Title

The Influence of Mean Product Ratings on Review Judgments and Search

Permalink

https://escholarship.org/uc/item/1w77z0j5

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Katz, Daniel
Bartels, Daniel

Publication Date

2022

Peer reviewed

The Influence of Mean Product Ratings on Review Judgments and Search
Daniel Katz and Daniel Bartels

Abstract

We investigate the way people judge how helpful a review is in informing their decision as to whether to make a purchase. In particular, we are interested in how the summary statistics an individual sees influences judgments of a review's helpfulness. We find perceived helpfulness of a given review decreases as the star rating of that review gets further from the mean rating. Additionally, participants were more likely to search for reviews close to the mean. Both of these findings are consistent with confirmation bias. We explore, but do not find support for, alternative possible explanations.

Keywords: confirmation bias, search behavior, belief updating, categorization, budgeting, mental accounting.

Introduction

As the use of online shopping continues to grow, people increasingly need to form expectations of product quality by searching through product information. One factor that can significantly influence these expectations is online reviews. On many major websites people are exposed to summary statistics for all the ratings the product has received. We explore how this summary information influences the judgments they make about the reviews they then read. Specifically, we find perceived helpfulness of a given review decreases as the star rating of that review gets further from the mean rating. We also investigate why this relationship occurs and how it can influence search behavior. First, we discuss existing theory that would predict these results, as well as other patterns of theoretical importance, which we do not find support for.

Confirmation Biases

Confirmation bias is an overarching term that encompasses several psychological tendencies (Klayman, 1995). There is a large body of work on various confirmation biases, and a complete review of this literature is beyond the scope of this project. However, there are several findings from this literature that are relevant the current research.

There are two broad classes of confirmation biases. The first of these is forward-looking biases that impact how people acquire information (Klayman, 1995). This line of research finds people tend to test cases that are expected to have a property of interest more than cases not expected to have a property of interest, otherwise known as a positive-test strategy. One canonical example is the Wason selection task, in which people disproportionately look for evidence to confirm a rule as opposed to evidence that could falsify the rule (Wason, 1968). This relates to the current research because people typically observe summary information before reading reviews. If people encode that information, in our case the mean rating, as a relevant property of the product, they may disproportionately seek out reviews that share that property.

A second class of confirmation biases includes backward-looking biases that impact how information is interpreted (Klayman, 1995). People may restrict their attention and/or give greater weight to evidence that confirms a hypothesis, independent of the information they seek out (Nickerson, 1998). We find evidence for both classes of confirmation bias in participants' judgments of review helpfulness.

Belief Updating and Information Search

Relative to a Bayesian benchmark, people tend to be overly conservative in their updating of beliefs. In the classic Hogarth and Einhorn (1992) model of belief updating with sequential information, new evidence is often encoded relative to a reference point and beliefs are updated in an anchor-and-adjust manner. In relation to the present research, if people adopt the mean rating as their initial reference point and encode subsequent information relative to that mean, one's beliefs about a product will tend to be biased toward the mean rating. Coupled with the aforementioned findings that people prefer information consistent with their beliefs, this would predict greater perceived helpfulness for reviews at or close to the mean.

In problems of information search, a common finding is people tend to search too little relative to normative search models (e.g., Zwick et. al., 2003). Conditional on knowing the mean rating, reviews that deviate far from the mean could be largely informative while reviews close to the mean provide only a small, but positive, amount of new information. Thus, these models would predict helpfulness would increase as reviews strayed further from the mean.

Reasoning by Representativeness

A canonical finding from the heuristics and biases literature is individuals often form judgments based on representativeness (Tversky and Kahneman, 1974). The most well-known case of this is the "Linda problem," where participants commit a conjunction fallacy. Participants read about a hypothetical person named Linda who is described as a feminist. People then tend to say it is more likely that Linda is a feminist bank teller than a bank teller, which is impossible. Similarly, people reading product reviews may look for reviews they believe come from a representative person. This would predict people would find modal reviews to be the most helpful.

Current Research

In this work, we explore this relationship between a review's deviation from the mean rating and its helpfulness. Four studies provide support for the fact that a review's helpfulness declines as the absolute difference between the review's rating and the mean increases. We also investigated the impact of other pieces of summary information, which we believed could impact judgments of review helpfulness.

Further, we attempted to understand why deviation from the mean rating impacts review helpfulness by probing potentially mediating factors. Finally, we observe how the mean rating influences search for reviews.

