of a higher likelihood of fraudulent reviews affecting average product ratings (Luca and Zervas 2016; Mayzlin, Dover, and Chevalier 2014). Thus, one might conclude that if consumers could be certain of the veracity of reviews for the higher-rated choice option with fewer reviews, the effect of the level of review numbers may be attenuated. We test this proposition by labeling the higher-rated option with fewer reviews as "Consumer Reports Verified" and demonstrate again that this label does not attenuate the effect of the level of review numbers.

Finally, in studies reported in Web Appendices W10 and W11, we tested whether the influence of the level of review numbers can be attenuated by providing consumers with a justification of why a product in a choice set might have fewer reviews (for reasons unrelated to product quality). In Web Appendix W10, we examine the role of production years of the goods. One reason that some products have more reviews than others is simply because they have been on the market longer. While this may be a positive signal for some products, for tech products, this signals outdated technology. However, we argue that the effect of review numbers is so strong, it leads consumers to discount the inferiority of older technology if the product has more reviews. We demonstrate that consumers are more apt to choose older products with more reviews (e.g., a 2013 DVD player, Galaxy S6) relative to newer products with fewer reviews (e.g., a 2015 DVD player, Galaxy S7) when the level of review numbers is low versus high. As such, it is quite likely that review numbers may bias consumers' decisions, leading to suboptimal outcomes.

Following the same logic as in the prior study, in Web Appendix W11, we examine the role of a "new arrival" label on the effect of the level of review numbers. Because new products have been on the market for a shorter period than other products, their fewer reviews should be justified. In theory, consumers should be more accepting of "new arrivals" with low review numbers relative to options that do not feature this label, as their low review number may not be justified. However, it is difficult to quantify the diagnostic value of a "new arrival" label and arrive at a proper discount rate for the review numbers relative to simply comparing the review numbers of competing products. As such, we theorize and demonstrate that the effect of the level of review numbers persists in the presence of a "new arrival" label. The studies presented in Web Appendices W7–W11 address several potential moderators identified in prior work, thus demonstrating that the interactive effect of the number of reviews and average product ratings does not appear to be affected by these factors.

Study 4: Ratings Valence Level and Level of Review Numbers Effects Across Small and Large Differences in Ratings

Study 4 accomplishes two goals. First, it provides an additional test of H_1 and H_2 by examining the effect of the level of review numbers at different valences of average product ratings. We expect to replicate the findings of Study 3, in which the effect of a low level of review numbers is attenuated when average product ratings are positive. Second, it investigates the role of the magnitude of difference between average product ratings in the joint influence of review numbers and ratings valence on preference.[4] Building on prior work that argues that the difference in attribute values between choice options can increase the salience of this attribute in decision making (Monroe 1971; Ross and Creyer 1992; Thaler 1980), we would expect that an increase in magnitude of difference between average product ratings of the options in the choice set would increase preference for the higher-rated product. However, consistent with H_3, we believe this effect will be more pronounced when

[4] We thank an anonymous reviewer for suggesting this study.

the diagnosticity of average product ratings is high. Thus, an increase in the difference between average product ratings in the choice set is more salient to consumers when both the level of review numbers and the diagnosticity of average product ratings are high than when the number of reviews is low and consumers place relatively less weight on the average product ratings. Examining the influence of the magnitude of a ratings difference between choice options at different levels of review numbers provides another test of H_3.

In our prior studies, we used relatively small differences between average product ratings (e.g., .2–.4); in this study, we expand the magnitude of difference to include .6 and .8. Specifically, we created choice sets with differences between choice options of magnitude of .2 (e.g., 3.8 vs. 3.6), .4 (e.g., 3.8 vs. 3.4), .6 (e.g., 3.8 vs. 3.2), and .8 (e.g., 3.8 vs. 3.0). We did so across two different valences of average product ratings: neutral and positive (e.g., 3.x or 4.x). This resulted in a necessarily unbalanced design because there are more differences of .2 within a level than .8; however, because there was no significant difference in effect within a given distance (for example, a difference of 4.2 and 4.4 led to similar preference as a choice set that had 4.4 and 4.6), we collapsed across those cells for the analysis. Thus, we collapsed across all small difference (.2–.4) and large difference (.6–.8) conditions to generalize our findings across relatively small and large differences in average product ratings.[5]

Participants and Design

We recruited 705 people from MTurk (M_{age} = 35.61 years; 47% female) to participate in the study in exchange for $.50. We randomly assigned them to a condition in a 2 (level of review numbers: low, high) × 2 (ratings valence levels: neutral, positive) × 2 (ratings difference size: small, large) between-subjects design.

Choice Set

Participants saw a choice set of two headphones. Choice options were nearly identical except for their average product ratings and number of reviews, as described previously (see Table 1). Relative preference was measured on the same seven-point scale as in Studies 1 and 3.

Results

A 2 (level of review numbers: low, high) × 2 (ratings valence levels: neutral, positive) × 2 (ratings difference size: small, large) ANOVA on preference yielded significant main effects of the level of review numbers ($F_{(1, 697)} = 40.66$, $p < .001$) and ratings valence levels ($F_{(1, 697)} = 6.70$, $p = .01$), qualified by the interaction of the level of review numbers and ratings difference size ($F_{(1, 697)} = 4.12$, $p = .043$) and a marginal interaction of the level of review numbers and ratings valence levels ($F_{(1, 697)} = 3.47$, $p = .063$). The main effect of the level of review numbers demonstrated that preference for the higher-rated option with fewer reviews was greater when the level of review numbers was high ($M_{low} = 3.63$, $M_{high} = 2.82$), which is consistent with prior studies. The main effect of valence demonstrated that preference for the higher-rated option with fewer reviews was greater when valence was positive relative to neutral ($M_{neutral} = 3.37$, $M_{positive} = 3.08$), consistent with Study 3.

In line with our expectations, the level of review numbers by ratings difference size interaction indicated that the effect of the ratings difference size on consumer preferences depended on the level of review numbers (see Figure 3, Panel A). When the level of review numbers was low, consumers were less sensitive to the ratings difference size between options ($M_{small} = 3.62$, $M_{large} = 3.68$; $p > .85$), suggesting that any ratings difference was considered a trade-off with the number of reviews. This is consistent with H_3, which proposes an increase in relative diagnosticity of the number of reviews when the level of the review numbers is low. By contrast, when the level of the review numbers was high, participants were more sensitive to the ratings difference size between options, thus increasing preference for the higher-rated choice option when the ratings difference size was large versus small ($M_{large} = 2.45$, $M_{small} = 2.98$; $F_{(1, 701)} = 6.72$, $p = .01$). This finding is also consistent with H_3, which proposes that the relative diagnosticity of average product ratings is greater when the level of review numbers is high (vs. low), thus increasing sensitivity to any difference in average product ratings.

The ratings valence levels by level of review numbers is consistent with the findings of Study 3 (see Figure 3, Panel B). Although a low level of review numbers decreased preference for the higher-rated choice option with fewer reviews, the magnitude of this effect was larger in the neutral condition relative to the positive valence condition, consistent with H_2 (neutral valence: $M_{low} = 3.98$, $M_{high} = 2.77$; $F_{(1, 701)} = 37.74$, $p < .001$; positive valence: $M_{low} = 3.32$, $M_{high} = 2.66$; $F_{(1, 701)} = 19.61$, $p = .012$). Note that the positive valence attenuated, rather than completely eliminated, the effect of the level of review numbers on preference (as we found in Study 3). While this nominal difference could be a result of the specific stimuli used for each study, both studies demonstrate an attenuation of the influence of the level of review numbers, which is consistent with our prediction, albeit to varying strengths.

