

Winter, K., Zapf, B., Hütter, M., Tichy, N., & Sassenberg, K. (2021). Selective exposure in action: Do visitors of product evaluation portals select reviews in a biased manner? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 15(1), article 4. <https://doi.org/10.5817/CP2021-1-4>

Selective Exposure in Action: Do Visitors of Product Evaluation Portals Select Reviews in a Biased Manner?

Kevin Winter¹, Birka Zapf^{1,2}, Mandy Hütter², Nicolas Tichy³, & Kai Sassenberg^{1,2}

¹ Leibniz-Institut für Wissensmedien, Tübingen, Germany

² University of Tübingen, Tübingen, Germany

³ Ludwig Maximilian University of Munich, Munich, Germany

Abstract

Most people in industrialized countries regularly purchase products online. Consumers often rely on previous customers' reviews to make purchasing decisions. The current research investigates whether potential online customers select these reviews in a biased way and whether typical interface properties of product evaluation portals foster biased selection. Based on selective exposure research, potential online customers should have a bias towards selecting positive reviews when they have an initial preference for a product. We tested this prediction across five studies (total N = 1376) while manipulating several typical properties of the review selection interface that should – according to earlier findings – facilitate biased selection. Across all studies, we found some evidence for a bias in favor of selecting positive reviews, but the aggregated effect was non-significant in an internal meta-analysis. Contrary to our hypothesis and not replicating previous research, none of the interface properties that were assumed to increase biased selection led to the predicted effects. Overall, the current research suggests that biased information selection, which has regularly been found in many other contexts, only plays a minor role in online review selection. Thus, there is no need to fear that product evaluation portals elicit biased impressions about products among consumers due to selective exposure.

Keywords: Product evaluation portals; customer reviews; information selection; selective exposure

Introduction

Purchasing products online has become routine for many people. According to a worldwide market survey from 2012, eight out of ten customers regularly buy items on the Internet (Fretwell et al., 2013). In 2019, 14.1% of retail sales worldwide fell under electronic commerce (Clement, 2019). Most of the online shops and product evaluation portals allow customers to rate and review products. As such, online customers have access to the experiences of earlier customers. Indeed, online customers consult reviews as the main source of information and prefer them over recommendations from friends and family (Fretwell et al., 2013). Scientific evidence further suggests that online customers make use of reviews to inform their decisions (Chen, 2008; Hu et al., 2008; Hu et al., 2014; Senecal & Nantel, 2004).

Large online shopping platforms feature product evaluation portals, where the number of reviews for a single product frequently goes beyond the number of reviews a customer would be interested in reading. Therefore, the interfaces of these platforms are designed to allow for the selection of reviews. To get as much information as possible from those reviews, customers should select them in an unbiased way. For instance, they should choose to read an equal number of positive and negative reviews. The literature on information selection (or selective

exposure) suggests that this might not be the case. In fact, an overwhelming body of research has demonstrated that people prefer information that is in line with their preferences over contradicting information (for a meta-analysis, see Hart et al., 2009). The current research investigates whether selective exposure also occurs in the context of review selection on product evaluation portals. In addition, we test whether certain typical interface properties of product evaluation portals boost selective exposure in line with the literature. Thus, the current research experimentally tests the validity of a well-established effect (i.e., selective exposure) in an applied setting (i.e., product evaluation portals), thereby, making a contribution that is highly informative for both theory and practice.

Selective Exposure in the Context of Product Evaluation Portals

Prior research shows that Internet users frequently select information in a biased manner. For instance, when searching for health information on the Internet, people's behavior is driven by their emotional state (Becker et al., 2018; Greving & Sassenberg, 2015, 2018) or interindividual differences like personality traits (Zhang et al., 2017). More importantly in the current context, users of online recommender systems (Buder & Schwind, 2012; Schwind et al., 2012; Schwind & Buder, 2012) and consumers of (online) media coverage (Knobloch-Westerwick & Kleinman, 2012; Lueders et al., 2019) are often biased towards selecting information that corresponds with their initial view on a given topic. This tendency to select or preferably process information confirming one's own attitude is called confirmation bias (Nickerson, 1998) and is well-established in basic psychological research (for a meta-analysis, see Hart et al., 2009). There, it has also been shown that biased information selection can have detrimental effects on subsequent decision making (Greitemeyer & Schulz-Hardt, 2003; Kray & Galinsky, 2003). In the context of product evaluation portals, relying on an unbalanced subset of all the available information could impair the decisions people make – for instance, purchasing an unsatisfactory product.

