Marketing Science Institute Working Paper Series 2019
Report No. 19-103

Measuring the Impact of a Single Negative Consumer Review on Online Search and Purchase Decisions Through a Quasi-Natural Experiment

Marton Varga and Paulo Albuquerque

Measuring the Impact of a Single Negative Consumer Review on Online Search and Purchase Decisions through a Quasi-Natural Experiment

Marton Varga and Paulo Albuquerque*

INSEAD

November 30, 2018

Abstract

How much does one negative consumer review impact online search behavior and purchase for a product? Using a quasi-natural experiment created by how retailers update consumer-generated reviews in product pages, we quantify the impact of a single negative review on consumer online shopping activity at a multi-category retailer. The paper proposes an identification approach that compares choices from consumers who searched for a product when, on its product page, there was a one-star rating, with decisions from consumers who searched for the same product when the same one-star rating moved to a second page due to the arrival of additional reviews. This identification strategy tackles the problem of spurious correlation between review creation and unobserved demand shocks. In technology and home and garden products, we find that the detrimental impact of encountering a single negative review is two-fold: on average, the probability of continuing search to competitors increases by 10.5%, while the purchase probability of the product drops by 18.3%. We derive own-negative-review and cross-negative-review elasticities that can be used by managers to understand how consumers respond to changes in the online word-of-mouth content. Our findings are illustrated on two-dimensional product maps, resulting in a new way to evaluate the vulnerability of products to critical word-of-mouth.

*We are grateful to Wharton Customer Analytics Initiative and its partners for providing us the data. We thank the participants of the 8th Workshop on Consumer Search and Switching Cost and of the 39th Annual ISMS Marketing Science Conference for their valuable comments.

"Please do not leave a negative review. Contact us and we will do everything for your satisfaction."

An eBay Seller

"...sellers may engage in fictitious or phantom transactions with themselves or collaborators in order to artificially inflate their search results rankings." in Ali Baba's IPO Prospectus.

1 Introduction

It is common practice for online retailers to provide customer ratings and reviews in their product pages. Consumers frequently resort to reviews for input across multiple stages of the purchase decision process, such as to obtain information about the products, evaluate and compare alternatives, and reach decisions of whether to purchase and which product to buy (Mudambi and Schuff 2010). In general, customers trust reviews more than advertising (Cheong and Morrison 2008, Hung and Li 2007) and brand signaling literature suggests that customer reviews provide an additional degree of credibility to brands not obtained through company communications (Erdem and Swait 1998, Montgomery and Wernerfelt 1992).

Because of increased importance of peer-generated content, manufacturers monitor the arrival of new ratings and contact some customers directly with requests to refrain from posting negative feedback or asking to remove past ones by offering compensation. In some online marketplaces - as the quotes above illustrate - the motivation to have good ratings and reviews is so high that it can occasionally lead to the practice of brushing, where a firm engages in fabricated transactions so that fictitious buyers leave positive feedback and improve search rankings (NPR 2018). Negative reviews have become so crucial that some sellers sued buyers for causing unrepairable damage with a critical opinion (Washington Post 2012). Likewise, online retailers, who provide space for user-generated information, pay close attention to their valence and content, as they wish to guarantee high level of trust that results in positive reputation and a meaningful experience on their site. For example, Amazon filed lawsuits against more than a thousand defendants who, for about $5 per unit, allegedly created fake reviews (Forbes 2017, New York Post 2017).

Given the attention that consumers, manufacturers, and retailers pay to user-generated content, it is of significant importance to accurately measure the impact an individual online review and respective rating may have on consumer search and purchase decisions during online shopping. The overwhelming majority of related past research, however, has primarily focused on the overall effect of reviews and their average rating on sales and not on the effect of a single review or rating.[1] The seminal

[1] For an exhaustive literature review on the topic, see Babic Rosario et al. (2016) and You et al. (2015).

paper by Chevalier and Mayzlin (2006) studies the effect of the number of reviews and average rating on the sales of a books at Amazon.com and BN.com. A more recent work by Hennig-Thurau et al. (2015) investigates how Twitter posts influence box-office numbers.[2] Although these papers contain examples of how one additional review rating could impact the average rating and hence demand, the role of each individual rating and its direct effect on online search and choice is largely ignored.[3]

In terms of the consumer purchase funnel, interested buyers go beyond consulting average review ratings while visiting the product page. Previous studies (BrightLocal 2017, Chen and Xie 2008a, Brown et al. 2005, Chen and Xie 2008b, Kozinets et al. 2010, Liu 2006) have shown that consumers consider individual review ratings and content as one of the most credible inputs in their purchase decision. We choose to analyze negative reviews - usually defined as a review rating with one star out of five following the seminal work of Chevalier and Mayzlin (2006) - instead of positive ones, because negative reviews have been shown to provide greater information value than positive reviews to consumers (Ahluwalia et al. 2000, Monga and John 2008, Lee et al. 2008, Ito et al. 1998, Ivanova et al. 2013, Soroka 2006). With this in mind, the goal of this paper is to answer the following research questions:

- What is the impact of a negative rating on the purchase decision of the reviewed product and of competitor products?
- How much does one negative rating impact search behavior, i.e. does it increase or decrease the consideration set of consumers, and by how much?
- What is the level of vulnerability of each product to the arrival of a negative rating, measured by elasticities to ratings?

We answer these questions through the use of a quasi-natural experiment created by how retailers update consumer-generated reviews in product pages and a rich dataset that tracks multiple steps of the consumer buying process. Our identification strategy involves comparing consumers who visited the product page, scrolled down to its first of reviews and found a low rating there, with consumers who visited the same product page, scrolled down to its first set of reviews but did not find the low rating there. The low rating is still accessible to the latter consumers but only upon clicking on a "read more reviews" button. This exogenous quasi-natural treatment of consumers happens be-

[2] Chevalier ans Mayzlin (2006) find that an additional 1-star review increases the sales rank difference of a book between Amazon.com and BN.com. Hennig-Thurau et al. (2015) mention that around 800 fewer negative Twitter reviews would have increased revenue of a popular movie by 3.5%. Assuming linearity, this might imply that a single negative review leads to a 0.004% drop in sales. Zhu an Zhang (2010) study the role of electronic word-of-mouth (eWOM) on the sales of video games, and find that those are more influential for less popular games. Although they do not provide numerical interpretation for the magnitude of the results, in an earlier version of the paper (Zhu and Zhang 2006) they write that a one-point increase in a game's average rating leads to a 4% increase in its sales in the following month. Doing some simple calculation with their reported data, we speculate that one negative review would pull down average rating of a representative game by 0.5. Therefore, according to their estimates, a single negative review decreases sales for video-games by 2%.

[3] Wu et al. (2015) analyze how consumers learn their preferences restaurants from online reviews. In their model, consumers update their quality and cost beliefs from individual reviews but the effects of those on the consideration set is not modeled. Neither do the authors use any exogenous variation to separate the impact of quality and ratings.

cause the number of reviews on the product page is fixed and newly submitted reviews automatically take the first positions, thereby relegating older ratings to higher order review-pages.

Our approach deliberately focuses on the relegation of the negative review to the second review-page instead of its arrival as unobserved events (e.g. delivery delays or stock outs) may lead to the creation of the negative review. Such unobserved events can lower demand for the product (Reinstein and Snyder 2005) independently of the arrival of the rating. As a consequence, if we compared consumer decisions before the low rating was posted with consumer decisions right after the arrival of that rating, we would measure the impact of the low rating plus some other unknown demand or supply shock(s). By focusing on the relegation instead, we avoid this problem because the negative review existed and was available at the website both in before-relegation and after-relegation periods.

We illustrate our method with data on search and purchase decisions from two broad product categories: Technology and Home-&-Garden–at a large British online retailer. Using the described quasi-experimental approach, we quantify how much consumers shift away from a product when faced with a negative review. Our results show that a single negative review increases the probability that the consumer adds further items to the consideration set by 10.5%. In our data, this corresponds to a 3.0% expected increase in the size of the consideration set. We also find that, on average, viewing a negative review causes a 18.3% drop in the sales of the focal product, and that consumers are more likely to switch to competing, often more expensive alternatives. Based on these estimates, we derive negative-review-elasticities that range between -2% and -23%. We visualize these elasticities on product maps, which can be a useful tool for managers to gauge the impact of negative reviews on their product offering.

The remainder of the paper is organized as follows. Section 2 provides theoretical background for the study. Section 3 explains the research design including our data, identification, and econometric methodology. Section 4 describes the results, and Section 5 presents managerial implications and demand elasticities to a negative review. Section 6 concludes.

2 Theoretical Background

2.1 Conceptualization

Both theoretical (Kahneman and Tversky 1979) and empirical (Siegrist and Cvetkovich 2001, Bizer et al. 2010, Mittal et al. 1998) literature suggest that negative information has a two-fold effect: it decreases an item's perceived quality and increases the uncertainty about that quality. In economic terms, these effects imply that a consumer's expected utility of a given product drops when a negative review is encountered and its pre-purchase quality variance increases. This is justified by a dislike for

worse products, but also by risk aversion from finding an opinion of negative valence.[4]

In addition, past literature has also shown that consumers have significant search costs while gathering information about products, even in an online setting. In fact, although consumers have access to large number of reviews, they often decide to see a small percentage of the total universe of products. This limited sampling of information has been shown to be inherently related to time and energy constraint, and a large body of research has shown this cost of collecting information to be significant (Kim et al. 2010, Honka 2014, Seiler 2013).

Our research builds on the aforementioned economic-based constructs, preferences, risk aversion, and search costs, and applies them to consumer search and purchase decisions. First, we argue that even one negative review can trigger uncertainty about product quality, besides a perception of worse quality. In part, this is justified by negative information about a product being less frequent in online platforms (Hu et al. 2009, Schoenmueller et al. 2018). Hence, negative reviews provide a stronger and divergent signal, when compared to the more frequent positive reviews (Ahluwalia et al. 2000).

We assume that consumers search for products in a sequential way (e.g. Kim et al. 2016; 2010), and find strong support for this assumption in our empirical application. In a sequential search framework, consumers continue to search as long as the expected marginal benefit of searching exceeds marginal cost. In this setting, a negative rating may cause search length to increase if the additional information in a product page decreases the utility of the product and increases the variance of initially unobserved attributes such as quality. In a fixed-sample (simultaneous) search framework (e.g. De los Santos et al. 2012), where consumers decide ex-ante how many products to sample, a negative review found in a product page would not change the size of the consideration set.

In a recent and closely related paper, Liu et al. (2017) apply deep-learning models to study the impact on customer reviews on conversion. Their identification strategy relies on the variation in the reviews due to newly added reviews as those create within-product variations of review features. The authors claim that using first-differences they are able to derive the causal impact of review on purchase. However, given that the arrival of a low rating can co-occur with other unobserved shocks, we argue that comparing consumers' behavior before the arrival of a negative review with the period after the arrival leads to biased estimates. Therefore, we believe our strategy is the one that is able to identify the unbiased causal effects of reviews.