Discussion

Study 4 replicates the findings in Study 3 by demonstrating that positive valences attenuate the influence of the level of review numbers on consumer decisions, thus providing additional support for H_2. Importantly, it also tests how the size of the difference in average product ratings between choice options changes consumer preference at different levels of review numbers. Consistent with the change in diagnosticity of attributes

[5] Additional results of extensive exploration of various levels of rating dispersion and the level of review numbers are available from the authors on request.

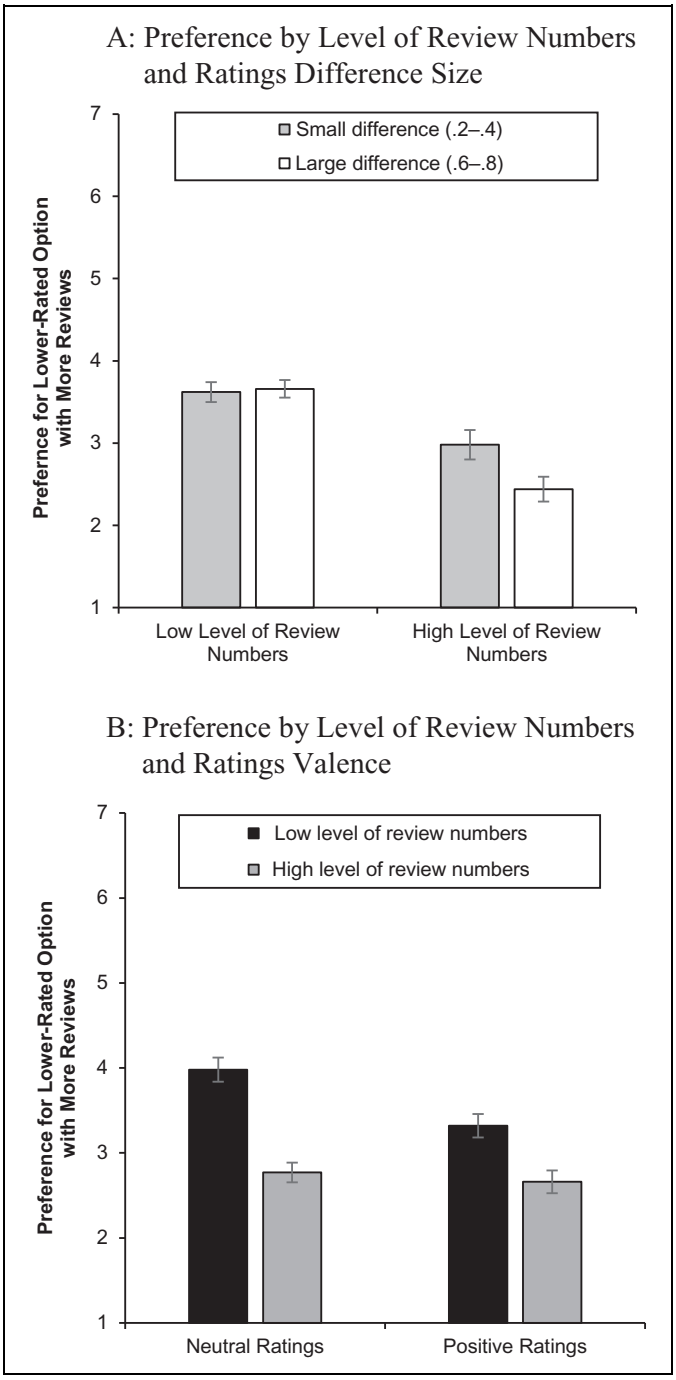

Figure 3. Study 4 results.

outlined in H_3, our results demonstrate that the ratings difference size is more important to consumers when the level of review numbers is high (vs. low). This happens because under these conditions, the diagnosticity of the number of reviews is low (relative to the diagnosticity of average product ratings). By contrast, consumers are less sensitive to the ratings difference size between choice options when the diagnosticity of the number of reviews increases (e.g., when the level of review numbers is low). This finding underscores the robustness of the effect of the level of review numbers on consumer choices.

Study 5: Level of Review Numbers Effects at Ratings Scale Boundaries

Study 5's objective is to provide additional tests of H_1 and H_2 by examining whether the influence of average product ratings' valence is enhanced at ratings scale boundaries. As we argued previously, ratings scales have defined boundaries (e.g., 1.0–5.0). Relative to unbound scales, bound scales make for easier comparisons of values, thus increasing the diagnosticity of average product ratings.

In this study, we investigate whether the ratings that exist at the endpoints of the scale increase the diagnostics of average product ratings relative to when ratings do not exist at the endpoints. We build this proposition on work by Isaac and Schindler (2013), which demonstrates that consumers often form mental boundaries of ranked lists around the numbers that end in zeroes (e.g., "top 10," "top 100"). Even if a list has more than 10 options, consumers will evaluate those within and outside of the top 10 differently. Because grouping options in this way affects the types of comparisons people make, as well as their final choices (Brenner, Rottenstreich, and Sood 1999), the 10th option is perceived significantly more differently than adjacent options. Isaac and Schindler (2013) demonstrate this effect in the context of student rankings. Imagine students in a classroom who are ranked on performance. The 11th-ranked student is perceived to be significantly worse than the 10th-ranked student; however, there is no difference in evaluation of the 11th- and 12th-ranked students. The "top 10" effect could be explained within the confines of prospect theory (Tversky and Kahneman 1992), in which losses are shown to loom larger (i.e., be more important) than gains. Thus, if consumers' point of reference is the top 10, 10th place would meet their standard, while 11th place would be considered a significant loss. However, if the consumers are evaluating 11th and 12th places, both fall below their reference point of the top 10, thus diminishing sensitivity to the loss on this attribute and leading to attenuation of the importance of their rankings.

Similarly, in the context of our research, we expect that when the rating of one product in a choice set lies at the boundary (i.e., has a rating of 1.0 or 5.0), the diagnosticity of average product ratings increases, thus attenuating the influence of the number of reviews. For example, in a choice set featuring ratings of 5.0 versus 4.7, consumers will be more influenced by the ratings rather than by the number of reviews relative to when they view a choice set featuring 4.9 versus 4.6.

While one might expect the same effect to occur at the negative boundary of the scale, the effect is likely to be smaller (i.e., consistent with the Prospect Theory), such that a loss relative to the reference point of 5.0 is more significant than an equivalent gain relative to the reference point of 1.0. Although consumers are unlikely to choose between options with extremely negative ratings, this study extends the examination of review volume effects to the extreme negative boundary of the ratings scale to provide complete analysis of all possible levels of product ratings.

Participants and Design

We recruited 410 people from MTurk (M_{age} = 37.88 years; 51% female) to participate in the study in exchange for $.50. They were randomly assigned to a condition in a 2 (level of review numbers: low, high) × 2 (ratings valence level: negative, positive) × 2 (scale boundary included: no, yes) between-subjects design.

Choice Set

Participants saw a choice set of two hand mixers. Choice options were nearly identical except for their average product ratings and number of reviews (see Table 1). We manipulated valence and scale boundary by the specific values used for average product ratings. The negative valence conditions were 1.3 versus 1.0 (boundary included) and 1.4 versus 1.1 (boundary not included). The positive valence conditions were 5.0 versus 4.7 (boundary included) and 4.9 versus 4.6 (boundary not included).

Measure

After viewing the choice options, participants indicated relative preference on the same seven-point scale used in previous studies.