Following the selective exposure literature, confirmatory tendencies arise from the motivation to defend one's own view in light of incoming information (Chaiken et al., 1989; Fischer & Greitemeyer, 2010; Hart et al., 2009). In order to avoid the cognitive dissonance that may result from contradicting information, people preferably select information that bolsters their initial preferences. Within the classical selective exposure paradigm, participants have to make an initial decision (e.g., for a job candidate) and later have the opportunity to select information that either confirms or contradicts their initial preference (Greitemeyer & Schulz-Hardt, 2003; Jonas et al., 2005; Jonas et al., 2008). While the phenomenon of selective exposure itself seems suitable to the context of product evaluation portals, visitors of these portals usually do not need to make an explicit initial decision in favor of a product. Therefore, the paradigms in selective exposure research and the conditions during the use of product evaluation portals differ systematically.

Still, it is likely that visitors of product evaluation portals form an initial preference based on immediately available product information. This is in line with the finding that having an initial preference is not necessary for selective exposure to occur, but that developing a preference based on external cues suffices (Russo et al., 1996). In this regard, the typical interface properties of product evaluation portals might play a crucial role for the occurrence of selective exposure effects. In the following sections, we will outline how these interface properties contribute to the development of an initial preference and potentially also foster selective exposure.

The Role of Typical Interface Properties

When visiting product evaluation portals, people are usually interested in reading customer reviews to inform their purchasing decisions. In this context, there are two likely candidates as sources of selective exposure. First, people might visit the product evaluation portal with a clear goal to buy a specific product (e.g., buying a particular toothbrush). These visitors have a strong initial preference for a product (i.e., a positive initial attitude) and should according to selective exposure literature prefer positive customer reviews (Hart et al., 2009). Second, people might not have a clear preference up front, but want to inform themselves and potentially buy a product of a particular category (e.g., any toothbrush). In this case, visitors should still form an initial attitude towards a specific product based on its perceived quality, which could bias their review selection.

Product evaluation portals typically provide visitors with plenty of information to gain a first impression and to draw conclusions regarding product quality. First and foremost, an *overall rating* of the product (e.g., depicted in

the form of stars) calculated from the individual ratings of previous customers provides direct information about the product quality and is commonly presented in a prominent position (Hu et al., 2014). In addition, an *image of the product* and a *textual description* of product features are often accessible at first sight. There is a large body of research showing that similar cues influence perceived product quality and, in turn, purchasing behavior (e.g., Jacoby et al., 1971; Szybillo & Jacoby, 1974; van Ooijen et al., 2017). In the current case, an attractive physical appearance (provided through the image) and desirable product features (provided through the description) suggest the product is high-quality and when combined with a positive overall rating, visitors should form a positive initial attitude. In contrast, an unattractive product image or undesirable product features combined with a negative overall rating should lead to visitors forming a negative initial attitude. Therefore, low-quality products should be irrelevant for the purchasing decision and not cause selective exposure. Visitors of product evaluation portals usually perceive product quality cues (i.e., image, description, overall rating) before selecting any customer reviews. The overall rating should be especially important as it is the most direct information on product quality and probably has the biggest influence on the initial attitude. Accordingly, if the overall rating is only presented after selecting reviews, this should not influence visitors' selection behavior. However, if high product quality can be derived from the available information before selecting reviews, this should lead visitors to preferably select positive customer reviews.

H1: Visitors of product evaluation portals will select predominantly positive customer reviews when exploring a high-quality product.

Other interface properties typically inherent in the procedure of selecting products and reviews might further strengthen biased review selection. One source of selective exposure on product evaluation portals is that visitors can *freely choose* which product to explore. This feature is inherent in most product evaluation portals, given that visitors cannot be forced to look at a particular product. Free choice comes closest to the initial preference that is often asked for in selective exposure studies and reflects a positive initial attitude towards the selected product, because it usually means that one considers buying said product. Deliberately deciding to explore a particular product should increase the commitment to this initial choice as a means of post-decisional dissonance reduction (Russo et al., 1996; Svenson, 1992). This can also enhance feelings of ownership towards the chosen product (Abelson, 1988; Hart et al., 2009). Commitment to an initial choice, in turn, increases the tendency to defend one's decision and, thus, to show selective exposure (Chaiken et al., 1989; Fischer & Greitemeyer, 2010; Hart et al., 2009). Therefore, we predict that the free choice (vs. assignment) of a product for further exploration facilitates selective exposure to predominantly positive customer reviews.