With these concepts as background, our hypotheses related to the impact of one negative review on online search and purchase are the following: we expect to find (1) a reduction of purchase probabilities of products receiving the a negative review; (2) an increase in search activity of competitive products due to the sequential nature of online search, the negative shock, and the risk aversion resulting

[4]In few cases, negative reviews were associated with a positive impact on sales (Vermuelen and Seeger 2009, Berger et al. 2010). However, in these studies, consumers tended to be initially unaware of the item and reviews were perceived as a form of publicity that raised awareness.

from finding negative information; (3) an increase in the probability of buying a substitute product; and (4) a higher price paid for that substitute, as higher prices signal better quality.

2.2 Spurious Correlation

One of the concerns when identifying the effect of online ratings on demand is the possibility that parameter estimates are biased due to spurious correlation. This happens when less favorable reviews are purely evidence of a negative demand or supply shock and a sales decline would have happened even without the existence of the review, as consumers could have found out about the negative shock from other information sources. This raises concerns about the use of methods that compare sales or search before and after new ratings arrive at the platform.

Some previous work has tackled similar challenges when measuring the reviews' effects on sales. Reinstein and Snyder (2005) warn that spurious correlation can be present between reviews and sales induced by an underlying correlation in unobservable quality signals. In their work, using a Difference-in-Differences approach – the difference between positive and negative evaluations for movies reviewed during their opening weekend and movies reviewed later on – the authors are able to isolate the influence of expert reviews from other unobservables. Chintagunta et al. (2010) also exploit the sequential nature of movie release strategies across markets. As reviews can only come from consumers where the movie was previously released, the researchers use information from these markets as instruments to measure impact of reviews at release times in other locations. Anderson and Magruder's (2012) take advantage of a managerial policy by Yelp, a local-search service powered by crowd-sourced review forum, to obtain the causal effect of the average rating on restaurant reservations. When Yelp computes the average rating of a product or service, the company rounds off to the nearest half-star. This enables the authors to propose a regression discontinuity design that takes advantage of this rounding policy.

In our setting, i.e. shopping at an online retailer, spurious correlation between review creation and sales might arise from a number of factors, and can be related to either permanent or temporary shocks. We now elaborate on and illustrate some of these sources with real-world examples of customer reviews that were submitted to retailers, and with social media messages by manufacturers.

Permanent Demand Shocks

> *We got this TV. The Video feature doesn't work as advertised. It simply freezes. We found on the manufacturers website where they said they simply stopped supporting this part of their TV.*

The manufacturer of a product present in the platform may have stopped offering maintenance and/or customer support, as exemplified by the above quote from a consumer review. Consumers can obtain

this information from other sources, such as the firm's website or its newsletter.

Temporary Demand Shocks

> *This game wouldn't play on our video console, looked on the Internet and there is a fault with them.*

The product may have a temporary malfunction, which could have been known by various consumers from other sources than our retailer. This malfunction after a while may be resolved - e.g. with a software update. Hence, a temporary demand drop is likely due to this demand shock, instead of this review.

Temporary Supply Shocks

> *To our valued customers - we're experiencing delays in the NY & NJ delivery areas and are working diligently to resolve these issues. Unfortunately, we're unable to provide any updated delivery details at the present time. Thank you for your patience!*

Temporary supply shocks can also affect demand. Manufacturers may have logistic or capacity issues that impede the satisfaction of demand. With this Twitter post by IKEA, consumers could have learned about an unpleasant supply shock. Those who read this post are less likely to order the product for at least as long as the delivery issues are not resolved. This reduction of demand could coincide with reviews that mention the same issue, hence creating a spurious effect.

3 Research Design

3.1 Empirical Setting

We have obtained from a large online retailer based in the United Kingdom a data set that tracks both consumer search and purchase decisions. The data include each individual visit between 1st of February and 31st of March of 2015 as click-stream information to all web pages from products in the Home & Garden and Technology categories. These products are further classified by the online retailer into 629 subcategories. We observe 31,284 products that have received at least one consumer visit during the analysis period, of which 4,375 have reviews.

A search session, as we define it, is unique to each consumer who is browsing for items in a product category. It starts when the consumer opens a product page and ends either with a purchase in that category or with the consumer's last observed product-related click in that category. Altogether in our dataset, 23,994 unique consumers carried out 222,682 search sessions, among which 7,774 sessions ended with purchase. During these search sessions consumers viewed a total of 410,628 product pages (including visits to the same product by different consumers). These consumers had access to 634,424 unique product reviews with a respective rating. Products are rated on a 1-to-5 discrete scale

represented by stars. 57% of the reviews had a 5-star rating, while 26% of them had a 4-star. Thus, the overwhelming majority of them were positive. 6% of the reviews had a 3-star ratings, 3% of them had a 2-star rating, and 8% had a 1-star rating. Following Chevalier and Mayzlin (2006), we classify a review as negative if that has a rating of 1-star, although we test the robustness of our findings with different definitions as well.

Price, number of reviews, and average rating of each product is visible to consumers while browsing a particular category page without opening the page of the particular product. If consumers decide to click on the product page, they see it organized into two parts: the product description and the review list and description (Figure 1). The product description section at the top of the product page is immediately visible when the product-page is opened. It repeats the information from the category-page (including brand, price, average rating, and number of reviews) and provides further details about the item, such as technical specifications, size, color, etc.

In contrast to the product description section, the review section is not immediately visible on arrival at the product page. The user must scroll down to view a first list of reviews (which we call the first review-page) that includes the five most recent reviews. After scrolling down to the bottom of the page, the consumer has the option to paginate to additional review-pages and view 20 more reviews on each page.[5] Altogether, consumers viewed the first review page 76,726 times, and paginated to the remaining review pages 12,773 times. In our data, we observe an indicator variable that captures the decision to scroll down to the reviews section and an indicator variable that measures pagination to additional review pages.[6]

In terms of browsing for products, slightly more than 60% of the consumers click on only one product page, while about 10% of them click on four or more product pages. Around one-fifth of all visits include scrolling down to the reviews section. In one of every five cases when the consumer scrolls to the reviews, he or she also paginates to the second page of reviews. In terms of time spent on the website, we observe that a visit lasts about 2 minutes on average.

In terms of product-specific information, we observe reviews, within-category market shares and page-view shares during the two months, and prices. Regarding reviews, among products with at least one review, the mean number of reviews is 55.9, while the median is 7. The average rating, provided that the product has any reviews, is 4.18, while the median rating is 4.34. Regarding market shares and page-view shares, we find strong market concentration: about 1% of items attract more than two-thirds of visits and purchases. Finally, the mean and median price are £139.2 and £50.0, respectively.[7]

[5] Consumers can also sort the reviews, for example by helpfulness or rating. However, we observe that only 3.3% of consumers who scrolled down to the review-section sort reviews thereafter by any criteria.

[6] We note that the data have the limitation that we do not observe how many of the reviews were exactly visible on a particular review-page as it might depend e.g. on the physical screen size of a particular visitor. However it is easy to see that the more consumers might have missed reading the 5th review from the first review set (on which our strategy focuses on), the more conservatives our estimates are in terms of effect size.

[7] For most products, we obtained average price directly from the data provider. We scraped archived data from https://archive.org/ to complete the data set. We were left with 1,693 products for which we could not obtain price

Figure 2 shows how price (in £), average ratings, and number of reviews of viewed products change as consumers move forward in the search process from the first product page seen in a browsing session to the tenth. There is a clear pattern of price increase as consumers move over product pages. In contrast, the number of seen reviews decreases as consumers keep browsing. Finally, the average rating of products seen early in the search process is higher than that of later products.

3.2 Identification Strategy

The unique setting of our data includes a quasi-natural experiment that helps us to quantify the effect of a 1-star rating on both search and choice. During the two months in our data set, the online retailer updated the product pages daily. These updates include the publication of product ratings and respective reviews by consumers that have purchased a product. New reviews take the first spots in the product page, while older reviews are moved further down in that page or become relegated to a review-specific page. This sequential arrival of reviews and ratings, without managerial influence,creates variation in the reviews that consumers are able to observe in different pages of the website.[8] Due to the publication of a new individual rating, the average rating of the product is also recalculated immediately. Combined with the fact that we observe whether consumers scrolled down to the review section and whether they paginated to other review-pages, we can use this quasi-natural experiment to separate consumers into treatment and control groups.

The identification of the effect of negative review relies on the comparison four groups of users that are the followings.

1. Consumers who visited the product-page, scrolled down to the review section at the bottom of the product-page, and found a 1-star rating there.
2. Consumers who visited the product-page while the same 1-star rating was accessible on the bottom of the product page but who did scroll to the review section (i.e. did not see any reviews).
3. Consumers who visited the same product page, scrolled down to its reviews, but did not find the same 1-star rating on the product page because of its relegation to a second review-page. (These consumers either decided to paginate to the second page or not.)
4. Consumers who visited the product-page while the same 1-star rating was already relegated to the second review-page but who did not scroll to the reviews section (i.e. did not see any reviews).

It is the relegation of the negative review that serves as the distinction between the treatment and the control groups. Our treatment group consists of consumers who visited the product page with a

data. In these latter cases, we use the category-average to impute the missing values.

[8] From our consultations with company representatives, we believe the firm does not manipulate the publication of the submitted reviews beyond censoring the ones with inappropriate wording, e.g. offensive language.

negative rating just before its relegation (and either read some reviews or not), while our control group consists of consumers who saw the page of the same product just after its product page was "cleaned" of the 1-star rating (and either read some reviews or not). Consequently, in our model the *products* are the treated ones – by the relegation of the bad review – and not the *consumers*. The latter are exogenously sorted to treatment and control browsing sessions depending on when they arrived to the focal product's page (i.e. before or after the relegation of the negative review). We define *focal products* as the set of products that have at least one treatment period with a corresponding control period. The rest of the products offered by the retailer constitutes the set of *other products*

In order to avoid biased estimates due to spurious correlation as described in section 2.2, we apply a Difference-in-Differences (DiD) approach (Blundell and Dias 2009). We estimate the difference in the behavior between those consumers in the treatment and control groups that did not scroll to the product reviews, and use it to control for demand changes between treatment and control periods. In other words, we first take the difference between groups (2) and (4), which we subtract from the difference between groups (1) and (3). This DiD setting enables us to obtain unbiased estimates for the impact of viewing the negative review on search and purchase outcomes.[9]

To implement our approach, we define, for each product, time intervals between rating arrivals, i.e. periods that start with the posting of a rating and end with the posting of a subsequent rating. All details related to the product in the product-page remain unchanged throughout each of these periods. Hence, a product is in its treatment period while a negative rating is on the product page, in the 5th position. A product is in its control period from the time when the negative review is relegated to a different page, and it stays in its control period as long as the product page remains absent from 1-star ratings. Although our proposed DiD strategy provides unbiased estimates for the impact of the negative at any position, we choose to focus on the clear discontinuity of having the negative review at the bottom of the first review-page vs. on the top of the second review-page. By doing so we can make sure that our approach is unrelated to the effect of the arrival of the negative review.