Results

Relative preference between choice options. A 2 (level of review numbers) × 2 (average product ratings valence) × 2 (scale boundary inclusion) ANOVA on relative preference between options yielded a significant main effect of valence ($F_{(1, 404)}$ = 3.98, p = .047), a marginal main effect of the level of review numbers ($F_{(1, 404)}$ = 3.12, p = .078), and an interaction between the level of the review numbers and scale boundary inclusion ($F_{(1, 404)}$ = 4.47, p = .035), qualified by the three-way interaction ($F_{(1, 404)}$ = 12.98, p < .001). To explain the relationship between these factors, we examine the two-way interactions between the level of review numbers and scale boundary inclusion at the negative and positive ends of the rating scale (see Figure 4).

At the positive end of the scale (4.6–5.0), a 2 (level of review numbers) × 2 (scale boundary inclusion) ANOVA on preference yielded a marginal effect of the level of review numbers ($F_{(1, 202)}$ = 2.84, p = .093) qualified by a significant interaction of the level of review numbers and scale boundary inclusion ($F_{(1, 202)}$ = 13.98, p < .001). When the scale boundary is not included (i.e., 4.9 vs. 4.6), a low level of review numbers weakens preference for the higher-rated option with fewer reviews (M_{low} = 4.08, M_{high} = 2.71; $F_{(1, 202)}$ = 14.86, p < .001), which is consistent with the results of our prior studies. By contrast, when the scale boundary is included (i.e., 5.0 vs. 4.7), the effect of the level of review numbers is attenuated (M_{low} = 3.48, M_{high} = 2.96; $F_{(1, 202)}$ = 2.09, p = .15). Consistent with Isaac and Schindler's (2013) findings, this suggests that when products have perfect ratings (5.0), the ratings become significantly more influential in the decision process relative to when products have near-perfect ratings (4.9).

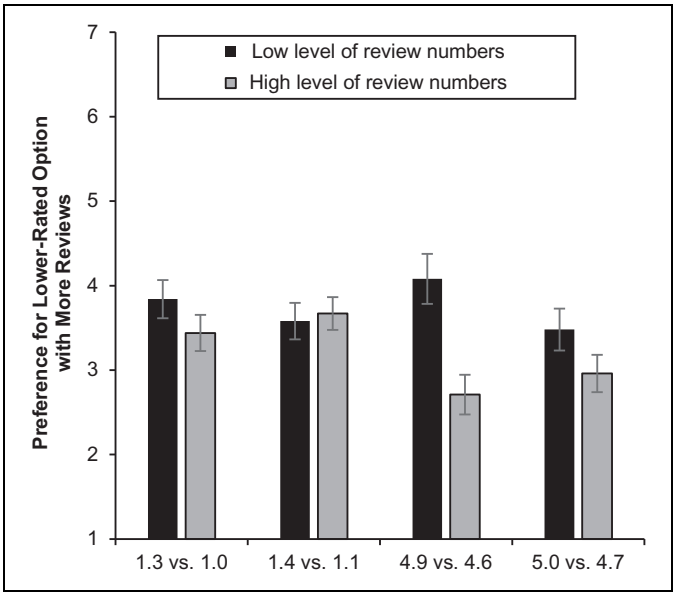

Figure 4. Study 5 results: Preference by ratings valence, level of review numbers, and scale boundary inclusion.

At the negative end of the scale (1.0–1.4), a 2 (level of review numbers) × 2 (scale boundary inclusion) ANOVA on preference yielded no significant effects (p > .20). Across all negative conditions, preference averaged 3.64, where four would indicate no preference between options. We did not predict this effect a priori, and we discuss it next.

Discussion. This study demonstrated that the effect of the level of review numbers is attenuated when the diagnosticity of average product ratings is high as a result of the inclusion of the rating scale boundary in the positive-valence choice set. This finding provides additional support to H_2. Furthermore, an interesting asymmetric valence effect emerged, such that, at the negative end of the scale, the effect of the level of review numbers was attenuated regardless of whether the scale boundary was included. This could be a function of consumers choosing between two unattractive options (Dhar and Sherman 1996), which is known to increase difficulty of making a choice and likelihood of choice deferral. To further explore this point, in a follow-up study we compared the effect of the level of review numbers across very negative ratings (1.3 vs. 1.0) and somewhat negative ratings (2.4 vs. 2.1). Replicating the findings in Study 5, we found no effect of the level of review numbers in the very negative condition (M_{low} = 3.51, M_{high} = 3.14; $F_{(1, 83)}$ = 1.13, p = .29). Yet, consistent with our findings in Study 3, we replicated the effect of the level of review numbers in the somewhat negative condition (M_{low} = 3.89, M_{high} = 3.16; $F_{(1, 85)}$ = 3.75, p = .056). Furthermore, consistent with the view that choosing between two extremely unattractive options increases choice difficulty, potentially leading to the random choice between low-valence options in the main study, the rate of choice deferral was significantly

higher in the very negative condition (45.9%) as compared with the somewhat negative condition (21.8%; $\chi^2(1) = 10.93$, $p = .001$). This finding suggests that consumer decision processes differ in the context of extremely negative and somewhat negative choice sets.

Study 6: Measuring the Difference in Diagnosticity of Average Product Ratings and the Number of Reviews Through Self-Reporting

Having established the robust effect of the level of review numbers on consumer decisions, we next shift our focus to directly measuring the relative diagnosticity of the average product ratings and number of reviews. H_3 suggests that the preference between choice options is driven by the difference in diagnosticity of average product ratings and the number of reviews. When the level of review numbers is low (vs. high), the diagnosticity of the number of reviews increases relative to the diagnosticity of average product ratings. To test this proposition directly, in this study we ask participants how important each attribute was to their decision and compute the difference between the importance of average product ratings and the number of reviews to demonstrate the changing diagnosticity of these attributes between conditions.

Participants and Design

We recruited 183 people from MTurk to participate in this study in exchange for $.50. We randomly assigned them to a condition with one of three levels of review numbers (low, high, control) in a between-subjects design.

Choice Set

Participants saw a choice set of two blenders. Choice options were nearly identical except for their average product ratings and number of reviews, as described previously (see Table 1). Similar to Study 2, we used discrete choice as our dependent measure. However, this time we integrated the choice and deferral measures into one, so participants were told that they could choose option A or option B; alternatively, they could "defer purchase and look elsewhere," as this more closely mirrors the consumer decision process. Unlike in our other studies, to test H_3, we then asked participants to "indicate the importance of each attribute in making your decision" for the five attributes (image, brand, price, average product rating, and review volume) on seven-point scales (1 = "not at all important," and 7 = "extremely important"). As we expected, there were no significant differences across conditions of the perceived diagnosticity of product image, brand, or price ($p > .10$) because they were relatively comparable across products. Next, we computed a difference score of the review attribute diagnosticities (diagnosticity of average product ratings minus the diagnosticity of the number of reviews) to demonstrate the changing diagnosticities of the review attributes as a function of the level of review numbers. Thus, a positive score indicates that average product ratings are more diagnostic than the number of reviews and vice versa. As the difference score approaches zero, this indicates that consumers would equally weigh average product ratings and the number of reviews in their decisions.

Results

Choice of the higher-rated option with fewer reviews. Examining only participants who chose one of the two product options (N = 153), a binary logistic regression in which we dummy coded our level of review numbers yielded an omnibus effect of the level of review numbers ($\chi^2(2) = 15.07$, $p = .001$). Consistent with our prior studies, when the level of review numbers was high ($P_{high} = 71\%$; $\chi^2(1) = 11.40$, $p = .001$) or absent ($P_{control} = 71\%$; $\chi^2(1) = 11.83$, $p = .001$) participants were significantly more likely to choose the higher-rated option with fewer reviews relative to when the level of review numbers was low ($P_{low} = 36\%$). There was no significant difference in the high and control conditions ($p > .95$).