In Study 1, we factorially manipulated several types of summary information, including the mean, to study the impact of summary information on helpfulness. In Study 2, we added a thought listing procedure to explore potential mediating processes that may give rise to the negative relationship between deviation from the mean and review helpfulness. In Study 3, we further investigated the mediating processes suggested by Study 2. In Study 4, we explore search for reviews as a behavioral consequence of the impact of deviation on helpfulness.

Study 1: The Impact of Summary Information on Review Helpfulness

The mean rating of a product is typically just one of several pieces of summary information people have when shopping online. Most websites also display the distribution of star ratings a product has received as well as the total number of ratings. Past research has found one's product preferences are significantly influenced by the number of total reviews a product has, a phenomenon dubbed "popularity bias" (Heck, Seiling, and Bröder, 2020; Powell et. al., 2017).

Experimental Design and Procedure

Study 1 used a 3(overall rating: 2, 3, or 4) x 2(distribution: mean is mode, mean is not mode) x 2(product: blender, book) x 2(total number of reviews: 84, 984) x 5(star rating of review: 1, 2, 3, 4, 5) between-subjects design. The products and total number of reviews were selected based on a prior norming study. Participants saw only one review and answered the following question: "How helpful would this review be when deciding whether to buy this [book/blender]?" (1 = Not helpful at all, 7 = Very helpful).

The distribution of reviews carries a lot of information about how people feel about a product. We chose to manipulate the mode as a potential alternative to the hypothesis involving the mean. If people reading the reviews are trying to gauge the experience they are most likely to have if they purchase the product, the mode seems like a sensible reference point they could adopt. This would be consistent with reasoning by representativeness. The distributions we used were constructed specifically to hold constant, within a mean rating, the percentage of reviews that were four or five stars and the percentage of reviews that were one or two stars. The reason for this was to control for binary bias (Fisher, Newman, and Dhar, 2018). This phenomenon finds people tend to group 4- and 5-star reviews together as "good" and group 1- and 2-star reviews together as "bad." Figure 1 shows this manipulation for an average rating of four stars.
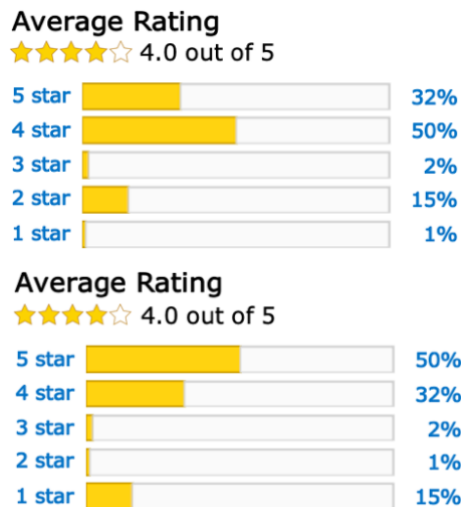

Figure 1: Distributions with 4.0 mean (top: mode is 4, bottom: mode is 5)

Participants

Three thousand six hundred and eleven participants completed the survey on Prolific (M_{age} = 35, 52% female). Eight were excluded due to a memory check failure. This left 3,603 valid completions.

Results

We ran the following ordinary least squares regression:

$$\text{Rated Helpfulness}_i = \beta_0 + \beta_1 * \text{Absolute Deviation}_i + \beta_2 * \text{Star Rating}_i + \beta_3 * \text{Book}_i + \beta_4 * \text{Total Reviews}_i + \beta_5 * \text{Mean is Mode}_i + \varepsilon_i \quad (1)$$

In Equation 1, Book is a dummy variable that takes a value of 1 for book and 0 for blender. Mean is Mode is a dummy variable that takes a value of 1 when the mean and mode of the distribution are equal and 0 when they are not. There was a significant negative relationship between a review's absolute deviation from the mean and its helpfulness rating ($\widehat{\beta_1}$ = -0.41, S.E. = 0.03, t = -14.51, p < .001).

There was a significant negative relationship between perceived helpfulness and the star rating of the review ($\widehat{\beta_2}$ = -0.17, S.E. = 0.02, t = -9.14, p < .001), which is consistent with prior literature on negativity bias in reviews (Yin, Mitra, and Zhang, 2016). Neither the total number of reviews nor the distribution significantly impacted helpfulness judgments (i.e., $\widehat{\beta_4}$ and $\widehat{\beta_5}$ were not significant). This is true in all studies, so, for brevity, we will not discuss it again. There were slight differences across the two products but the coefficient of interest, which is the effect of deviation, was similar across both. This is also true in all studies.