Figure 3 provides an illustrative example. In this case, at the beginning of the observation window, the product page contains a negative review with a 1-star at position four. Consumers arriving at this page will not yet be considered in the treatment group. Later, a review with a 5-star rating is posted, and older reviews move down; the bad review is now at position 5 of the first review-page and the treatment period begins. Still later, a 4-star review is posted and the 1-star review is relegated to the review-specific page. To observe this rating, consumers now need to paginate the review-page. At this point, the treatment period ends, and the control period starts; from now on consumers that arrive at the page but who do not paginate will not see the negative review. The control period lasts as long

[9] As long as unobserved shocks that are correlated with the arrival of the review affect consumers who read reviews similarly to consumers who do not read reviews, the DiD approach is unbiased. We believe we can safely assume this, as we see no reason why the demand or supply related disturbances (on which we elaborated in Chapter 2.2) would affect consumers differently in a systematic way.

as a new bad review does not arrive. In the figure, the end of the control period happens on the the arrival of another 1-star rating. The remaining of the time periods for this product are neither a treatment nor a control period. (In the absence of this second negative review the control period would, in this example, last until the end of our observation window.)

In our data, 24.1% of the consumers scroll to the first review-page whenever that was available - i.e. whenever the product had a nonzero quantity of reviews. In addition, only 21.2% of the consumers who have seen the first review-page will paginate to the second review-page by clicking the "next review page" button. As a result, the bad review was *hidden* from the eyes of many, including those who saw the first set of reviews. This provides us the opportunity to quantify the effect of finding a negative review while browsing. Estimates derived using the proposed strategy may be considered a lower bound to the causal effect of facing a negative review. This is because when the negative review is relegated to the second page, it is still available to the consumers, although the large majority of them will not paginate to that page.

We observe 1,136 products that have at least on treatment and a corresponding control period. These are the key products to our analysis and identification of the effect of the negative review - hence we denote them as focal products. During the observation period, these products accounted for 23.3% of sales and for 17.1% of the overall traffic (measured as number of page-views) in categories that include at least one focal item. Such categories received 80.3% of the overall traffic at the retailer.

The mean duration of the treatment periods is 10.8 days, while of the control periods is 23.0 days. There are 8,196 consumers who searched for a focal product in its treatment period, and 27,640 consumers who searched for a focal product in the control period. 22,668 visitors of the focal products saw those items in other periods.

We noticed in the data that February brought higher sales to the retailer than March. In total 3,397 products were sold in March, while 4,377 were sold in February. This still applies if we look at categories with at least one focal item - 647 pieces of focal products were sold in March, 896 in February.

3.3 Econometric Specification

In a logit framework, we model the following consumer decisions that may be affected by the presence of a negative review: (1) the purchase of the focal product that has a negative review during some treatment period; (2) subsequent browsing for competitor products as a consequence of a negative review of the focal product; (3) the purchase of a competitor product.

3.3.1 Purchase Decisions

Consumer i's utility for purchasing product j on day t is defined as

$$U_{ijt}^F = \beta_0^F + \beta_1^F T_{jt} R_{ijt} + \beta_2^F T_{jt}(1 - R_{ijt}) + \beta_3^F C_{jt} R_{ijt} + \beta_4^F C_{jt}(1 - R_{ijt}) + \beta_5^F R_{ijt} + \mu_j^F + \gamma^{F'} X_{ijt} + \epsilon_{ijt}^F. \quad (1)$$

In the equation above, the term T_{jt} takes the value of 1 if the product is one of the focal products visited in its treatment period, 0 otherwise, and C_{jt} takes the value of 1 if the product is one of the focal products visited in its control period, 0 otherwise. Thus, the baseline group consists of products that have no treatment or control period (we include all products the retailer offers). The variable R_{ijt} measures scrolling down behavior, taking the value of 1 if the consumer scrolled down to the review area in the product page, 0 otherwise. θ_j^F is a product fixed effect. The term X denotes a vector of time-varying control variables, namely (i) logarithm of product j's number of reviews + 1; (ii) average rating of j; (iii) log of the number of products consumer i has browsed for until product j including j; (iv) log of search sessions consumer i has initiated so far including the current one; (v) an indicator variable whether j has a positive number of reviews at the time of visit. We set the average rating to zero whenever a product had no reviews. ϵ_{ijt}^F is an i.i.d. unobserved component following a type I extreme value distribution.

Denoting y^F as an indicator variable for product j purchase, we have

$$y_{ijt}^F = \begin{cases} 1 & \text{if } U_{ijt}^F \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

In this formulation, $\beta_1^F - \beta_3^F$ measures the utility difference between consumers in the treatment and control groups that did scroll to the focal product's reviews. On the other hand, $\beta_2^F - \beta_4^F$ measures the utility difference between those consumers in the treatment and control groups who did not scroll to the focal product's reviews. Using the latter quantity to control for unobserved demand changes between control and treatment periods leads us to the following DiD estimator:

$$\delta^F = (\beta_1^F - \beta_3^F) - (\beta_2^F - \beta_4^F).$$

Due to the properties of logit regression (see e.g. Train 2009), δ^F is the log of odds ratio between purchasing product j having seen a recent negative review about it and purchasing product j not having seen that negative review about it. In other words, the expected change in the log odds of selling product j to consumer i due to the submission of a 1-star review is δ^F. Equivalently, the effect of a single bad review on focal product's purchase is $\exp(\delta^F)$.

The discovered negative review about product j might increase the probability that consumer i is going to purchase another the retailer offers in product j's category. By denoting y^C_{ijt} the action of choosing a competing product instead of j after product j was browsed for, we can model the impact of product j's characteristics on substitution to any competing item. In our logit model, the utility of switching away to competitors as a function of the characteristics of j then becomes

$$U^S_{ijt} = \beta^S_0 + \beta^S_1 T_{jt} S_{ijt} + \beta^R_2 T_{jt}(1 - R_{ijt}) + \beta^S_3 C_{jt} R_{ijt} + \beta^S_4 C_{jt}(1 - R_{ijt}) + \beta^S_5 R_{ijt} + \theta^F_j + \gamma^{S'} X_{ijt} + \epsilon^S_{ijt}, \quad (2)$$

where ϵ^S_{ijt} is an i.i.d. extreme value term.

The DiD estimator for the impact of a discovered 1-star review (received by j) on purchasing a competitor instead of the focal product becomes:

$$\delta^S = (\beta^S_1 - \beta^S_3) - (\beta^S_2 - \beta^S_4).$$

3.3.2 Search Decisions

In the case of search, we consider two cases. First, we model search decisions analogously to the purchase decision above. Consumers decide whether visit pages of products after visiting the page of the focal product j, i.e., whether to continue search. In this case, the dependent variable y^B_{ijt} takes a value of 1 if consumer i at time t decides to continue her search after the focal product's page, and 0 otherwise. Utility from continued browsing is

$$U^B_{ijt} = \beta^B_0 + \beta^B_1 T_{jt} R_{ijt} + \beta^B_2 T_{jt}(1 - R_{ijt}) + \beta^B_3 C_{jt} R_{ijt} + \beta^B_4 C_{jt}(1 - R_{ijt}) + \beta^B_5 R_{ijt} + \theta^F_j + \gamma^{B'} X_{ijt} + \epsilon^B_{ijt}, \quad (3)$$

where ϵ^B_{ijt} is an i.i.d. extreme value random variable,

while

$$y^B_{ijt} = \begin{cases} 1 & \text{if } U^B_{ijt} \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Alternatively, we can measure the reviews' impact on further browsing by the duration of search after observing the focal product's page. In this case, the dependent variable will be $\log D_{ijt}$, the log-

arithm of one plus the number of product pages visited after product j was viewed. Here, given that the dependent variable is continuous, we use ordinary least squares to estimate the respective parameters.

$$\log D_{ijt} = \beta_0^D + \beta_1^D T_{jt} R_{ijt} + \beta_2^D T_{jt}(1 - R_{ijt}) + \beta_3^D C_{jt} R_{ijt} + \beta_4^D C_{jt}(1 - R_{ijt}) + \beta_5^D R_{ijt} + \theta_j^F + \gamma^{D'} X_{ijt} + \nu_{ijt}, \quad (4)$$

where ν_{ijt} is an unobserved random variable following the standard normal distribution.

3.4 Estimation

Denote

$$\mu_{ijt}^z(\beta^z, \gamma^z, \theta^z) = \beta_0^z + \beta_1^z T_{jt} R_{ijt} + \beta_2^z T_{jt}(1 - R_{ijt}) + \beta_3^z C_{jt} R_{ijt} + \beta_4^z C_{jt}(1 - R_{ijt}) + \beta_5^z R_{ijt} + \mu_j^z + \gamma^{z'} X_{ijt} \text{ for } z \in \{F, S, B, D\}.$$

To estimate the parameters for the binary purchase and search decisions, we maximize the following log-likelihood functions for $z \in \{F, C, B\}$:

$$LL^z(\beta^z, \gamma^z, \theta^z) = \frac{1}{2ITJ} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} y_{ijt}^z \log[Pr(y_{ijt}^z = 1)] + (1 - y_{ijt}^z) \log[1 - Pr(y_{ijt}^z = 1)], \quad (5)$$

where $y_{ijt}^z = 1$ if consumer i takes the respective action and 0 otherwise, and

$$Pr(y_{ijt}^z = 1) = \frac{1}{1 + e^{-\mu_{ijt}(\beta^z, \gamma^z)}}.$$

For $z = D$, we find the maximum of the likelihood for the Gaussian family

$$LL^D(\beta^D, \gamma^D, \theta^D) = \frac{1}{2ITJ} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} (\mu_{ijt}^D - y_{ijt}^D)^2. \quad (6)$$

For the estimation of each of the above models, we take advantage of the H20 package interface, a scalable open source machine learning platform that offers parallelized implementations of various machine learning algorithms, including Generalized Linear Models (Nykodym et al. 2016).[10] Given the

[10] For more information visit http://docs.h2o.ai/ or http://docs.h2o.ai/h2o/latest-stable/h2o-docs/

size of the data and especially the number of covariates (more than 30,000 product fixed-effects), we found that traditional maximum likelihood estimation was slow and demanded significantly more computational resources than a regular personal computer. The H20 approach is efficient because, unlike traditional methods, it runs maximum likelihood estimation via iteratively reweighed least squares (Burrus 2012). Consequently, we decided to run our models trough connecting to the H2O cluster. Standard errors are obtained by bootstrapping samples of consumers at the product level.[11]

To further increase performance, we use Revolution R Enterprise, a big-data analytics solution using R language.[12]

4 Results

4.1 Model-Free Evidence

This section provides descriptive evidence of our hypotheses by looking at the average behavior of consumers who searched for the focal products in the treatment periods, when a negative review was accessible upon scrolling down, and in their control periods, when it was not. Table 1 shows how frequently on average consumers (i) purchased the focal item, (ii) purchased a competitor, (iii) searched for more alternatives following seeing the page of the focal product, and also (iv) the number of products searched after the focal one. These statistics are broken down by treatment status (i.e. whether the visit occurs in a treatment of a control period, other periods not shown) and by scrolling decision (i.e. whether the consumer saw the first review-page of the focal item or not). Altogether, we observe 2,034 scrolls in treatments periods (6,162 page visits without scrolls) and 7,198 scrolls in control periods (20,442 page visits without scrolls). In the table, we indicate the change in the behavior between treatment and control periods, and respective significance.