Rate of choice deferral. A binary logistic regression in which we dummy coded our level of review number yielded an omnibus effect of the level of review numbers ($\chi^2(2) = 9.42$, $p = .009$). When the level of review numbers was high ($P_{high} = 13\%$; $\chi^2(1) = 4.64$, $p = .031$) or absent ($P_{control} = 8\%$; $\chi^2(1) = 7.60$, $p = .006$), participants were significantly less likely to defer choice relative to when the level of review numbers was low ($P_{low} = 29\%$), replicating results of Study 3. There was no significant difference in choice deferral between the high and control conditions, consistent with our prior studies ($p > .40$).

Difference in diagnosticity of review attributes. A one-way (level of review numbers: low, high, control) ANOVA on the difference in diagnosticity of average product ratings and the number of reviews yielded a marginal omnibus effect ($F_{(2, 150)} = 4.59$, $p = .061$). Consistent with H_3, the difference in perceived diagnosticity between the review attributes is smaller when the level of review numbers was low ($M_{low} = .26$) relative to high ($M_{high} = .85$; $t_{(150)} = 2.28$, $p = .024$) or absent ($M_{absent} = .75$; $t_{(150)} = 1.89$, $p = .061$). There was no significant difference between the absent and high level of review numbers ($p > .65$). In other words, average product ratings are considered significantly more diagnostic than the number of reviews when the level of review numbers is absent or high, relative to when the level of review numbers is low (i.e., when the two attributes are more equally diagnostic).

Preference mediation through the difference in diagnosticity of review attributes. We used mediation analysis (Model 4; Preacher, Rucker, and Hayes 2007) to demonstrate that the effect of the level of review numbers (low vs. high) on consumer preference is driven by the difference in the diagnosticity of average product ratings and number of reviews. As we expected, the model demonstrated that the effect of the level of review numbers on consumer preference was mediated through

the difference in perceived diagnosticity of average product ratings and the number of reviews ($B = -.49$; $CI_{95\%} = [-1.22, -.10]$).

Discussion. This study provided support for H_3 by demonstrating that the effect of the level of review numbers on choice option preference was driven by the difference in diagnosticity of average product ratings and number of reviews: as the diagnosticity of the number of reviews increases (i.e., when the level of review numbers is low), the preference shifts away from the higher-rated option with fewer reviews toward the lower-rated option with more reviews. We further showed, consistent with H_3, that average product ratings are considered more diagnostic than the number of reviews, but this difference in diagnosticity is attenuated when the level of review numbers is low. Finally, we demonstrated that as the diagnosticity of average product ratings and the number of reviews become more equal, choice deferral increases, which is consistent with the findings of Study 2 and prior work demonstrating the link between trade-off difficulty and choice deferral (Dhar and Nowlis 1999; Etkin and Ghosh 2018; Tversky and Shafir 1992).

Study 7: Measuring the Difference in Diagnosticity of Average Product Ratings and Number of Reviews Through Eye Tracking

The objective of Study 7 was to further test H_3 by demonstrating how the level of review numbers differentially affects consumers' attention to average product ratings and the number of reviews. An information uptake measure, such as eye movement, has been validated to provide insight into cognitive processes underlying choice (Ashby et al. 2016; Orquin and Loose 2013). Specifically, research has shown that attention is often top-down driven and decision makers are more likely to attend to stimuli with higher task relevance (Orquin and Loose 2013). Thus, in this study we capture consumer attention (measured with eye tracking) to provide further evidence for the differential diagnosticity of average product ratings and the number of reviews in consumer decisions.

We have argued that consumers infer different diagnostic values of average product ratings and number of reviews as a function of the level of the review numbers (H_3). Specifically, in choice sets with neutral and low average product ratings, when consumers see that the number of reviews is low, it signals to them that average product ratings may not be as diagnostic of product quality as when the number of reviews is high or when no review number information is displayed. Therefore, in terms of consumers' attention, when examining review attributes, we would expect that consumers would be more likely to reexamine average product ratings after viewing a low (vs. high) number of reviews. This happens because the diagnosticity of the number of reviews increases under these conditions (H_3), leading consumers to reappraise the diagnosticity of the average product ratings in light of the information garnered from the number of reviews.

To test this argument, we use eye-tracking measurements to determine not only gaze times (i.e., time spent looking at) for each attribute but also the sequence of fixations (i.e., order looked at) for all attributes to determine whether consumers are more likely to reexamine average product ratings after viewing low versus high levels of review numbers.

Participants and Design

Ninety-two undergraduate students participated in the study in exchange for course credit. They were randomly assigned to conditions with one of two levels of review numbers (low, high) in a between-subjects design. Participants were randomly selected two at a time from a larger sample of research participants to participate in the eye-tracking study. After engaging in a short eye-tracking calibration task, participants followed a similar paradigm to prior studies.

Choice Set

Participants saw a choice set of two microwave ovens. Choice options were nearly identical except for their average product ratings and number of reviews as described previously (see Table 1). Relative preference between choice options was measured on the same seven-point scale as in our previous studies.

Additional Measures

We defined areas of interest as parts of the screen where corresponding product attributes were displayed and measured the number of eye fixations and gaze times for each attribute. Fixations refer to the frequency with which participants would look at a given attribute, and gaze times refer to the amount of time participants spent looking at the specific attributes. As we expected, there were no significant differences across conditions for fixations or gaze times of product images, brand names, prices, or highlighted information ($p > .10$); thus, we do not discuss these further.

Results

Relative preference between choice options. A one-way (level of review numbers: low, high) ANOVA on preference yielded a significant effect ($F_{(1, 90)} = 10.32$, $p = .002$). Consistent with prior studies, preference for the higher-rated option with fewer reviews was weaker when the level of review numbers was low ($M_{low} = 4.89$) versus high ($M_{high} = 3.68$).

Transition matrices. To provide further support for this process, we also derive transition matrices from the eye-tracking data. Doing so enables us to demonstrate the probabilities of participants transitioning their attention from one attribute to the next. As we discussed previously, we argue that a low level (relative to a high level) of review numbers increases the diagnosticity of the number of reviews, and this causes consumers to reevaluate average product ratings. To demonstrate this, we assessed the differential probabilities of participants shifting

Table 3. Study 7: Eye-Tracking Attribute Transition Matrices by Level of Review Numbers.

	To Image	To Brand and Price	To Rating	To Number of Reviews	To Additional Information	To End
A: Low Level of Reviews						
From image	68.23%	14.47%	4.51%	2.26%	6.02%	4.51%
From brand and price	8.94%	72.81%	12.23%	2.37%	2.92%	0.73%
From rating	3.50%	10.07%	62.58%	19.69%	3.72%	0.44%
From volume	2.65%	3.41%	24.24%	49.62%	18.94%	1.14%
From additional information	9.66%	1.87%	1.71%	2.80%	82.09%	1.87%
B: High Level of Reviews						
From image	71.74%	12.89%	3.64%	2.15%	6.12%	3.47%
From brand and price	9.35%	73.28%	12.02%	1.53%	2.48%	1.34%
From rating	4.26%	9.31%	64.10%	14.63%	6.12%	1.60%
From volume	4.69%	2.17%	13.36%	60.65%	17.33%	1.81%
From additional info	8.20%	1.64%	1.37%	4.37%	83.47%	.96%

their attention from the number of reviews to average product ratings as a function of the level of review numbers (for the complete transition matrices, see Table 3). Consistent with our theory, participants were significantly more likely to return their attention to the average product ratings after viewing the number of reviews when the level of review numbers was low relative to high ($P_{low} = .24$, $P_{high} = .13$; $z = 3.25$, $p < .01$). This suggests that participants were nearly twice as likely to return their attention to average product ratings when the level of review numbers was low versus high. Importantly, the transition proportions from the number of reviews to all other attributes were not significantly different across conditions ($p > .10$).