Discussion

The results from this study support the notion that greater deviation from the mean rating causes people to perceive reviews as less helpful. The evidence from Study 1 does not

support the hypotheses that most helpful reviews are those that contain the most new information (i.e., reviews far from the mean) or those that are most representative (i.e., the modal reviews). We continue to manipulate the mode in future studies but never observe an effect on helpfulness.

Study 2: Exploratory Thought Listing

Study 2 had two main goals. The first goal was to replicate the finding from Study 1 that greater deviation from the mean led to lower helpfulness ratings in a within-subjects design. Study 1 only presented participants with one review each because of the rich literature on assimilation and contrast effects, where the judgement or valuation of a stimulus is influenced by the judgement or valuation of a prior stimulus (see Rudolph, 1994 for a review). However, an advantage of presenting participants with several reviews is to control for individual differences in perceptions of review helpfulness. The second goal of the study was to explore the cognitive processes that give rise to this effect.

Experimental Design and Procedure

This study used a 3(overall rating: 2, 3, or 4; between) x 2(distribution: mean is mode, mean is not mode; between) x 2(product: blender, book; between) x 5(star rating of review: 1, 2, 3, 4, 5; within) mixed design. Participants read five reviews, one from each possible star rating. The order of the reviews participants saw was randomized.

After giving helpfulness judgements for all five reviews, participants engaged in a thought listing exercise. For each review, we displayed the helpfulness judgement participants previously gave and asked them to list thoughts that came to mind when forming that judgement. Participants were required to list at least four thoughts. Following the thought listing, we asked participants to code their thoughts on several dimensions in an attempt to uncover the information participants used when making helpfulness judgements and whether the information they used differed depending on a review's deviation from the mean. Thus, we asked participants to indicate, for each thought, if it was about the star rating of the review, the text of the review, the mean rating, or the distribution of ratings. Lastly, for each thought, participants indicated whether the thought was positive, negative, or neutral.

Participants

Three hundred and eighty-nine participants completed the survey on Prolific (M_{age} = 32, 57% female). One participant was excluded due to a memory check failure. This left 388 valid completions.

Results

First, we attempted to replicate the results from Study 1 by running equation 1, adding a participant-level random intercept. There was a significant negative relationship between a review's absolute deviation from the mean and its helpfulness rating ($\widehat{\beta_1}$ = -0.30, S.E. = 0.03, t = -8.73, p <

.001). Again, we see evidence of negativity bias, as low-rated reviews were more helpful ($\widehat{\beta_2}$ = -0.13, S.E. = 0.02, t = -5.97, p < .001). These are all consistent with Study 1.

We next examined whether absolute deviation from the mean influenced the thoughts that came to mind when participants formed their helpfulness judgements. For the valence of the thoughts and most of the thought coding dimensions, no such relationship emerged. However, there was an effect of absolute deviation and the total thoughts about the text of the review. We ran equation 1 with a participant-level random intercept and total thoughts about the text of a review as the dependent variable. Like perceived helpfulness, there was a significant negative relationship between a review's absolute deviation from the mean and the total number of thoughts about the text ($\widehat{\beta_1}$ = -0.045, S.E. = 0.017, t = -2.52, p = .012). No negativity bias emerged for these thoughts.

Discussion

Study 2 replicated the finding that the deviation of a review from the mean rating is negatively related to its perceived helpfulness. It also provided preliminary process evidence for why this relationship exists. In the thought listing exercise, participants generally listed more thoughts about the text of the review when the star rating was close to the mean. This finding suggests absolute deviation from the mean influenced the amount of attention participants gave to the reviews, but additional process evidence is needed to support that claim.

Study 3: Cue Weights in Helpfulness Judgments

The goal of Study 3 was to probe the process suggested from Study 2 in a more targeted manner. Specifically, because Study 2 suggested deviation from the mean can affect engagement with a review's text, we wanted to measure this by directly asking participants. When forming the helpfulness judgments for a given review participants have two cues, the star rating and the text. Thus, we test for the relative weight participants give to each cue in their helpfulness judgments.

Experimental Design and Procedure

This study used the same design as Study 2, with one modification. After the helpfulness judgements, instead of going through the thought listing exercise, participants answered the following question: "What was the most helpful part of the review?" (1 = definitely star rating, 7 = definitely text; counterbalanced). These questions were presented after the helpfulness ratings were collected.