The descriptive statistics show no significant difference between purchase rates of the focal products in the treatment periods and in the control periods among consumers who did not scroll, neither between the purchase rate of competitors. These insignificant difference in sales between treatment and control periods are not surprising any more if we take into account the general sales pattern of the retailer. As noted in Section 3.1, the retailer experienced a general sales decline from February to March. As 56.1% of the visits during treatment periods take place in February, while only 26.3% of the control periods do so, this pattern is naturally reflected in these statistics. Notice, however, that both focal products and competitors have significantly higher purchase rates among scrolling con-

data-science/glm.html.

[11] We do not use any penalty in the likelihood function that would shrink some parameter estimates towards zero and/or limit the number of covariates (alias regularization penalty). Without these penalties, the likelihood functions in Equation 5 and 6 do not change and the standard econometric, full specifications are maintained (Friedman et al. 2001). Hence, all coefficients are estimated, including the various product fixed effects.

[12] A recent version is available at https://www.microsoft.com/en-us/download/details.aspx?id=51204; accessed on July 19th, 2018

sumers in the treatment periods. As these consumers did not see the reviews, these numbers clearly reflect the general sales decline. Consequently, the descriptive evidence suggests that *not* encountering any 1-star rating is powerful enough to counterbalance the dropping global trend in sales, at least in terms of the focal products.

To control for unobserved trends and shocks, our strategy is to difference out the behavior of those consumers who did not see the reviews. The last column of the table, thus, simply shows the difference between the respective values in the "Difference" column and illustrates our DiD strategy. These numbers (although come without statistical inference) suggest that those who encounter a negative rating on the first review-page are less likely to purchase the focal product and are also less likely to purchase a competitor. Furthermore, viewing a negative review increases the average propensity to search for alternatives, and the size of the overall consideration set.

4.2 Regression Coefficients

Going beyond the model-free evidence, we estimate the proposed model that controls for product-specific effects and other observable covariates. We run separate estimations for our four dependent variables: (1) purchase of a product; (2) purchase of other products other than the focal one; (3) continuation of search; (4) log of the number of product page browsed after the focal product plus one. We include all consumers in our data and add product fixed effects to each specification as described in Chapter 3.3.

The estimated coefficients are collected in Table 2. These tell us that consumers who scroll to the reviews section have a larger consideration set than those who do not scroll and are also more likely to purchase an item. Earlier discovered products are more likely to be bought as we find a negative coefficient for the number of products searched until the focal one. Returning consumers - i.e. those who have previously bought a product in the same category - make a purchase at the website with a higher likelihood and with a shorter search as shown by the positive coefficients of the number of search sessions. The marginal impacts of the average rating is generally insignificant, indicating that these are largely captured by the product fixed effects. More reviews are correlated with fewer purchases, which rather reflects a falling demand trend than preference for a low number of reviews as we have found sales to be generally higher at earlier dates in our data.

Table 3 lists the estimated δ coefficients alongside their standard errors and implied elasticities.[13]

[13] Elasticities are calculated for the average consumer who reads the first reviews of a focal product during its control period. Denote $m \in M$ as the index for consumer-time-product combinations that match this group description, and T_m as an indicator variable of value 1 if at the time of search m a negative review was present at the product page, 0 otherwise. The average of the predicted probability for response $y^z = 1$ in this group is

$$P_C^z = \frac{1}{M}\sum_{m=1}^{M} \Pr(y_m^z = 1 | T_m = 0) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{1 + e^{-\mu_m^z}}$$

for $z \in \{F, S, B\}$. Had these group members seen the negative review (which they could not as at the time of their search it was not present at the product page), the average of the predicted probability for response $y^z = 1$ in this group

These show a significant negative impact of 1-star ratings on purchase likelihood, implying a decrease in the probability of focal product's purchase of 18.3%. On the other hand, a negative rating has no significant impact on the purchase of competing products in our data. The latter result suggests that although some consumers might be willing to substitute the focal product with (potentially more expensive, later browsed) items, a large fraction of them might decide not to buy anything from the retailer once a bad review is encountered.

Regarding further search, positive DiD estimators reveal that consumers are about 10.5% more likely to continue their search for other products after encountering a negative review. In that case, consumers are also expected to visit 3.0% more additional product pages. Overall, both search-related dependent variables indicate an increased browsing effort when consumers view a low rating during online shopping.

4.3 Robustness Checks

To investigate the robustness of our findings, we rerun the analysis described in the previous chapter using alternative treatment definitions. When performing these exercises, we only change one piece at a time, while keeping the rest of the main model unchanged. For parsimony, in Table 4 we present only the estimated DiD coefficients (δ) from these alternative specifications.

In the *first alternative* specification, periods in which the negative review is on the 4th position are also considered as treatment periods, while control periods are unchanged. The effects of a negative review on purchase and further search decisions are similar to the ones found in the preceding chapter. A single negative review decreases the probability of purchasing the focal product, but leaves the probability of purchasing competitors unchanged. It significantly increases further browsing, just as we found previously.

In our *second alternative*, we restrict the control period to last only as long as the 1-star review is on the first position of the second review-page. This shrinks the number of control period product page visits to 6,578. This change could be the reason why the DiD focal purchase estimator is not significant in this case - although it has the expected sign with an implied elasticity comparable to the ones we have seen before. Our search coefficients, however, are robust to this specification as well.

In the *third alternative*, we define a negative review as one that has a no larger rating than a 2-star, while in the *fourth alternative* as one that has a no larger rating than a 3-star. The estimated

would become

$$P_T^z = \frac{1}{M}\sum_{m=1}^{M} \Pr(y_m^z = 1 | T_m = 1) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{1 + e^{-\mu_m^z - \delta^z}}.$$

The elasticities are then obtained by

$$E^z = \frac{P_T^z}{P_C^z} - 1,$$

for $z \in \{F, S, B\}$. For $z = D$, the elasticity is simply δ^D due to the log specification in the OLS.

focal purchase coefficient is significant and negative in the latter specification (insignificant negative in the former), while the substitute purchase coefficient is significantly positive in the former case. Search coefficients are unanimously positive and significant.

In our final *fifth alternative*, we let the 1-star review to occupy any position on the first review-page, but we delay the start of the treatment period by 3 days. Although the start of the treatment period is nonetheless closer here to the arrival of the negative review than in the base model, a few days might be enough for temporary shocks to dissipate. The four significant estimates from this specification support all our hypotheses.

With a variety of robustness checks then, we have been able to show that negative reviews are likely to cause a considerable drop in the purchase likelihood of the focal product. These checks also suggest that one critical opinion could affect the sales of competitors, although this evidence is not conclusive in our data. Furthermore, we have consistently found across all specification that a single negative rating is enough to trigger an intensified browsing behavior. Due bad reviews, consumers are more likely to search for further items in the hope of finding a better alternative than the focal one.

4.4 Effect of a Negative Rating on Price Paid

In this section, we discuss the possibility that consumers end up paying a higher price after facing a 1-star rating in a product page visited. If a negative rating increases the sense of risk aversion, the higher price of competitors of the product with a negative rating may be perceived as a signal of better quality and consequently a safer purchase option (Rao 2005). Thus, those consumers who decide not to abandon the website without purchase following the discovery of a negative review about the focal product might "flee to safety" but could in fact, spend more.

To empirically investigate this question, we run several regressions explaining the log of the price paid for the purchased item. The sample for this specification consists of consumers who made a purchase at the end of their search session, which leaves us with 5,564 consumers who viewed 16,290 items - among which 1,715 focal - in total (in the main specification). We keep the same model covariates and robustness specifications as before, and illustrate our findings in Table 5 by displaying the resulted DiD coefficients in a similar fashion to Table 4.

We find a significantly positive DiD estimator in our main specification, indicating that the price paid is larger - on average by 9.6% - among those who have seen a negative review about the focal product. This is the result of people switching away from the focal item towards often more expensive competitors. Despite the considerably smaller sample size than before, robustness checks 3 and 4 support the increased paid price hypothesis.

Altogether, results suggest that consumers are willing to pay a close to 10% premium for a competing product due to a discovered negative review, provided they have found a suitable alternative on

the retailer's website.

5 Elasticities to Negative Reviews

In this section, we describe the variability of responsiveness to negative reviews across products. To do so, we use the estimates of our model to approximate a drop in sales in GBP-terms due to a 1-star rating. Given this variation in sales, we can compute the elasticity-to-negative-review for each product, a measure analogous to price elasticity. We then show examples of product maps based these elasticity measures and probabilities to search induced by the low rating.

Variation in products' elasticity to negative review comes from consumers' heterogeneous search behavior. For each of them, we measure how the likelihood of continued search after viewing the focal item changes due to the negative review. From these, we infer the changes in consideration sets. Then, by predicting the eventual choice under the base scenario with the negative review on the first page and the counterfactual scenario with the negative review on the second, we are able to obtain own-negative-review and cross-negative-review elasticities. These indicate how sensitive the focal product is to its own reviews, and how vulnerable other products are to the focal product's reviews.

In order to derive the elasticities, we apply the following strategy. We take all consumers who visited the focal product during its treatment period, regardless whether they scrolled to its reviews. (For the products that have more than one treatment periods, here we take the chronologically first one.) Using our regression coefficients from 2, we calculate the log odds of searching for additional products after viewing the focal item

$$L^S_{ijt} = \log\left(\frac{Pr(y^S_{ijt} = 1)}{Pr(y^S_{ijt} = 0)}\right)$$

and the log odds of purchasing the focal item

$$L^F_{ijt} = \log\left(\frac{Pr(y^F_{ijt} = 1)}{Pr(y^F_{ijt} = 0)}\right)$$

for each consumer.