Difference in fixation counts for the review attributes. Because our variable of interest is the difference in attention paid to average product ratings and the number of reviews, we calculated the difference in fixations between average product ratings and the number of reviews. A one-way (level of review numbers: low, high) ANOVA on the difference in fixation counts yielded a significant effect ($F_{(1, 90)} = 7.05$, $p = .009$). As we expected, the difference in fixations between average product ratings and the number of reviews was greater when the level of review numbers was low ($M_{low} = 4.29$ fixations) relative to high ($M_{high} = 2.11$ fixations). This is consistent with our view that a low level of review numbers causes consumers to reevaluate the average product ratings attribute, thus increasing overall attention paid to that attribute.

Difference in gaze times for the review attributes. A one-way (level of review numbers: low, high) ANOVA on the difference in gaze times yielded a significant effect ($F_{(1, 90)} = 10.59$, $p = .002$). As we expected, the difference in gaze times between average product ratings and the number of reviews was greater when the level of review numbers was low ($M_{low} = 12.40$ seconds) relative to high ($M_{high} = 5.78$ seconds). Consistent with our prior findings, consumers seem to pay more attention to average product ratings when the level of review numbers is low, and we argue that this occurs because a low number of reviews causes consumers to reevaluate the inferences from average product ratings.

Mediation through the difference in gaze times. We argue that gaze times are a more precise measure of attention relative to fixations because they quantify the time spent on an attribute. As such, we demonstrate that consumers are likely to pay more attention to average product ratings when the level of review numbers is low; thus, this difference in gaze times would mediate the influence of the level of review numbers on consumer preference. Using mediation analysis (Model 4; Preacher, Rucker, and Hayes 2007), we find support for the argument that the difference in gaze times between average product ratings and the number of reviews mediates the effect of the level of review numbers on consumer preference between choice options ($B = .25$; $CI_{95\%} = [.04, .62]$).

Discussion. This study provided further evidence for H_3 by demonstrating that the difference in gaze times for the average product ratings and the number of reviews dictates consumer preference between options. When the level of review numbers is low, it signals to consumers that the average product ratings may be relatively closer in diagnosticity to that of the number of reviews, leading consumers to reevaluate this attribute before reaching a decision. Yet when the level of review numbers is high, there is a hierarchy in diagnosticity between the two attributes, and consumers can reach a decision faster without reevaluating average product ratings. Thus, this study compliments Study 6 in providing a different measure of diagnosticity of average product reviews and number of reviews.

General Discussion

Across seven studies, we find consistent support for our propositions that average product ratings are a more diagnostic cue of product quality than the number of reviews and that the number of reviews can become more influential in consumer decisions when (1) average product ratings are negative or neutral and (2) the level of review numbers is low. This change in attribute diagnosticity leads to a systematic shift in

preference away from higher-rated options with fewer reviews toward lower-rated options with more reviews. Furthermore, when the diagnosticities of these two review attributes are closest to each other, consumers experience trade-off difficulty, as evidenced by increased choice deferral.

This work makes several contributions. Theoretically, we contribute to the literature on numerical cognition, which investigates how consumers process numbers as information; we do so by examining distinctions between interpretations of continuous and discrete variables, scalar variability, and the relationships between absolute and relative number comparisons (Gallistel and Gelman 2000). We add to this literature by demonstrating how consumers integrate multiple attributes that use different numeric scales. Specifically, we investigate how consumers integrate average product ratings, which are usually clearly bound on a one- to five-point scale, and numbers of reviews, which are usually unbound. Prior work on attribute diagnosticity (Feldman and Lynch 1988; Herr, Kardes, and Kim 1991; Purohit and Srivastava 2001) has suggested that consumers weigh attributes in their decisions differently as a function of these attributes' perceived diagnosticity as signals of product quality. Consistent with this research, we demonstrate that average product ratings are more likely to be incorporated in consumer judgments as a signal of product quality than the number of reviews because of the natural differences in the scaling of these attributes. Importantly, we further outline conditions under which the perceived diagnosticity of the number of reviews can increase, leading to joint influences of average product ratings and number of reviews in consumers' choices. Our investigation into these conditions advances the understanding of how consumers incorporate numerical cues on different scales into their judgments.

While we consistently demonstrate the persistent effect of the level of review numbers on consumer preferences across seven studies in the main article and eight in the Web Appendix, several other possible moderators exist that we did not address. Recent work has demonstrated that individual review sentiment has an influence on consumer decisions above and beyond aggregate review information (Ludwig et al. 2013; Villarroel Ordenes et al. 2017), yet this work has not looked at the three-way interactive relationship among product ratings, number of reviews, and textual content of individual reviews. Naturally, we believe there are several purchase decisions in which consumers would be more likely to consult the reviews' text in addition to aggregate measures of reviews. For instance, it would seem that, when average product ratings and the level of review numbers hold similar diagnosticity (e.g., when the level of review numbers is low) consumers would be more likely to seek additional information to differentiate options. This would be consistent with prior work demonstrating that making choices is more difficult when consumers are making trade-offs between attributes of similar importance (Chatterjee and Heath 1996; Dhar and Simonson 2003), therefore increasing the need for additional information (as we demonstrate in Study 2). If consumers choose to seek more information about product options, such as by reading individual reviews, such information may become more influential in judgments as compared with the previously accessed information (Tversky and Shafir 1992). This could result in an increase in the diagnosticity of textual information relative to numerical information contained in aggregate reviews. Exactly how consumers integrate numerical and textual information to form a single judgment remains an worthwhile avenue for future research.

Managerial Implications

Next, we provide an illustration of how the results of our study can be incorporated into business practices. We interviewed two managers involved with product review acquisition strategies for their respective brands. One, the chief executive officer of a nutritional supplement company, instructs his team to employ a proactive strategy in which it aggressively pursues reviews from customers through email nudges after purchase and offers them steep product discounts. The discounts are used to increase the level of review numbers as a function of more sales, whereas the email nudges are intended to increase the sales-to-review conversion ratio. The other, an associate brand manager for a leading home electrics company, relies on a reactive strategy in which the company offers free products in exchange for honest reviews. The interviews confirmed our perception that business practitioners appreciate the importance of both average product ratings and number of reviews in driving product sales.

We should note that the aforementioned strategies, while certainly effective in increasing the number of reviews, do not necessarily result in improved average product ratings. Indeed, depending on the review stimulation strategy employed as well as the dynamics in the particular rating environment (e.g., see Moe and Trusov 2011), the average product rating can increase, decrease, or remain unchanged. Thus, in this section we focus solely on the number of reviews, making a simplifying assumption that the average product rating is not affected by the stimulation. As various stimulation instruments typically carry some associated costs (e.g. reduced earnings from deep discounts or free products), we argue that managers need to trade off the potential benefits of receiving additional reviews against the costs associated with these efforts. Because the detailed financial analysis would call for additional assumptions about the product profit margins and the cost of stimulation, for the sake of generalizability, we explore how the incremental number of reviews may drive sales volume under different choice scenarios. Specifically, we show how the firm could benefit from increasing the number of reviews in early- and mid-stage product life cycles when they face competitors with more reviews.