Participants

Six hundred and thirty-two completed the survey on Prolific (M_{age} = 32, 51% female). Two participants were excluded due to a memory check failure. This left 630 valid completions.

Results

We ran equation 1 with a participant-level random intercept for two dependent variables: perceived helpfulness and the importance of text relative to star rating. There was again a significant negative relationship between a review's absolute deviation from the mean and its helpfulness rating ($\widehat{\beta_1}$ = -0.31, S.E. = 0.03, t = -12.33, p < .001). Again, lower-rated reviews were more helpful ($\widehat{\beta_2}$ = -0.14, S.E. = 0.02, t = -8.17, p < .001). We also found the importance of the text relative to the star rating was negatively related to absolute deviation from the mean ($\widehat{\beta_1}$ = -0.16, S.E. = 0.03, t = -5.81, p < .001).

Discussion

The results of Study 3 replicated the findings of Studies 1 and 2 and gave further insight into the processes underlying participants' judgments of review helpfulness. We saw the text of the review received greater weight in helpfulness judgments for reviews close to the mean. This is consistent with backward-looking confirmation bias, whereby the mean rating influences the way the information in the reviews are interpreted.

Study 4: Search for Reviews

One behavioral consequence of these results may involve the way people search for reviews. Confirmation bias can significantly impact search because people tend to look for evidence that supports a hypothesis rather than evidence that refutes it (Klayman, 1995). This study looks into the effect of deviation from the mean on search and helpfulness.

Experimental Design and Procedure

This study used a 3(overall rating: 2, 3, or 4; between) x 2(distribution: mean is mode, mean is not mode; between) x 6(product: book, painting, wine, blender, trash can, hangers; between) x 5(star rating of reviews that could be searched: 1, 2, 3, 4, 5; within) mixed design. The main difference between this design and those in prior studies is that participants chose which reviews to read (as opposed to being randomly assigned). Participants were required to search at least one review, and after that they could terminate search at any time. The maximum number of reviews they could search is five (one from each star rating). After the search part of the study, participants saw and gave helpfulness ratings to all five reviews.

Participants

Six hundred and four participants completed the survey on Prolific (M_{age} = 33, 52% female). Ten participants were excluded due to a memory check failure. This left 600 valid completions.

Results

First, we ran equation 1 with a participant-level random intercept and replicated our previous finding that helpfulness decreases with increasing absolute deviation from the mean rating ($\widehat{\beta_1}$ = -0.47, S.E. = 0.03, t = -12.12, p

< .001). There was again evidence of negativity bias ($\widehat{\beta_2}$ = -0.22, S.E. = 0.03, t = -8.47, p < .001).

To examine search behavior, we calculated the mean of all the ratings a participant searched and tested whether that differed as a function of the mean product rating. Thus, we ran equation 2:

$$\text{Mean Rating of Searched Reviews}_i = \beta_0 \\ + \beta_1 * \text{Mean Product Rating}_i + \beta_2 * \text{Book}_i \qquad (2) \\ + \beta_3 * \text{Mean is Mode}_i + \varepsilon_i + \varepsilon_i$$

We find a significant positive relationship between the mean product rating and the mean rating of the reviews a participant searched ($\widehat{\beta_1}$ = 0.26, S.E. = 0.03, t = 7.44, p < .001).

Discussion

Results from this study again replicate our finding that helpfulness is negatively related to deviation from the mean. Additionally, we find this has significant consequences for the reviews participants choose to search. The mean rating of the reviews participants searched in this study was significantly, positively related to the mean product rating participants were assigned to. In other words, participants chose to read reviews that were close to the mean rating. This is one way in which the effect of deviation from the mean on perceived review helpfulness can have a significant impact on behavior. This is consistent with forward-looking confirmation biases, as participants generally searched for confirmatory evidence.

General Discussion

Four studies provided converging evidence for a negative relationship between the perceived helpfulness of a review and its deviation from the mean product rating. This is consistent with research on confirmation biases. In addition to documenting this relationship, Studies 2 and 3 provided evidence for cognitive processes responsible. The text of a review was more important to helpfulness judgments the closer its rating was to the mean rating. Lastly, Study 4 showed evidence of participants using a positive-test strategy when searching for reviews. These findings contribute to the existing literatures on consumer reviews and confirmation biases.