To obtain probabilities under the counterfactual scenario where no negative review is present on the first review-page of the focal product, we take these log odds values and correct them by subtracting our DiD estimates. Only after this correction do we transfer the log odds values to probabilities the following way:

$$Pr^C(S_{ijt} = 1|T_{ijt} = 1) = \exp(L^S_{ijt} + \delta^S)$$

$$Pr^C(F_{ijt} = 1|T_{ijt} = 1) = \exp(L^F_{ijt} + \delta^F).$$

This ensures that, for those who saw the negative review, the probability of further search is lower while the probability of focal purchase is higher if the bad review is *not* displayed on the first review-page. (For those who did not scroll, we do not correct the values.)

The probability of purchasing a product searched before the focal is the same under the base and counterfactual scenarios. However, the probability of purchasing a product viewed after the focal one is lower under the counterfactual scenario. This is because the product's probability of being searched is lower. For a given consumer under the counterfactual, the probability of purchase drops equally for all items sampled after the focal one. However, the magnitude of this drop is specific to each consumer. Once we have calculated the changes in the purchase probabilities for everyone in the sample who visited the focal product during its treatment period, and for each product (searched before focal, focal, searched after focal), we predict the expected changes in sales due to the relegation of the focal product's negative review. Own-negative-review elasticity is then calculated as the percentage change in the market share of the focal product under base and counterfactual scenarios, multiplied by -1.

On Figure 4, we display the own-negative-review elasticities of focal products in the most visited Home-&-Garden categories. The box-plots reveal that sales are expected to drop by 10-15% on average. However, there is a significant dispersion across products. We observe that e.g. curtains are sensitive to negative reviews, while wardrobes are impacted less. If we look at the most popular technology categories (Figure 5), we among others see that headphones tend to have larger elasticities than televisions.

One way managers can apply our methodology to gain insight about how products the market responds to negative review is to visualize these reactions on a map. Figure 6 shows one such map, with own-negative-review elasticity on the horizontal axis and induced search probability on the vertical one for tech products, and another map for Home-&-Garden products. Each point represents a focal item. Maps are divided into four quadrant by vertical and horizontal lines, representing median values. The bottom-right quadrant includes products with low elasticity and low induced search. The sales of these items does not drop much if a negative review appears on their page. Furthermore, consumers encountering a bad review about them are not prone to search for further products. Thus, this is the most desirable quarter for a product to be located. On the other hand, the top-left corner is the least desirable for managers, as products here are at risk of loosing market share due to negative posts and are also likely to face with more de facto alternatives. As the color of the points in the top-quarters indicates, popular products consumers often view tend to be sensitive to critical opinion in terms of search behavior in the sense that a negative review induces consumers to consider further alternatives.

On Figure 7, we plot a similar chart but for one product category - kettles - only. Capital letters

denote the brand (made anonymous). We differentiate between national brands, premium brands and private labels. Notice that the majority of premium brands tend to appear close to the bottom-right quarter of the map, the segment with the smallest response to negative reviews. This matches the view that demand for premium brands is less elastic than for normal goods (Reddy et al. 2009). Our example shows that the demand of premium brands might remain inelastic when elasticity is measured to negative reviews instead of to price, although not necessarily without exceptions. On the other hand, private labels tend to appear on the upper/left side of the map, suggesting that these products could be especially vulnerable to the critical opinions posted on the retailer's website.

One might prefer a different categorization along other attributes - e.g. durability, energy consumption or comfort. With our method, managers can classify their products (and their competitors) according to the elasticity of the items to negative reviews, in order to find the attributes that vulnerable products possess. This can give valuable insight about where to intervene, for instance by trying to avoid the arrival of a low rating or by changing the characteristics of selected products.

Furthermore, we calculate the competitive clout (Kamakura and Russell 1989) of the focal products over their category alternatives. To do so, we simply take the sum of squared cross-review elasticities of competitors and divide by the total number of alternatives in the category. The range of the competitive clout values informs us about market structure. If it is small, no focal product in the category has a power over its rivals. On the other hand, if it is large, the distribution of power among focal products is unequal. We find that this power distribution is especially heterogeneous e.g. among blenders and laptops, while it is quite concentrated among e.g. curtains and SIM free phones. (Figure A.1 and A.2).

Finally, we dedicate some words to fake reviews. On average, a negative review stays on the product page for around 14 days, meaning, roughly, that it affects search and purchase over the following 2 weeks in case of a typical product. We calculate sales of the retailer's products during our 2-month observation window in order to gauge the 2-week sales in general. We find that most products were not sold at all, and that the top 10% best-sellers had 2-week sales in the range between £5-£875. Using our estimates and the information that only around 1 out of 5 people see the reviews, we speculate that a negative review would cause a £0.1-£30 sales drop among the best-sellers over the next two weeks, with a median decline of less than a pound. From further speculation and disregarding any legal or reputational consequences, if a manager of a well-established product was about to buy 5 fake good review to relegate a recent true bad review to the second page, paying more than £5 for those would presumably not be a good investment (especially not for other-than-bestseller products). Still, according to the New York Post (2017) the price of 5 fake good reviews at Amazon is $25. As a consequence, we surmise that fake reviews are overpriced in addition of being hazardous and unethical.

6 Conclusion

Although a variety of previous studies investigated the impact of eWOM on product sales, we found that authors often elaborate on the direction of it and much less on its magnitude. We also could not find any research that would have explicitly focused on quantifying the effect of a single negative review on purchase. Based on existing work, one could best speculate about the magnitude of this latter effect so far.

In this paper, we pinpoint this magnitude among technology and home-&-garden products. We take advantage of a quasi-natural experiment arising from the "newest first" display policy by the retailer to obtain our results. We compare users who searched for the product when the negative review was among the first reviews shown with users who searched for the same product when the same review was among the second set of reviews shown. By looking at where the review is displayed instead of whether the review is posted, our estimates do not suffer from the problem of endogeneity of review valence with other demand factors.

Our results reveal that a single negative review encountered by the consumer decreases purchase probability by 18.3%, on average. We also analyze how consumer search behavior changes after reading a critical opinion. We find that if that happens, the probability of further search for alternative items rises by 10.5%.

Building on these estimates we predict changes in the consideration sets and derive unique elasticity-to-review values for a variety of products. We find that among all consumers considering the product - the ones who read its reviews plus the ones who did not read them - the elasticity of sales to a single bad review is between -2% and -23%.[14]

The proposed methodology can be of particular interest for product managers. Using it, one can gain further insight into the competitive market. We suggest to visualize products on a map that shows how sensitive each of them is to negative eWOM in terms of sales and alternatives considered. Platform managers might also be able to gain some insights about how different review display policies could influence the firm's revenue. The free availability of a machine learning package makes our approach feasible to run on a personal computer despite the large number of covariates that grow together with the number of products.

Future work might address how managers could shape their platform such that they gain the most from it. For instance, a different set of displayed reviews might influence what consumers look at, and also what they eventually purchase. However, a change in platform design might also induce a change in consumer behavior (e.g. more or less review readings). How the introduction of an alternative display policy would modify consumers' review reading practices and perception about the retailer is an important question we leave for future research. Another potentially fruitful direction to study would

[14] The elasticities we derived are then larger than the ones we calculated from the tables of previous papers.

be, following Moe and Schweidel (2012), to estimate whether and how low ratings influence the arrival frequency of new reviews through decreased purchase rates.

References

Ahluwalia, R., R. E. Burnkrant, and H. R. Unnava (2000). Consumer response to negative publicity: The moderating role of commitment. *Journal of Marketing Research 37(2)*, 203–214.

Anderson, M. and J. Magruder (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal 122*(563), 957–989.

Babić Rosario, A., F. Sotgiu, K. De Valck, and T. H. Bijmolt (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research 53*(3), 297–318.

Berger, J., A. T. Sorensen, and S. J. Rasmussen (2010). Positive effects on negative publicity: When negative reviews increase sales. *Marketing Science 29(5)*, 815–827.

Bizer, G., J. T. Larsen, and R. E. Petty (2010). Exploring the valence-framing effect: Negative framing enhances attitude strength. *Political Psychology 32(1)*, 59–80.

Blundell, R. and M. C. Dias (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources 44*(3), 565–640.

BrightLocal (2017). Local consumer review survey 2017.

Brown, T. J., T. E. Barry, P. A. Dacin, and R. F. Gunst (2005). Spreading the word: Investigating antecedents of consumers' positive word-of-mouth intentions and behaviors in a retailing context. *Journal of the Academy of Marketing Science 33*(2), 123–138.

Burrus, C. S. (2012). Iterative reweighted least squares. *OpenStax CNX. Available online: http://cnx.org/contents/92b90377-2b34-49e4-b26f-7fe572db78a1 12.*

Chen, Y. and J. Xie (2008a). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science 54(3)*, 477–491.

Chen, Y. and J. Xie (2008b). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science 54*(3), 477–491.

Cheong, H. J. and M. A. Morrison (2008). Consumers' reliance on product information and recommendations found in ugc. *Journal of Interactive Advertising 8*(2), 38–49.

Chevalier, J. A. and D. Mayzlin (2006). The effects of word of mouth on sales: Online book reviews. *Journal of Marketing Research 43(3)*, 345–354.

Chintagunta, P. K., S. Gopinath, and S. 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)*, 944–957.

De los Santos, B., A. Hortacsu, and M. R. Wildenbeest (2012). Search model of consumer search using data on web browsing and purchasing behavior. *The American Economic Review 102(6)*, 2955–2980.

Erdem, T. and J. Swait (1998). Brand equity as a signaling phenomenon. *Journal of consumer Psychology 7*(2), 131–157.

Forbes (2017). Amazon's fake review problem is now worse than ever, study suggests.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics New York.

Hennig-Thurau, T., C. Wiertz, and F. Feldhaus (2015). Does twitter matter? the impact of microblogging word of mouth on consumers' adoption of new movies. *Journal of the Academy of Marketing Science 43*(3), 375–394.

Honka, E. (2014). Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics 45*(4), 847–884.

Hu, N., J. Zhang, and P. A. Pavlou (2009). Overcoming the j-shaped distribution of product reviews. *Communications of the ACM 52*(10), 144–147.

Hung, K. H. and S. Y. Li (2007). The influence of ewom on virtual consumer communities: Social capital, consumer learning, and behavioral outcomes. *Journal of advertising research 47*(4), 485–495.

Ito, T. A., J. T. Larsen, N. K. Smith, and J. 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.

Ivanova, O., M. Scholz, and V. Dorner (2013). Does amazon scare off customers? the effect of negative spotlight reviews on purchase intention. *Wirtschaftsinformatik Proceedings*.