We conducted a simulation in which the higher-rated good with fewer reviews accrued reviews at the same rate as the lower-rated good with more reviews or at twice the rate (e.g., when a review acquisition strategy was employed). We then crossed this with whether the firm employed a proactive review

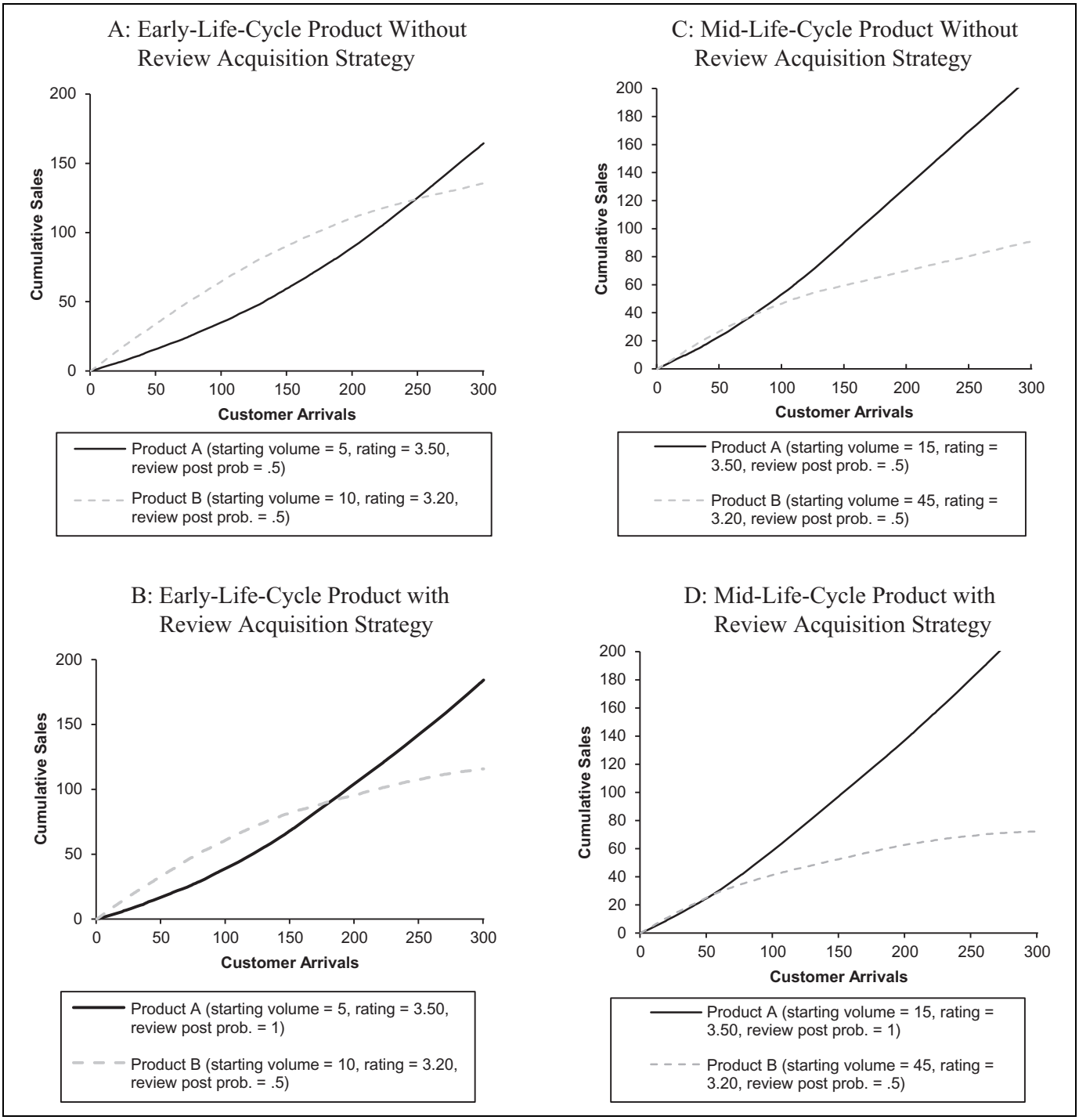

Figure 5. Implications simulations for early- and mid-life-cycle products.

acquisition strategy early on in the product life cycle (e.g., at 5 reviews) or a reactive strategy later on in the life cycle (e.g., at 15 reviews). The simulation was conducted in MATLAB using choice shares computed from Study 1. We took the relative preference measure used in that study and pooled the numbers that indicated preference for the higher-rated option with fewer reviews (i.e., 1–3) and the numbers that indicated preference for the lower-rated option with more reviews (i.e., 5–7). Participants who indicated no preference (i.e., 4) were excluded from the choice share computation. This resulted in several data points ranging from low to mid- to high in the level of review numbers. Then, we used the MATLAB surface fitting function ("fit") to interpolate choice shares for various combinations of review volume pairs.

Our simulations were agnostic as to what type of review acquisition strategy was employed (e.g., email nudges, price

promotions, free product), but we modeled the effect of employing an acquisition strategy by doubling the likelihood of a review being written. We seeded the baseline likelihood of receiving a review after a purchase at 50%, and the likelihood of receiving a review when a review acquisition strategy is employed at 100%. The likelihood of leaving a review has been estimated as low as .001% depending on the product category. Thus, the results we discuss here are extremely conservative estimates of the impact the number of reviews might have on choice shares over time.

In the first two simulations, we investigated a proactive strategy in which the higher-rated product had 5 reviews, whereas the lower-rated product had 10 reviews. In Scenario 1, we held the likelihood of receiving a review constant across both products at 50%. Thus, the first scenario investigated how long it would take for a product with a higher rating, but five fewer reviews, to accumulate more sales than the other products. As we illustrate in Figure 5, Panel A, it would take approximately 240 consumers purchasing products in this category for this to occur. In Scenario 2 (Figure 5, Panel B), we assume that the manager of the higher-rated product employs an active review management strategy and doubles the likelihood of receiving product reviews. In this scenario, the higher-rated product reaches the dominant sales position roughly 33% more quickly—around 160 consumer category purchases.

In Scenarios 3 and 4 (Figure 5, Panels C and D), we investigate a reactive strategy in which the market has somewhat matured (i.e., 15 and 45 reviews, respectively). Again, in Scenario 3, we assume that each product has a 50% likelihood of receiving a review after it is purchased. Here, it takes approximately 75 additional customer purchases in the product category before the higher-rated product surpasses the lower-rated product in sales. In Scenario 4, we once again assume that the manager of the higher-rated product actively manages its reviews and can double the likelihood of receiving a review after purchase. This results in the higher-rated product surpassing the lower-rated product in sales roughly 20% earlier, or after 60 consumer category purchases.

While these scenarios are just a few of the many possible that exist in the marketplace, they provide additional support for our claim that the number of reviews is highly influential in consumer decisions, and managers would be wise to oversee their growth. As new products with fewer reviews enter against incumbents with more reviews, employing a proactive strategy to spur an increase in product reviews can quickly decrease the disadvantage that a product manager faces when competing against established products. Obviously, a comprehensive optimization of the review stimulation strategy should account for various additional factors (e.g., costs associated with discounted products or additional review nudging, anticipated competitive response, product profit margins); nonetheless, we argue that the results of our study provide an important input to managerial decision making.

Conclusions

This research outlines conditions in which the diagnosticity of the number of reviews increases relative to the diagnosticity of average product ratings (as cues of product quality), potentially leading to suboptimal decisions for consumers and an increase in choice deferral for brands and retailers. Theoretically, we argue that an inherent difference in the types of scales in which these attributes are presented (bound and unbound) leads to the observed difference in their diagnosticity; by demonstrating how consumers integrate attributes on both numeric scales into a single judgment, we contribute to the literature in numerical cognition. Furthermore, we provide clarity to the debate on the relative influence of average product ratings and the number of reviews (see Floyd et al. 2014; You, Vadakkepatt, and Joshi 2015) by demonstrating the conditional influences of the number of reviews under various average product rating valence conditions.