H2a: Free choice (vs. assignment) of a product should lead to the selection of more positive customer reviews.

Another potential driver of selective exposure on product evaluation portals is the *procedure* used to select and read customer reviews. Visitors of these environments typically select and read one customer review before selecting and reading the next piece of information. In selective exposure research, this procedure is referred to as sequential selection (Jonas et al., 2001). A blocked selection procedure – where one first selects all pieces of information at once and then reads all of them in a second step – is more frequently used in selective exposure studies (e.g., Jonas & Frey, 2003; Schulz-Hardt et al., 2000), but rarely occurs in product evaluation portals or other real-life contexts (Jonas et al., 2001). Importantly, sequential selection increases selective exposure compared to blocked procedures (Jonas et al., 2001). Under typical conditions of product evaluation portals, selective exposure is, thus, more likely. Therefore, we predict that under conditions of free product choice the sequential (vs. blocked) selection should amplify the selection of positive reviews.

H2b: Sequential (vs. blocked) selection of customer reviews for a freely chosen product should lead to the selection of more positive reviews.

Overview of the Current Research

In the current research, we investigate whether the selection of reviews from product evaluation portals is biased towards the selection of positive reviews when exploring a high-quality product (Hypothesis 1). In addition, we test the effects of two typical interface design properties on this biased selection tendency: free choice (vs. assignment) of a product (Hypothesis 2a), and sequential (vs. blocked) selection (Hypothesis 2b).

To test Hypothesis 1, we provided participants with a positive *overall rating* of the product by previous customers as well as an image and product information that suggested the product was *high-quality* in each study. In Studies 1a and 1b, we tested whether the presentation of a positive overall rating before (vs. after) review selection would lead to the selection of more positive reviews. In Study 2, we explicitly compared whether the presentation of a high-quality (vs. low-quality) product including the corresponding overall rating would lead to more positive reviews being selected. In both these cases, support for Hypothesis 1 would be shown by a difference between the experimental conditions, given that we would not expect a bias to occur when the overall rating is only presented after selection (i.e., limited information on product quality) or when participants are told to explore a low-quality product (i.e., low interest in exploring the product).

Studies 3 and 4 were designed to test both Hypotheses 1 and 2a/b. Here, we established factors that should lead to a positive bias across all conditions (i.e., presentation of a positive overall rating for a high-quality product). Accordingly, a bias towards selecting positive reviews (compared to the midpoint of the scale), in line with Hypothesis 1, should occur across all conditions. In addition, the typical (vs. alternative) interface design properties assumed to foster positive review selection were implemented as experimental manipulations in these studies: free choice (vs. assignment) of a product (Study 3; Hypothesis 2a), and sequential (vs. blocked) selection on top of free choice (Study 4; Hypothesis 2b).

We report all manipulations and measures included in the studies as well as all exclusion criteria. Furthermore, for every study, we state whether including or excluding participants who fulfill our exclusion criteria alters the results. Across all studies, we excluded statistical outliers based on their studentized deleted residuals (SDR) in a linear multiple regression of the main dependent variable (i.e., average selected reviews) on the main predictors (i.e., factors of the respective design). Participants with an absolute SDR > 2.65 were treated as statistical outliers and, thus, excluded from further analyses as is common in our lab (e.g., Ditrich et al., 2019; Scholl & Sassenberg, 2014; based on Neter et al., 1996). Additional preregistered exclusion criteria are reported for each individual study.

Study 1a

Method

Design and Participants

Study 1a was an online experiment with one between-groups factor (presentation of overall rating: before vs. after selecting reviews). To initially test our hypothesis, we aimed at collecting data from $N = 150$ participants. Assuming a medium effect size ($d = .50$) in an independent two-sample t -test (two-sided) with $\alpha = .05$, this would result in statistical power of $(1-\beta) > .80$. One-hundred-and-fifty-eight participants completed our study in response to a call sent via the university's mailing list. As this initial study was not preregistered, we did not apply any exclusion criteria apart from the outlier analysis described above. This led to the exclusion of two participants, but including them did not change the results reported below. Thus, the final sample consisted of $N = 156$ participants (106 female, age: $M = 23.65$ years, $SD = 3.44$, range: 18-38). As a reward, participants were offered the chance to take part in a lottery of 10 vouchers for an online shopping portal each worth 10 Euros.