Kahneman, D. and A. Tversky (1979). Prospoect theory: An analysis of decision under risk. *Econometrica 47(2)*, 263–292.

Kamakura, W. A. and G. J. Russell (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of marketing research*, 379–390.

Kim, J. B., P. Albequerque, and B. J. Bronnenberg (2010). Online demand under limited consumer search. *Marketing Science 29(6)*, 1001–1023.

Kim, J. B., P. Albequerque, and B. J. Bronnenberg (2016). The probit choice model under sequential search with an application to online retailing. *Management Science*.

Kozinets, R. V., K. de Valck, A. C. Wojnicki, and S. J. Wilner (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing 74(2)*, 71–89.

Lee, J., D.-H. Park, and I. Han (2008). The effect of negative online consumer reviews in product attitude: An information processing view. *Electronic Commerce Research and Applications 7(3)*, 341–352.

Liu, X., D. Lee, and K. Srinivasan (2017). Large scale cross category analysis of consumer review content on sales conversion leveraging deep learning. *NET Institute Working Paper No. 16-09.*.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing 70*(3), 74–89.

Mittal, V., T. R. William, and M. B. Patrick (1998). The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing 62(1)*, 33–47.

Moe, W. W. and D. A. Schweidel (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science 31*(3), 372–386.

Monga, A. B. and D. R. John (2008). When does negative brand publicity hurt? the moderating influence of analytic versus holistic thinking. *Journal of Consumer Psychology 18(4)*, 320–332.

Montgomery, C. A. and B. Wernerfelt (1992). Risk reduction and umbrella branding. *Journal of Business*, 31–50.

Mudambi, S. M. and D. Schuff (2010). What makes a helpful online review? a study of customer reviews on amazon.com. *MIS Quarterly 34(1)*, 185–200.

NPR (2018). A series of mysterious packages.

Nykodym, T., T. Kraljevic, N. Hussami, A. Rao, and A. Wang (2016). Generalized linear modeling with h2o. *Published by H2O. ai, Inc.*

Post, N. Y. (2017). Scammers elude amazon crackdown on fake reviews with new tricks.

Post, W. (2012). Judge says homeowner must delete some accusations on yelp, angie's list.

Rao, A. R. (2005). The quality of price as a quality cue. *Journal of marketing research 42*(4), 401–405.

Reddy, M., N. Terblanche, L. Pitt, and M. Parent (2009). How far can luxury brands travel? avoiding the pitfalls of luxury brand extension. *Business Horizons 52*(2), 187–197.

Reinstein, D. A. and C. M. Snyder (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The journal of industrial economics 53*(1), 27–51.

Schoenmueller, V., O. Netzer, and F. Stahl (2018). The extreme distribution of online reviews: Prevalence, drivers and implications.

Seiler, S. (2013). The impact of search costs on consumer behavior: A dynamic approach. *Quantitative Marketing and Economics 11*(2), 155–203.

Siegrist, M. and G. Cvetkovich (2001). Better negative than positive? evidence of a bias for negative information about possible health dangers. *Risk Analysis 21(1)*, 199–206.

Soroka, S. N. (2006). Good news and bad news: Asymmetric rresponse to economic information. *The Journal of Politics 68(2)*, 372–385.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

Vermeulen, I. E. and D. Seegers (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management 30*(1), 123–127.

Wu, C., H. Che, T. Cha, and X. Lu (2015). The economic value of online reviews. *Marketing Science 5*(34), 739–754.

You, Y., G. G. Vadakkepatt, and A. M. Joshi (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing 79*(2), 19–39.

Zhu, F. and X. Zhang (2006). The influence of online consumer reviews on the demand for experience goods: The case of video games. *ICIS 2006 Proceedings*, 25.

Zhu, F. and X. Zhang (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing 74*(2), 133–148.

Figure 1: The product page and the review pages

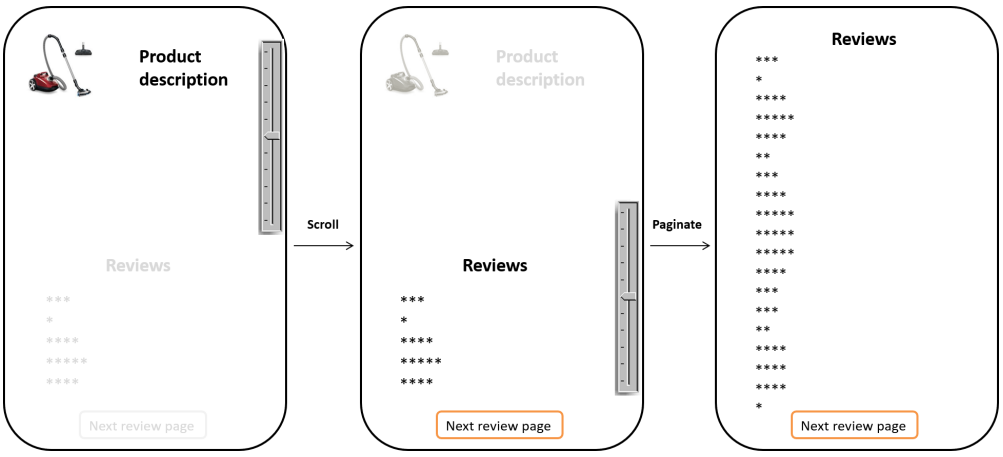
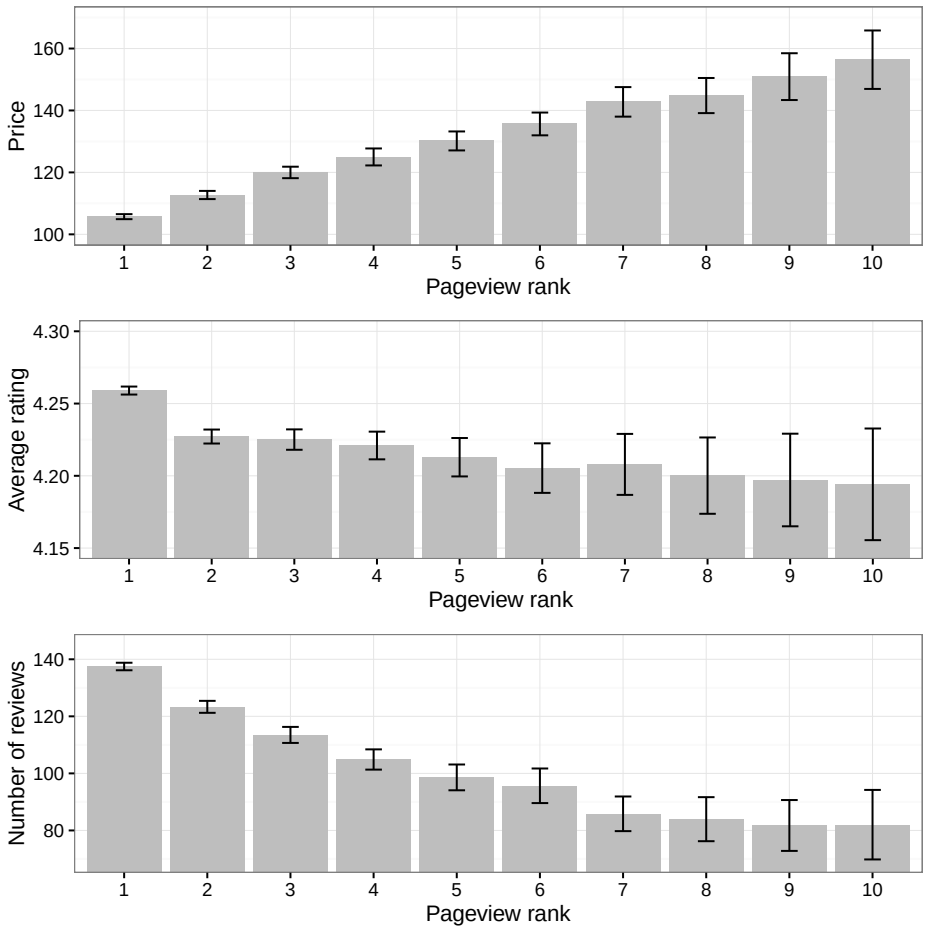


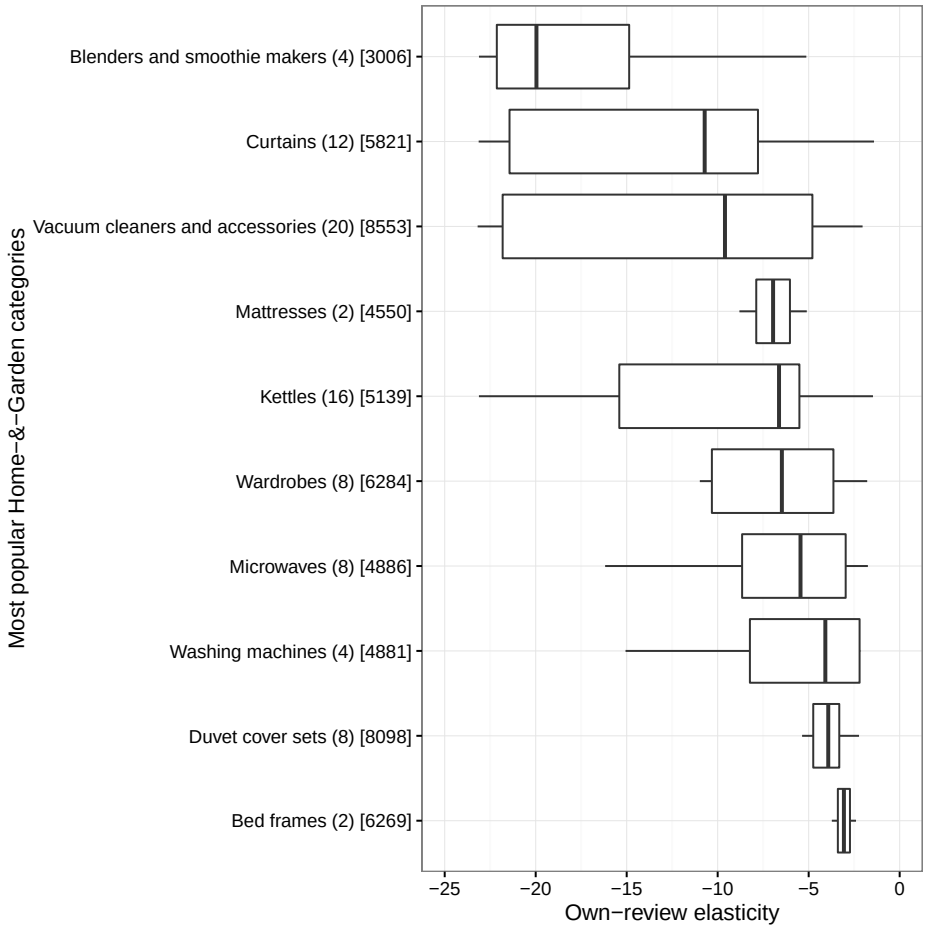
Figure 2: Mean of product characteristics as search progresses



Note: Error bars represent 95% confidence interval.