Acknowledgments

The authors are grateful to the *JM* review team, who provided invaluable feedback and guidance in each stage of the review process.

Area Editor

Alexander Chernev served as area editor for this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Ashby, Nathaniel J. S., Joseph G. Johnson, Ian Krajbich, and Michel Wedel (2016), "Applications and Innovations of Eye-Movement Research in Judgment and Decision Making," *Journal of Behavioral Decision Making*, 29 (2/3), 96–102.

Bagchi, Rajesh, and Xingbo Li (2010), "Illusionary Progress in Loyalty Programs: Magnitudes, Reward Distances, and Step-Size Ambiguity," *Journal of Consumer Research*, 37 (2), 888–901.

Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs (2001), "Bad Is Stronger Than Good," *Review of General Psychology*, 5 (4), 323–70.

Bockenholt, Ulf, Dietrich Albert, Michael Aschenbrenner, and Franz Schmalhofer (1991), "The Effects of Attractiveness, Dominance, and Attribute Differences on Information Acquisition in Multiattribute Binary Choice," *Organizational Behavior and Human Decision Processes*, 49 (2), 258–81.

Brenner, Lyle, Yuval Rottenstreich, and Sanjay Sood (1999), "Comparison, Grouping, and Preference," *Psychological Science*, 10 (3), 225–9.

BrightLocal (2017), "Local Consumer Review Survey," (accessed February 24, 2018), https://www.brightlocal.com/learn/local-consumer-review-survey.

Burson, Katherine A., Richard P. Larrick, and John G. Lynch Jr. (2009), "Six of One, Half Dozen of the Other: Expanding and Contracting Numerical Dimensions Produces Preference Reversals," *Psychological Science*, 20 (9), 1074–78.

Chandon, Pierre, and Nailya Ordabayeva (2017), "The Accuracy of Less: Natural Bounds Explain Why Quantity Decreases Are Estimated More Accurately than Quantity Increases," *Journal of Experimental Psychology: General*, 146 (2), 250–68.

Chatterjee, Subimal, and Timothy B. Heath (1996), "Conflict and Loss Aversion in Multiattribute Choice: The Effects of Trade-Off Size and Reference Dependence on Decision Difficulty," *Organizational Behavior and Human Decision Processes*, 67 (2), 144–55.

Chen, Pei-Yu, Samita Dhanasobhon, and Michael D. Smith (2008), "All Reviews Are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.com," working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=918083.

Chen, Zoey, and Nicholas H. Lurie (2013), "Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth," *Journal of Marketing Research*, 50 (4), 463–76.

Chen, P.Y., S.Y. Wu, and J. Yoon (2004). The impact of online recommendations and consumer feedback on sales. *ICIS 2004 Proceedings*, 58.

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.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets," *Marketing Science*, 29 (5), 944–57.

Cho, Eunice K., Uzma Khan, and Ravi Dhar (2013), "*Comparing Apples to Apples or Apples to Oranges: The Role of Mental Representation in Choice Difficulty*," *Journal of Marketing Research*, 50 (4), 505–16.

Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," *Journal of Management Information Systems*, 23 (2), 149–71.

De, Langhe., Bart Philip M. Fernbach, and Donald R. Lichtenstein (2015), "Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings," *Journal of Consumer Research*, 42 (2), 817–33.

Dellarocas, Chrysanthos, Xiaoquan Michael Zhang, and Neveen F. Awad (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21 (4), 23–45.

Dhar, Ravi, and Stephen M. Nowlis (1999), "The Effect of Time Pressure on Consumer Choice Deferral," *Journal of Consumer Research*, 25 (4), 369–84.

Dhar, Ravi, and Steven J. Sherman (1996), "The Effect of Common and Unique Features in Consumer Choice," *Journal of Consumer Research*, 23 (3), 193–203.

Dhar, Ravi, and Itamar Simonson (2003), "The Effect of Forced Choice on Choice," *Journal of Marketing Research*, 40 (2), 146–60.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do Online Reviews Matter? An Empirical Investigation of Panel Data," *Decision Support Systems*, 45 (4), 1007–16.

Etkin, Jordan, and Anastasiya Pocheptsova Ghosh (2018), "When Being in a Positive Mood Increases Choice Deferral," *Journal of Consumer Research*, 45 (1), 208–25.

Feldman, Jack M., and John G. Lynch (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (3), 421–35.

Fisher, Matthew, George E. Newman, Ravi Dhar, and Stijn van Osselaer (2018), "Seeing Stars: How the Binary Bias Distorts the Interpretation of Customer Ratings," *Journal of Consumer Research*, 45 (3), 471–89.

Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling (2014), "How Online Product Reviews Affect Retail Sales: A Meta-Analysis," *Journal of Retailing*, 90 (2), 217–32.

Gallistel, Charles R., and Rochel Gelman (2000), "Non-Verbal Numerical Cognition: From Reals to Integers," *Trends in Cognitive Sciences*, 4 (2), 59–65.

Godes, David, and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

González-Vallejo, Claudia, and Elizabeth Moran (2001), "The Evaluability Hypothesis Revisited: Joint and Separate Evaluation Preference Reversal as a Function of Attribute Importance," *Organizational Behavior and Human Decision Processes*, 86 (2), 216–33.

He, Stephen X., and Samuel D. Bond (2015), "Why is the Crowd Divided? Attribution for Dispersion in Online Word of Mouth," *Journal of Consumer Research*, 41 (6), 1509–27.

Herr, Paul M., Frank R. Kardes, and John Kim (1991), "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective," *Journal of Consumer Research*, 17 (4), 454–62.

Ho-Dac, Nga N., Stephen J. Carson, and William L. Moore (2013), "The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter?" *Journal of Marketing*, 77 (6), 37–53.

Hsee, Christopher K. (1996), "The Evaluability Hypothesis: An Explanation for Preference Reversals Between Joint and Separate Evaluations of Alternatives," *Organizational Behavior and Human Decision Processes*, 67 (3), 247–57.

Hsee, Christopher K. (2000), "Attribute Evaluability and Its Implications for Joint-Separate Evaluation Reversals and Beyond," working paper, Booth School of Business, University of Chicago, http://faculty.chicagobooth.edu/christopher.hsee/vita/papers/attributeevaluability.pdf].

Hsee, Christopher K., George F. Loewenstein, Sally Blount, and Max H. Bazerman (1999), "Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis," *Psychological Bulletin*, 125 (5), 576–90.

Hsee, Christopher K., and Jiao Zhang (2010), "General Evaluability Theory," *Perspectives on Psychological Science*, 5 (4), 343–55.

Isaac, Mathew S., and Robert M. Schindler (2013), "The Top-Ten Effect: Consumers' Subjective Categorization of Ranked Lists," *Journal of Consumer Research*, 40 (6), 1181–202.

Ito, Tiffany A., Jeff T. Larsen, N. Kyle Smith, and John T. Cacioppo (1998), "Negative Information Weighs More Heavily on the Brain: The Negativity Bias in Evaluative Categorizations," *Journal of Personality and Social Psychology*, 75 (4), 887–900.

Johnson, Eric J., and Robert J. Meyer (1984), "Compensatory Choice Models of Noncompensatory Processes: The Effect of Varying Context," *Journal of Consumer Research*, 11 (1), 528–41.

Kahneman, Daniel, and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47 (2), 263–91.

Khare, Adwait, Lauren I. Labrecque, and Anthony K. Asare (2011), "The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings," *Journal of Retailing*, 87 (1), 111–26.

King, Dan, and Chris Janiszewski (2011), "The Sources and Consequences of the Fluent Processing of Numbers," *Journal of Marketing Research*, 48 (2), 327–41.