Our results did not support the two other hypotheses we initially mentioned. Contrary to predictions of many normative models of search, participants did not find reviews far from the mean to be most helpful, nor did they seek out those reviews. Contrary to the prediction of reasoning by representativeness, the mode (i.e., the most representative person) was not a significant predictor of review helpfulness or search.

The total number of reviews also did not significantly influence helpfulness judgments. This was surprising given prior work on the high weight people often give to this total when choosing products (Heck, Seiling, and Bröder, 2020;

Powell et. al., 2017). However, it is certainly possible the total number of reviews influences perceptions about product quality without influencing perceptions of review helpfulness.

Theoretical Contributions

The literature on judgments of review helpfulness is growing quickly, but it is still relatively small when considering the large role reviews play in ever-expanding online shopping. There has been very limited research into the effect of summary information on review helpfulness. We add to this literature in several ways. First, we corroborate the results from a few papers that use observational data to show that deviation from the mean is negatively correlated with helpfulness. We also add to these findings by exploring the cognitive processes that underlie this relationship as well as behavioral consequences that arise.

This work also contributes to the vast literature on confirmation biases. As previously noted, confirmation bias is not a single phenomenon but rather a class of psychological predispositions whereby people seek or attend to information that confirms a hypothesis more than information that refutes a hypothesis (Klayman, 1995). We provide evidence that the mean rating for a product can lead to confirmation bias in review judgments and search.

Limitations and Future Directions

We believe this work suggests fruitful opportunities for future research on how people use product reviews. One important factor to explore, outside the scope of this paper, is one's goals or motives for reading reviews, and whether this differs based on a review's deviation from the mean. Consumer A may have a goal of gathering as much information as possible before making a purchase decision. In this case, a review will be most helpful when it adds a maximal amount of new information about the product. Consumer B may have already decided whether they are likely to purchase the product or not. In this case, the person will be more likely to gather and/or attend to information that confirms this decision (Fischer, 2011; Nickerson, 1998). These two people will likely have very different criteria for determining whether a review is helpful. Future research could explore how goal setting interacts with absolute deviation during search.

An inherent limitation in these experiments is the stylized stimuli. During online shopping outside these experiments, people see a wide array of mean ratings, distributions, and written text. While the controlled experimental paradigms are useful to study causality and underlying cognitive processes, there is always a possibility that the specific stimuli contributed to certain patterns (or lack thereof) in our data. We stimulus sample several products, but it is still an infinitesimal set compared to the set of products people buy. Future research could examine observational data across a wide array of products to explore if there are certain categories with larger or smaller effect sizes.

In addition to the results presented in this research, we believe the impact of summary review information on the way reviews impact one's beliefs and preferences is a topic that could yield further exciting research. Only the mean rating had a significant effect in our studies, but there could be different tasks or contexts where other pieces of summary information are important.

Conclusion

Confirmation bias appears to play a large role in how people use product reviews. People tend to find reviews more helpful when they are close to the mean. Additionally, people tend to search for reviews close to the mean. These findings highlight the importance of summary information in judgments and decisions regarding product reviews.

References

Fischer, P. (2011). Selective exposure, decision uncertainty, and cognitive economy: A new theoretical perspective on confirmatory information search. *Social and Personality Psychology Compass*, *5*(10), 751-762.

Fisher, M., Newman, G. E., and Dhar, R. (2018). Seeing stars: How the binary bias distorts the interpretation of customer ratings. *Journal of Consumer Research*, *45*(3), 471-489.

Hogarth, R. M., and Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, *24*(1), 1-55.

Heck, D. W., Seiling, L., and Bröder, A. (2020). The love of large numbers revisited: A coherence model of the popularity bias. *Cognition*, *195*, 104069.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, *32*, 385-418.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175-220.

Powell, D., Yu, J., DeWolf, M., and Holyoak, K. J. (2017). The love of large numbers: A popularity bias in consumer choice. *Psychological science*, *28*(10), 1432-1442.

Rudolph, A. S. (1994). *Contrast and assimilation effects: A meta-analytic review* (Doctoral dissertation).

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly journal of experimental psychology*, *20*(3), 273-281.

Yin, D., Mitra, S., and Zhang, H. (2016). Research note—When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. Information Systems Research, 27(1), 131-144.

Zwick, R., Rapoport, A., Lo, A. K. C., and Muthukrishnan, A. V. (2003). Consumer sequential search: Not enough or too much? *Marketing Science*, *22*(4), 503-519.