Figure 3: Treatment and control periods

Figure 4: Review elasticities - Gardening



Note: Number of focal products in parentheses, number of total visits in brackets

Figure 5: Review elasticities - Tech

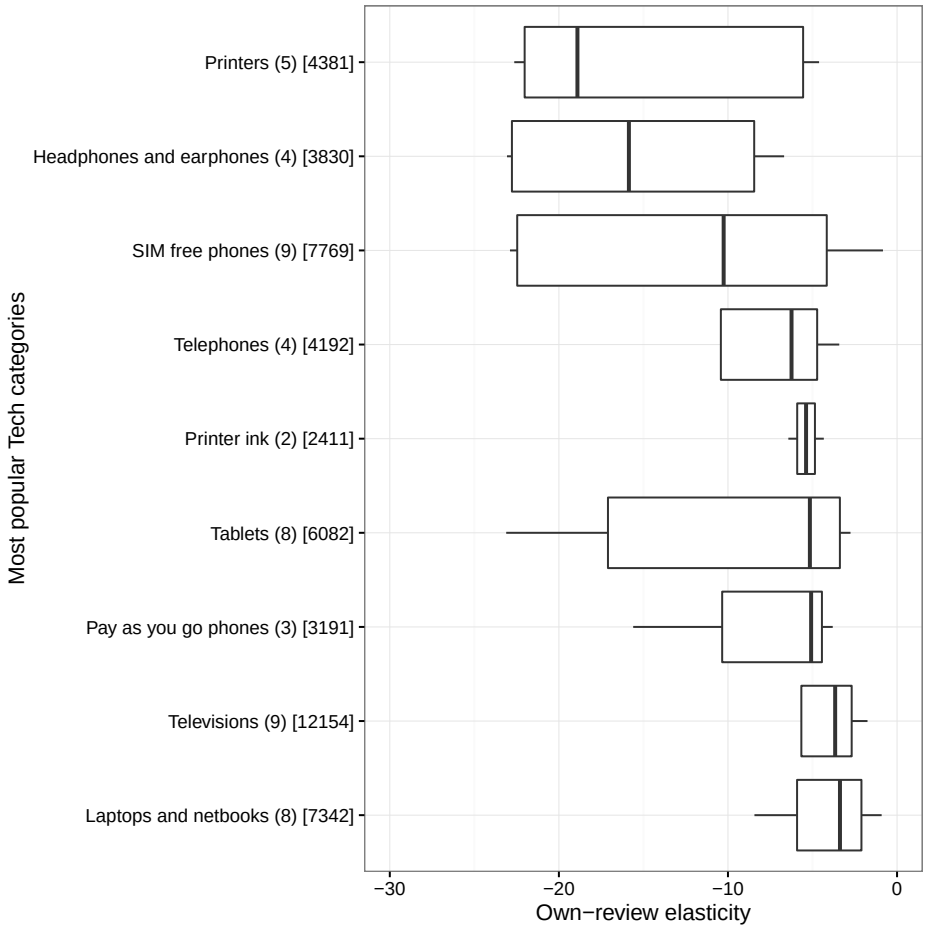
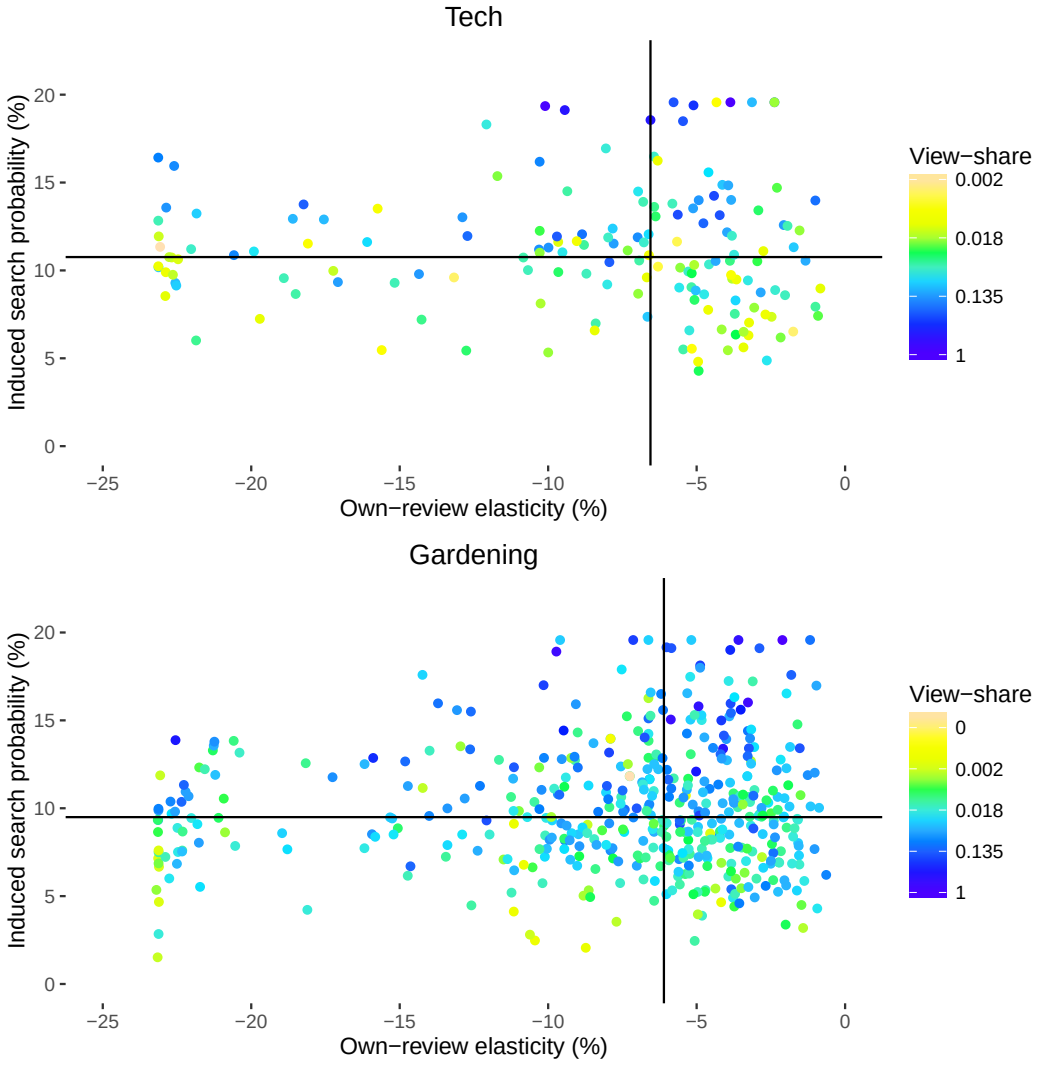
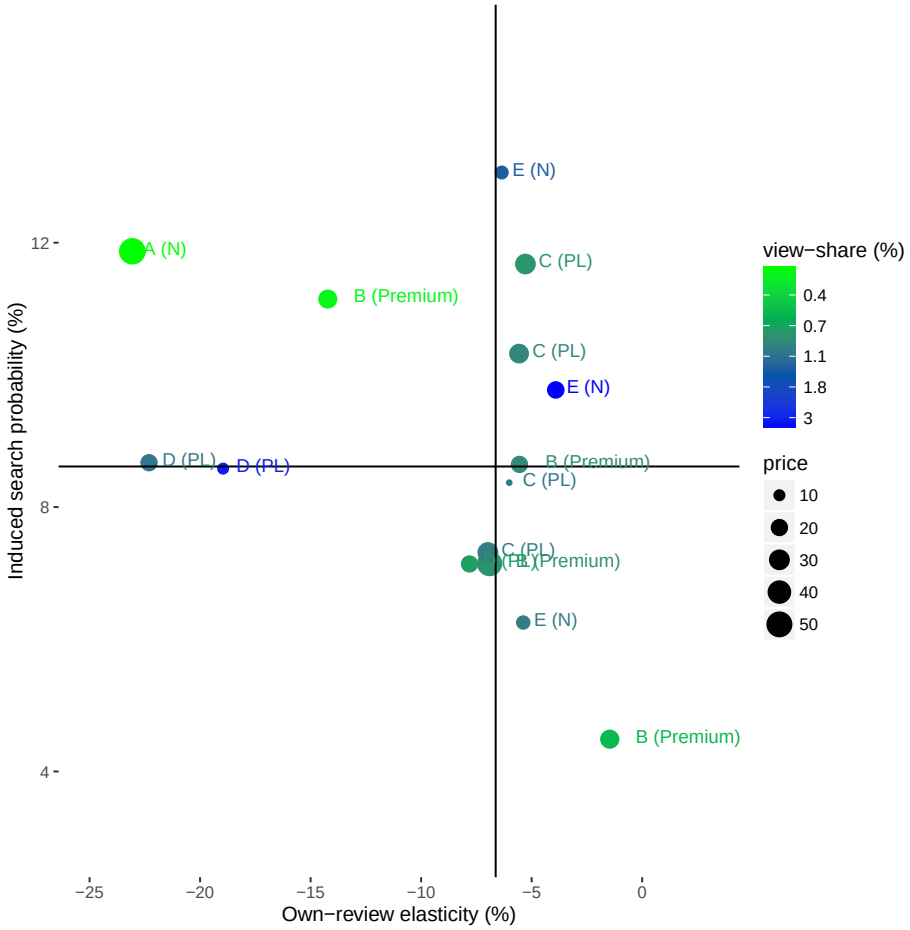


Figure 6: Induced search and review elasticities



Note: Number of focal products in parentheses, number of total visits in brackets

Figure 7: Kettles



Note: Letters denote brands. (PL=Private label, N=National brand, Premium=Premium brand)

Table 1: Average shopping behavior of visitors

Purchase	Scroll	Treatment	Control	Difference	t-stat	DiD
Buys focal product	Yes	0.026	0.025	0.001	0.193	-0.010
	No	0.034	0.024	0.011	3.998	
Buys substitute	Yes	0.024	0.021	0.003	0.831	-0.005
	No	0.024	0.016	0.008	3.715	
Search after focal product	Yes	0.484	0.465	0.019	1.503	0.040
	No	0.385	0.406	-0.021	2.968	
# Products searched after focal	Yes	1.076	1.142	0.034	1.143	0.142
	No	0.876	0.984	-0.108	4.135	