Procedure

First, participants were introduced to a scenario in which they imagined buying a gift for their grandfather with a pooled family fund via an online shopping portal. The family fund was chosen to ensure financial issues were less influential when evaluating the product. Buying the product for someone else allowed the review selection to be relevant, in case participants themselves were not interested in the particular product. Then, participants saw an image of an electric toothbrush and some basic product information. While presenting the toothbrush for the first time, we manipulated the experimental condition. That is, in one condition, the overall rating of the product from ostensible previous customers was presented on one screen *before* participants selected the reviews for this product. In the other condition, this overall rating was only presented *after* reviews had already been selected. The design of the overall rating was adapted to match widely used product evaluation portals, thus, displaying up to five stars (with five stars being the best possible rating). The overall rating of the toothbrush was very positive (i.e.,

4.3 stars based on 340 customers). Thus, we expected participants to form a positive initial attitude towards the product when seeing the overall rating before review selection (but not to the same extent when seeing it after the review selection).

On the next page, participants were asked to select eight out of twenty reviews for further reading. They were informed that this was not a representative composition of reviews but rather sought to display different opinions on the product. Accordingly, all five rating categories (i.e., one to five stars) were covered by four reviews each. The reviews were randomly arranged in two columns with ten rows each. Every single review was presented in a box and contained only the rating of the customer plus the title of the review fitting the rating (e.g., "hands off" or "absolute recommendation"). The text of the review was not visible as it was covered by grey bars. Thus, participants had to make their selection before being able to read the complete reviews.

Then, participants received the eight reviews they selected and were asked to read them (for examples see Figure 1). After a self-paced reading period, participants were asked to give their own rating of the product and to indicate their certainty regarding this evaluation. In the next step, all participants (i.e., irrespective of whether they had already seen it before or not) received information about how previous customers rated the product and indicated how surprising this rating was to them. Next, participants responded to some measures not central to the current research purpose (i.e., review selection, evaluation of a second product, and memory performance¹) and indicated their demographic information. Finally, they were thanked, debriefed, and directed to the lottery.

Figure 1. *Examples of Customer Reviews Used in Studies 1a and 1b With the Lowest (First Review) and Highest (Second Review) Ratings.*



Please keep your hands off it

My old toothbrush conked out after 6 years and with this model, I wanted to go at full steam into the modern times. Fiddlesticks! The only innovation is the display for the actual battery condition. Awesome. The battery is after a short time perceptibly dead. The customer service is a cheek, I haven't seen anything like that in a long time. Concerning the color, the thing is very far off the track. But the worst thing is the condition of my gum. Please learn from my mistakes. This has absolutely nothing to do with price-performance.



Absolute recommendation

I criticize the quality of the battery, which after all perceptibly lost performance after 2 years. Instead the electrical toothbrush convinced me with a long endurance per full charge. Second, I enjoy the range of clean programs, which you don't have with simpler models. Since I use this toothbrush, my boyfriend said, that my teeth clearly brightened up. Due to environmental protection considerations, I think that with an electrical toothbrush with quality you're better off than constantly disposing of single-use toothbrushes: Of course, the toothbrush is not being gifted, but in my opinion, it has an excellent price-performance ratio.

Measures

Review Selection. Our main dependent measure was the average rating of the selected reviews. Given that eight reviews had to be selected and that each rating category was present four times, the average rating of the selected reviews could in principal range from 1.5 (i.e., selecting four reviews with one star and four reviews with two stars) to 4.5 stars (i.e., selecting four reviews with four stars and four reviews with five stars).

Product Evaluation. The product evaluation was assessed with one item ("How do you rate this product?") using the same five-stars scale provided by the customer reviews and the overall rating; more stars represented a better rating. The certainty of this rating was assessed on the same page with one item ("How certain are you regarding

this rating?”) using a graphical slider allowing for integers between 0 and 100 with five labelled scores (0 = *uncertain*, 25 = *rather uncertain*, 50 = *mixed*, 75 = *rather certain*, 100 = *very certain*). The starting position of the slider was at the scale midpoint (i.e., 50).

Surprise. Participants were asked how surprising they found the overall rating of the product that ostensibly stemmed from previous customers. This question was included to check whether participants were aware of the overall rating in the *before* condition. We provided a similar slider to the one used for the certainty ratings, however, only using two verbal anchors (i.e., 0 = *not at all*, 100 = *very much*).