Kirmani, Amna, and Akshay R. Rao (2000), "No Pain, No Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality," *Journal of Marketing*, 64 (2), 66–79.

Lembregts, Christophe, and Mario Pandelaere (2013), "Falling Back on Numbers: Quantitative Specifications as Sources of Control," in *Proceedings of the 2013 Society for Consumer Psychology Annual Winter Conference*, San Antonio, TX.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (3), 74–89.

Luca, Michael, and Georgios Zervas (2016), "Fake It Till You Make it: Reputation, Competition, and Yelp Review Fraud," *Management Science*, 62 (12), 3412–27.

Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels, and Gerard Pfann (2013), "More than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing*, 77 (1), 87–103.

Lynch, John G., Howard Marmorstein, and Michael F. Weigold (1988), "Choices from Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations," *Journal of Consumer Research*, 15 (2), 169–84.

Markman, Arthur B., and Douglas L. Medin (1995), "Similarity and Alignment in Choice," *Organizational Behavior and Human Decision Processes*, 63 (2), 117–30.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014), "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, 104 (8), 2421–55.

McAuley, Julian, Rahul Pandey, and Jure Leskovec (2015), "Inferring Networks of Substitutable and Complementary Products," in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 785–94.

Moe, Wendy W., and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (3), 444–56.

Monga, Ashwani, and Rajesh Bagchi (2011), "Years, Months, and Days Versus 1, 12, and 365: The Influence of Units Versus Numbers," *Journal of Consumer Research*, 39 (2), 185–98.

Monroe, Kent B. (1971), "Measuring Price Thresholds by Psychophysics and Latitudes of Acceptance," *Journal of Marketing Research*, 8 (4), 460–4.

Moore, Sarah G. (2015), "Attitude Predictability and Helpfulness in Online Reviews: The Role of Explained Actions and Reactions," *Journal of Consumer Research*, 42 (1), 30–44.

National Retail Federation (2010), "E-Commerce Sales," (accessed November 28, 2016), [available at http://research.nrffoundation.com].

Nowlis, Stephen M., and Itamar Simonson (1997), "Attribute-task Compatibility as a Determinant of Consumer Preference Reversals," *Journal of Marketing Research*, 205–18.

Nowlis, Stephen M., Ravi Dhar, and Itamar Simonson (2010), "The Effect of Decision Order on Purchase Quantity Decisions," *Journal of Marketing Research*, 47 (4), 725–37.

Orquin, Jacob L., and Simone Mueller Loose (2013), "Attention and Choice: A Review on Eye Movements in Decision Making," *Acta Psychologica*, 144 (1), 190–206.

Pandelaere, Mario, Barbara Briers, and Christophe Lembregts (2011), "How to Make a 29% Increase Look Bigger: The Unit Effect in Option Comparisons," *Journal of Consumer Research*, 38 (5), 308–22.

Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*, 16 (2), 366–87.

Preacher, Kristopher J., Derek D. Rucker, and Andrew F. Hayes (2007), "Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions," *Multivariate Behavioral Research*, 42 (1), 185–227.

Purohit, Devavrat, and Joydeep Srivastava (2001), "Effect of Manufacturer Reputation, Retailer Reputation, and Product Warranty on Consumer Judgments of Product Quality: A Cue Diagnosticity Framework," *Journal of Consumer Psychology*, 10 (3), 123–34.

Rao, Akshay R., and Kent B. Monroe (1988), "The Moderating Effect of Prior Knowledge on Cue Utilization in Product Evaluations," *Journal of Consumer Research*, 15 (2), 253–64.

Rao, Akshay R., and Kent B. Monroe (1989), "The Effect of Price, Brand Name, and Store Name on Buyers' Perceptions of Product Quality: An Integrative Review," *Journal of Marketing Research*, 26 (3), 351–57.

Richardson, Paul S., Alan S. Dick, and Arun K. Jain (1994), "Extrinsic and Intrinsic Cue Effects on Perceptions of Store Brand Quality," *Journal of Marketing*, 58 (4), 28–36.

Ross, Jr. William T., and Elizabeth H. Creyer (1992), "Making Inferences about Missing Information: The Effects of Existing Information," *Journal of Consumer Research*, 19 (1), 14–25.

Rozenkrants, Bella, S. Christian Wheeler, and Baba Shiv (2017), "Self-Expression Cues in Product Rating Distributions: When

People Prefer Polarizing Products," *Journal of Consumer Research*, 44 (4), 759–77.

Rozin, Paul, and Edward B. Royzman (2001), "Negativity Bias, Negativity Dominance, and Contagion," *Personality and Social Psychology Review*, 5 (4), 296–320.

Schley, Dan R., Christophe Lembregts, and Ellen Peters (2017), "The Role of Evaluation Mode on the Unit Effect," *Journal of Consumer Psychology*, 27 (2), 278–86.

Schlosser, Ann E. (2011), "Can Including Pros and Cons Increase the Helpfulness and Persuasiveness of Online Reviews? The Interactive Effects of Ratings and Arguments," *Journal of Consumer Psychology*, 21 (3), 226–39.

Slovic, Paul (1966), "Cue-Consistency and Cue-Utilization in Judgment," *American Journal of Psychology*, 79 (3), 427–34.

Slovic, Paul, and Sarah Lichtenstein (1971), "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *Organizational Behavior and Human Performance*, 6 (6), 649–744.

Slovic, Paul, and Douglas MacPhillamy (1974), "Dimensional Commensurability and Cue Utilization in Comparative Judgment," *Organizational Behavior and Human Performance*, 11 (2), 172–94.

Sun, Monic (2012), "How Does the Variance of Product Ratings Matter?" *Management Science*, 58 (4), 696–707.

Thaler, Richard (1980), "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior & Organization*, 1 (1), 39–60.

Tsiros, Michael, and Vikas Mittal (2000), "Regret: A Model of Its Antecedents and Consequences in Consumer Decision Making," *Journal of Consumer Research*, 26 (4), 401–17.

Tversky, Amos, and Daniel Kahneman (1985), "The Framing of Decisions and the Psychology of Choice," in *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*, Vincent T. Covello, Jeryl L. Mumpower, Pieter Jan M. Stallent and V.R.R. Uppuluri, eds. Berlin: Springer, 107–29.

Tversky, Amos, and Daniel Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5 (4), 297–323.

Tversky, Amos, and Eldar Shafir (1992), "Choice Under Conflict: The Dynamics of Deferred Decision," *Psychological Science*, 3 (6), 358–61.

Villarroel Ordenes, Francisco, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (2017), "Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media," *Journal of Consumer Research*, 43 (6), 875–94.

Yeung, Catherine WM, and Dilip Soman (2005), "Attribute Evaluability and the Range Effect," *Journal of Consumer Research*, 32 (3), 363–9.

You, Ya, Gautham G. Vadakkepatt, and Amit M. Joshi (2015), "A Meta-Analysis of Electronic Word-of-Mouth Elasticity," *Journal of Marketing*, 79 (2), 19–39.

Zeelenberg, Marcel, and Rik Pieters (2007), "A Theory of Regret Regulation 1.0," *Journal of Consumer Psychology*, 17 (1), 3–18.

Zhang, Shi, and Arthur B. Markman (1998), "Overcoming the Early Entrant Advantage: The Role of Alignable and Nonalignable Differences," *Journal of Marketing Research*, 413–426.

Zhang, Shi, and Arthur B. Markman (2001), "Processing Product Unique Features: Alignability and Involvement in Preference Construction," *Journal of Consumer Psychology*, 11 (1), 13–27.

Zhu, Feng, and Xiaoquan Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74 (2), 133–48.