Table 2: Regression estimates

	Focal purchase	Other purchase	Search after focal	Log products searched after focal
Intercept	-2.980	-4.288	-0.145	0.538
	(0.532)	(0.286)	(0.330)	(0.086)
Scroll	0.073	0.041	0.260	0.070
	(0.017)	(0.022)	(0.010)	(0.002)
Treatment*Scroll	-0.057	-0.029	0.005	-0.027
	(0.091)	(0.121)	(0.057)	(0.012)
Control*Scroll	-0.183	-0.104	-0.105	-0.039
	(0.060)	(0.070)	(0.034)	(0.008)
Treatment*NoScroll	0.239	-0.014	-0.213	-0.070
	(0.074)	(0.078)	(0.034)	(0.007)
Control*NoScroll	-0.096	-0.243	-0.145	-0.049
	(0.049)	(0.051)	(0.028)	(0.006)
Log of products searched until focal	-0.217	0.426	0.482	0.190
	(0.019)	(0.027)	(0.006)	(0.001)
Log of search sessions	2.684	2.181	-0.336	-0.075
	(0.088	(0.098)	(0.061)	(0.015)
Log of number of reviews	-0.101	-0.148	-0.071	-0.031
	(0.032)	(0.030)	(0.017)	(0.004)
Average rating	0.039	0.036	-0.008	0.000
	(0.041)	(0.051)	(0.027)	(0.007)
Has reviews	-0.000	0.022	-0.010	0.000
	(0.056)	(0.056)	(0.030)	(0.010)
Product fixed effects	Yes	Yes	Yes	Yes
Observations	410,628	410,628	410,628	410,628

Note: Logit models for binary decisions, OLS for the Log of products searched after focal. Bootstrapped standard error, clustered at product level in parentheses. Products that have no treatment or control periods serve as reference group.

Table 3: Estimated DiD coefficients

	$\delta = (\beta_1 - \beta_3) - (\beta_2 - \beta_4)$	
Decision	Coefficient	Elasticity
Buys focal product	**-0.209** (0.115)	**-18.3%**
Buys substitute	-0.153 (0.143)	-13.9%
Search after focal product	**0.178** (0.070)	**10.5%**
Log of # products searched after focal	**0.034** (0.012)	**3.0%**

Note: Bootstrapped standard errors, clustered at product level in parentheses. Estimates in bold are significant at 5% (one-tailed).

Table 4: DiD estimators from alternative definitions

	(I) Buys focal product	(II) Buys substitute	(III) Search after focal product	(IV) Log of # products searched after focal
Alternative 1: Negative review is at position 4 or 5				
Coefficient	**-0.289**	0.004	**0.148**	**0.032**
	(0.090)	(0.091)	(0.053)	(0.012)
Elasticity	**-24.6%**	3.9%	**8.1%**	**3.2%**
Alternative 2: Control period's review is at the first position only				
Coefficient	-0.172	-0.038	**0.157**	**0.033**
	(0.134)	(0.118)	(0.076)	(0.019)
Elasticity	-15.4%	-3.6%	**8.6%**	**3.3%**
Alternative 3: Negative review has a 1-star or 2-star rating				
Coefficient	-0.077	**0.197**	**0.165**	**0.027**
	(0.085)	(0.118)	(0.057)	(0.012)
Elasticity	-7.2%	**21.2%**	**9.0%**	**2.7%**
Alternative 4: Negative review has a 1, 2, or 3-star rating				
Coefficient	**-0.238**	0.038	**0.175**	**0.027**
	(0.076)	(0.100)	(0.042)	(0.009)
Elasticity	**-20.7%**	3.7%	**9.8%**	**2.7%**
Alternative 5: Negative review is at any position with a 3-day delay				
Coefficient	**-0.139**	**0.199**	**0.127**	**0.037**
	(0.073)	(0.093)	(0.046)	(0.011)
Elasticity	**-12.6%**	**21.5%**	**7.1%**	**3.7%**

Note: Bootstrapped standard errors, clustered at product level in parentheses. Estimates in bold are significant at 5% (one-tailed).

Table 5: Regression on price paid

Main specification

Coefficient	S.e.	Elasticity
0.092	0.054	**9.6%**

Alternative 1: Negative review is at position 4 or 5

Coefficient	S.e.	Elasticity
0.006	0.051	6.0%

Alternative 2: Control period's review is at the first position only

Coefficient	S.e.	Elasticity
0.055	0.076	5.6%

Alternative 3: Negative review has a 1-star or 2-star rating

Coefficient	S.e.	Elasticity
0.102	0.042	**10.7%**

Alternative 4: Negative review has a 1, 2, or 3-star rating

Coefficient	S.e.	Elasticity
0.135	0.036	**14.4%**

Alternative 5: Negative review is at any position with a 3-day delay

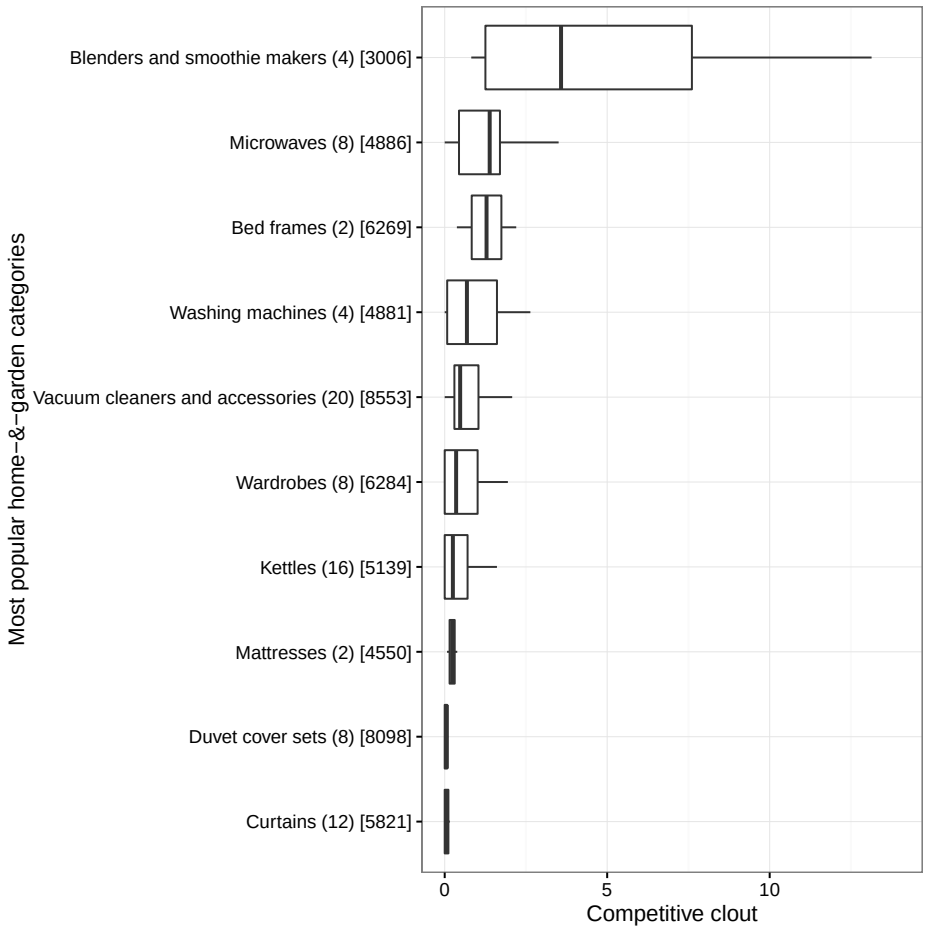
Coefficient	S.e.	Elasticity
-0.000	0.043	0.02%

Note: Bootstrap standard errors in parentheses, clustered at product level.

Estimates in bold are significant at 5% (one-tailed).

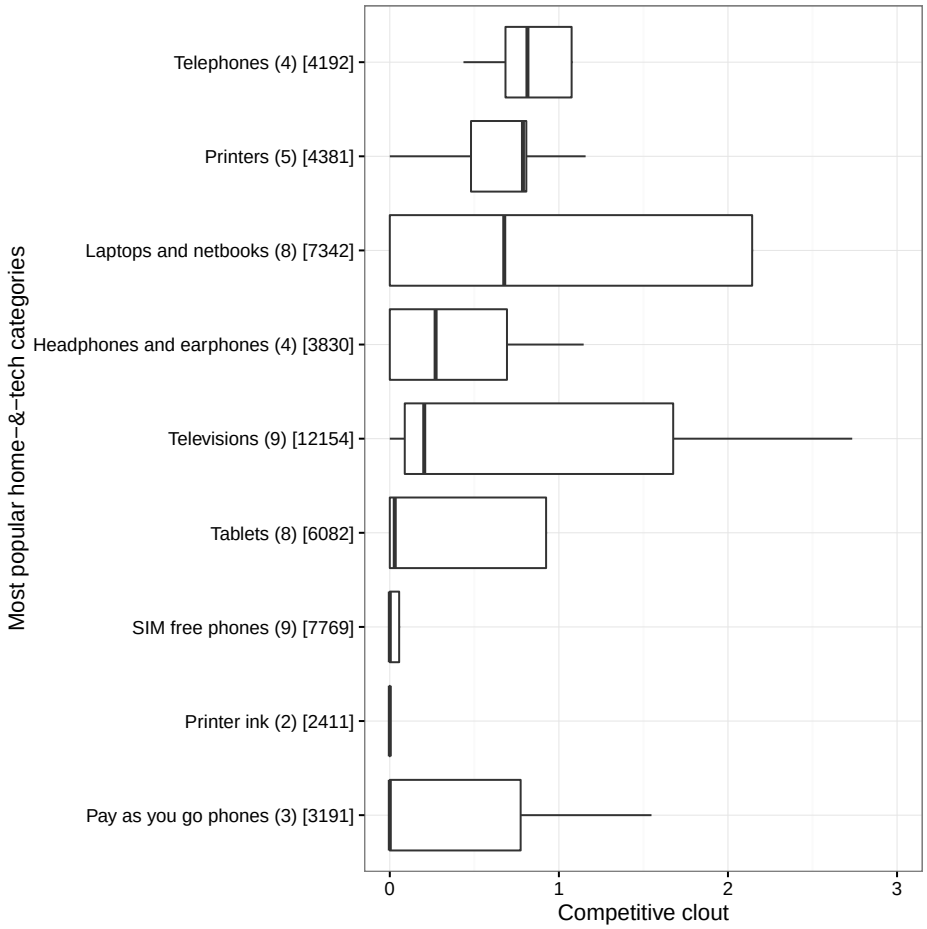
Appendix

Figure A.1: Competitive clout - Gardening



Note: Number of focal products in parentheses, number of total visits in brackets

Figure A.2: Competitive clout - Tech



Note: Number of focal products in parentheses, number of total visits in brackets