Results

Review Selection

We tested Hypothesis 1 that predominantly positive reviews would be selected when exploring a high-quality product. Crucially, participants only had an obvious quality cue available at the time of review selection when the positive overall rating was presented before (vs. after) review selection. Thus, we compared the two conditions with an independent two-sample *t*-test. Against our hypothesis, participants in the two conditions did not differ regarding the average rating of the reviews they chose, $d = 0.01$, 95% CI [-0.31, 0.33], $t(154) = 0.06$, $p = .950$. The descriptive statistics of the selected reviews as a function of experimental condition across all studies are displayed in Table 1.

As we did not observe any differences between the conditions, we further investigated participants' overall tendency of selecting reviews with a one-sample *t*-test against the midpoint of the scale (i.e., 3). There was a small but significant difference from the midpoint for the whole sample, $d = 0.31$, 95% CI [0.15, 0.47], $t(155) = 3.91$, $p < .001$. That is, on average, participants showed a tendency towards selecting positive reviews ($M = 3.11$, $SD = 0.35$) in line with Hypothesis 1. Looking at the distribution of selected reviews revealed that participants considered reviews from all categories. On average, they chose 1.37 reviews with one star, 1.57 reviews with two stars, 1.49 reviews with three stars, 1.96 reviews with four stars, and 1.62 reviews with five stars.

Additional Analyses.

An independent two-sample *t*-test revealed no significant difference between the experimental conditions regarding participants' surprise towards the overall rating, $d = -0.31$, 95% CI [-0.62, 0.01], $t(154) = -1.91$, $p = .057$. Participants who had already seen the overall rating *before* they selected reviews ($M = 37.38$, $SD = 26.91$) were only tendentially less surprised by the overall rating compared to those who only saw it *after* reading the reviews ($M = 45.87$, $SD = 27.77$). Likewise, regarding both product evaluation and certainty no differences between conditions could be observed, $t_s < 0.23$, $p_s > .828$. This indicates that participants either did not recognize the initial overall rating or that this did not change their initial perception of the product. Finally, the average rating of the selected reviews predicted the subjective evaluation of the product, $r(156) = .34$, $p < .001$. That is, the more positive the selected reviews were on average, the more positive the final evaluation of the product was.

Discussion

Our initial test of our hypothesis that people select predominantly positive customer reviews when viewing a high-quality product provided mixed results. Presenting a positive overall rating before (vs. after) review selection did not alter participants' behavior, which is against our expectation. However, we observed a small positive deviation from the midpoint of the scale across all participants, which favors Hypothesis 1, but questions our assumption that the overall rating is a necessary feature to provoke this bias. Rather the results suggest that the positive appearance of a product might be enough to elicit selective exposure.

Still, the presentation of an overall rating should have been an even stronger cue of product quality. Results on perceived *surprise*, however, call into question whether participants actually recognized this cue. When the overall rating was presented *after* the subjective product evaluation, participants were only tendentially more surprised compared to when they had already seen it *before*. Thus, participants might not have paid attention to the overall

rating in the first place. To address this, we conducted a conceptual replication study in which we made the overall rating more salient in Study 1b.

Study 1b

Method

Design and Participants

Study 1b had the same between-groups factor (overall rating: before vs. after selecting reviews) as Study 1a. To make this a conceptual replication, we tested a similar number of participants. We were able to recruit 188 participants via social media platforms. However, 79 participants did not complete the study and their data could, thus, not be analyzed. We did not preregister Study 1b and, therefore, did not exclude any participants besides statistical outliers ($N = 3$). Including these three participants did not change our results. This left us with a final sample of $N = 106$ participants (73 female, age: $M = 30.42$ years, $SD = 10.92$, range: 18-66). Of this general-population sample, 49% were students, 47% were employed, and 4% did not answer this question. Like Study 1a, participants took part in a lottery for one of 10 vouchers for an online shopping portal worth 10 Euros each.

Procedure and Measures

The procedure and the measures were identical to Study 1a except for one crucial difference regarding the experimental manipulation. In the condition where the overall rating was presented *before* the review selection, we changed the visualization of this rating. Instead of presenting the overall rating on the same page with the product image and description, we now presented it on a separate page after the product image and description were displayed to ensure that the overall rating received more attention. Furthermore, the presentation of the overall rating was accompanied by a short explanatory text and the distribution of the 340 ostensible customer reviews.

Results

Review Selection

First, we calculated an independent two-sample t -test between experimental conditions to test Hypothesis 1 like we did in Study 1a. However, we did not find a difference between the two experimental conditions, $d = -0.18$, 95% CI [-0.57, 0.20], $t(104) = -0.94$, $p = .350$ (see Table 1).

As an alternative test of Hypothesis 1, we tested whether participants on average preferably selected positive reviews. A one-sample t -test revealed a significant deviation from the midpoint of the scale, $d = 0.56$, 95% CI [0.35, 0.76], $t(105) = 5.77$, $p < .001$. Replicating the findings of Study 1a, participants were biased towards a positive selection of reviews ($M = 3.22$, $SD = 0.39$). On average, participants selected 1.20 reviews with one star, 1.43 reviews with two stars, 1.47 reviews with three stars, 2.26 reviews with four stars, and 1.65 reviews with five stars.

Additional Analyses

Participants in the two experimental conditions did not differ in their levels of surprise towards the overall rating, $d = -0.06$, 95% CI [-0.26, 0.14], $t(98)^2 = -0.62$, $p = .537$. Likewise, there were no differences between experimental conditions regarding the subjective product evaluation and the certainty of this judgement, $|t|s < 0.65$, $ps > .522$. Again, there was a positive relationship between review selection and product evaluation, $r(106) = .28$, $p = .004$. The more positive reviews they selected, the more positively participants evaluated the product.

Discussion

Like Study 1a, we did not find evidence for a bias that is caused by the presentation of a positive overall rating before (vs. after) review selection. However, as before, participants on average (i.e., irrespective of whether they

saw the overall rating before or after review selection) showed a bias in favor of selecting more positive reviews. Given that the current study's setup makes it very unlikely that participants did not recognize the overall rating in the first place, it might be the case that the positive appearance of the product was sufficient to form a positive initial attitude towards it. This would also explain why participants were not more surprised when the rating was presented after selecting and reading the reviews. Participants might just expect a positive overall rating based on the appearance of the (admittedly fancy) product and their own attitude. As before, participants' product evaluations corresponded to the valence of selected reviews. Choosing and reading more positive reviews was related to evaluating the product more positively. Taken together, Studies 1a and 1b indicate that seeing an overall rating of a product before selecting reviews is not necessary for biased review selection. Rather participants tended to select positive reviews across the board which supports Hypothesis 1.

Study 2

To address the limitations of Studies 1a and 1b, we implemented some changes to the basic paradigm in Study 2. To avoid all participants forming a positive attitude because they see a high-quality product, we manipulated the quality of the product. To this end, we varied both the appearance of the to-be-explored product and the accompanying overall rating (i.e., suggesting either high or low quality of the product). Evidence for Hypothesis 1 would be provided by participants selecting more positive reviews in the high-quality product condition than in the low-quality product condition.

Method

Design and Participants

Study 2 was a laboratory experiment with two conditions (product quality: high vs. low). Based on the effects observed in Studies 1a and 1b, we ran an a-priori power analysis using G*Power (Faul et al., 2007). Assuming a small-to-medium effect size ($d = 0.40$) in an independent two-sample t -test (two-sided) with $\alpha = .05$ and $(1-\beta) = .80$, this analysis revealed a desired sample size of $N = 200$ participants. A total of 199 participants were recruited via a local participant pool that undergraduates can voluntarily sign up to. As preregistered (see <https://aspredicted.org/3r7d6.pdf>) for this study, we excluded people who were not fluent in German (due to the language sensitivity of the materials), psychology students (as they might be suspicious about study procedures and hypotheses), and participants older than 35 years (because these are a minority within the student body and might thus react differently to the material) in addition to statistical outliers. Including these participants did not change our results. The final sample consisted of $N = 192$ participants (147 female, age: $M = 23.42$ years, $SD = 3.40$, range: 18-33). Participants received 8 Euros as compensation for their participation in a lab session of about one hour (with this study being one of two independent parts)³.

Procedure

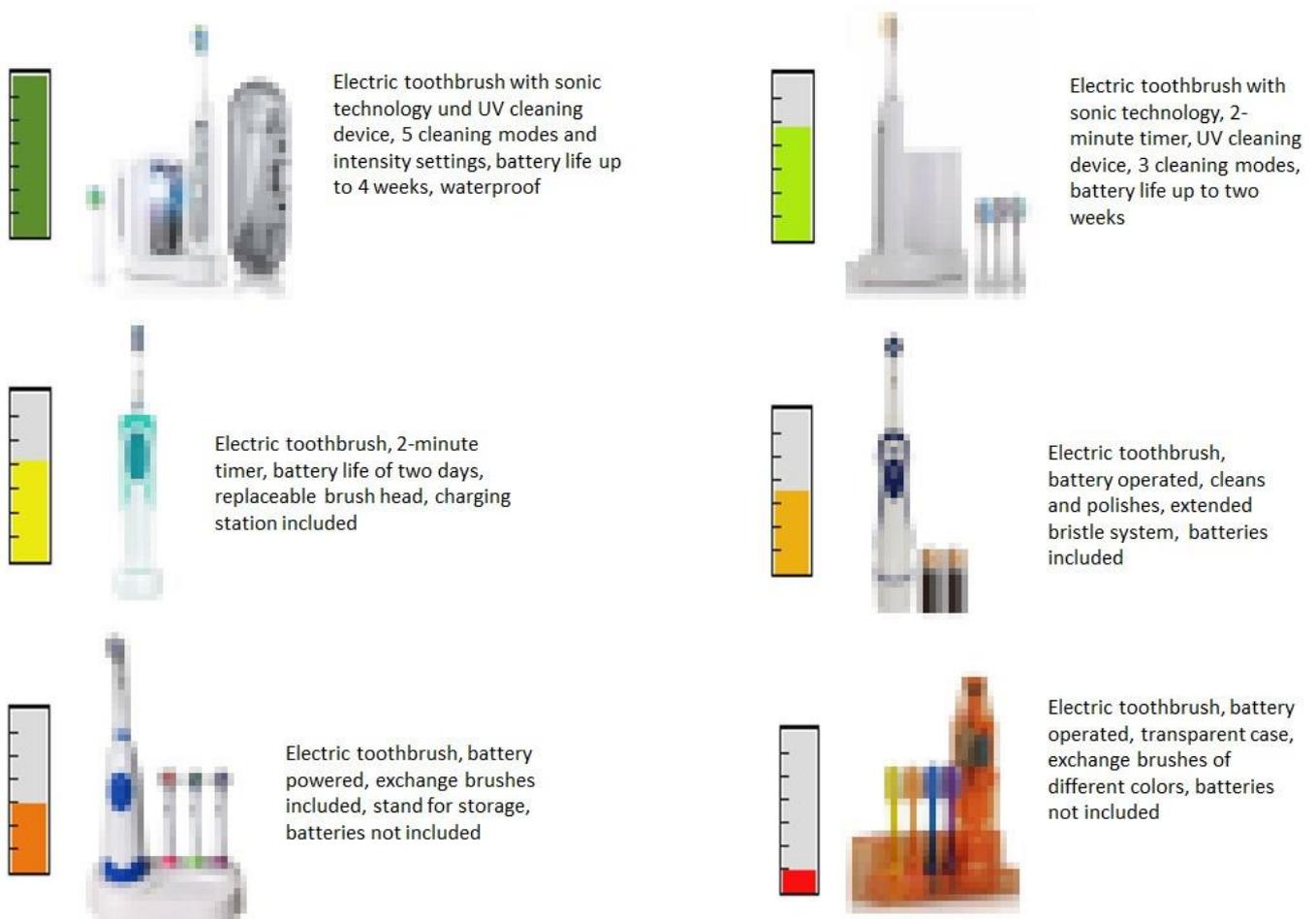
After informing participants that this study simulated an online shopping situation, they saw a page displaying six different electric toothbrushes. Next to the image of each product, we presented a short description and a vertical bar indicating the overall rating by previous participants. In choosing bars instead of stars, we intentionally deviated from the widely used design for product evaluation portals (as used in Study 1a and 1b) to avoid the results mirroring people's habitual use of product evaluation portals. The bars allowed for a differentiation of product quality in seven graduations (i.e., reflecting seven rating categories) with fuller bars representing better ratings (see Figure 2). Additionally, the bars were filled in different colors to represent different rating categories (e.g., deep green for the best Category 7 and deep red for the worst Category 1).

On the next page, participants were told that in the following sections they would learn more about one specific product. In the *high-quality* condition, the top left product with the highest rating was assigned, whereas the bottom right product with the lowest rating was assigned in the *low-quality* condition. In doing so, we aimed to manipulate the initial attitude participants had towards the product they ought to explore. We assumed a positive attitude would result from the assignment of the high-quality product and a negative attitude from the assignment of the low-quality product. The products were pre-tested to ensure that they were, indeed, perceived as high or low in quality by participants similar to our sample. Based on the results of this pre-test ($N = 51$), we used the toothbrush with the best rating ($M = 75.71$, $SD = 23.16$; scale from 0 to 100) as the high-quality product and the

toothbrush with the worst rating ($M = 27.94$, $SD = 19.08$) as the low-quality product. This supports our assumption that a high-quality (low-quality) product entails a positive (negative) initial attitude.

Next, participants received instructions to further explore the assigned product by choosing and reading customer reviews and to behave as if they were making a purchasing decision for no particular target (i.e., not specifying for whom they would buy the product). Then, participants selected a total of eight reviews stemming from any rating category they wanted. To this end, they entered a number (from 0 to 8) into a form field for each category captioned by the question "How many reviews out of Category X?". After participants completed review selection, they read the reviews on subsequent pages in a self-paced manner. After reading the reviews, participants were asked to rate the product – which served as a manipulation check in this study – and to indicate the certainty of their evaluation. Finally, some exploratory measures and demographic data were retrieved. Finally, participants were thanked and debriefed, and received their compensation.

Figure 2. Overview Page of the Available Products Used in Studies 2-4.



Note. Images are pixelated due to copyright reasons.

Measures

Review Selection. As before, our main dependent variable was the average rating of the selected reviews. Given that in this study rating categories ranged from 1 to 7 and in principal participants could choose eight reviews from the same category, average ratings could range from 1 to 7 as well.

Product Evaluation. As before, the product evaluation was assessed with one item ("How do you rate this product?"). However, in this study participants indicated their subjective rating of the product using a slider (from 0 = *very negative* to 100 = *very positive*) instead of the stars scale used in Studies 1a and 1b). The evaluation's certainty was assessed in the same way ("How certain are you about this rating?"; from 0 = *not at all certain* to 100 = *very certain*). The starting position of the slider was the scale midpoint (i.e., 50).

Results

Manipulation Check

Comparing the two experimental conditions in an independent two-sample *t*-test revealed that they led to different subjective evaluations of the respective product, $d = 1.88$, 95% CI [1.54, 2.88], $t(190) = 13.01$, $p < .001$. As intended, the high-quality toothbrush was rated more positively ($M = 65.19$, $SD = 18.83$) than the low-quality toothbrush ($M = 30.87$, $SD = 17.71$). Thus, our manipulation of product quality was successful. No difference emerged regarding the certainty of product evaluations, $t(190) = -1.01$, $p = .313$.

Review Selection

To test Hypothesis 1, we conducted an independent two-sample *t*-test that revealed no difference between experimental conditions, $d = 0.03$, 95% CI [-0.25, 0.32], $t(190) = 0.21$, $p = .833$. Review selection did not differ as a function of product quality (see Table 1).

Given that no differences between conditions occurred, we tested whether participants' selection did on average differ from the midpoint of the scale (i.e., 4). A one-sample *t*-test yielded a non-significant result which pointed in the opposite direction to our prediction, $d = -0.13$, 95% CI [-0.28, 0.01], $t(191) = -1.85$, $p = .066$. Thus, participants' review selection on average did not significantly differ from the midpoint of the scale. If anything, they descriptively selected negative reviews ($M = 3.91$, $SD = 0.70$) in both conditions, contrary to what we predicted. On average, participants selected reviews distributed as follows: 2.06 from Category 1 (i.e., the worst rating), 0.98 from Category 2, 0.64 from Category 3, 0.82 from Category 4, 0.73 from Category 5, 0.96 from Category 6, and 1.80 from Category 7 (i.e., the best rating).

Beyond that, the correlation between the valence of selected reviews and the product evaluation was significant, $r(192) = .22$, $p = .002$. The more positive the reviews selected (and read) were, the more positive the evaluation of the toothbrush was.

Discussion

In Study 2, we did not find evidence for the predicted positive bias in selecting customer reviews when exploring a high-quality product. Manipulation checks revealed that participants evaluated the products in line with our manipulation which resembles the results of the pre-test. However, this did not influence the direction of review selection. We did not find a more positive review selection when exploring a high-quality (vs. low-quality) product. Rather, the results pointed in the direction of a negative selection bias. This is inconsistent with Studies 1a and 1b, where we found a positively biased review selection when exploring a high-quality product. This mixed evidence regarding Hypothesis 1 calls for further investigations of the predicted bias as well as potentially promoting factors.

Study 3

Study 3 set out to test Hypothesis 2a that freely choosing (vs. assignment of) the to-be-explored product would facilitate positive review selection. To be able to test the effect of free choice versus assignment of a product, participants in the assignment condition also received a high-quality product. Differences between both conditions would provide evidence for Hypothesis 2a. Hypothesis 1 would be supported by an average review selection score above the scale mean (for the whole sample).

Method

Design and Participants

We conducted an online experiment with one between-groups factor (product selection: free choice vs. assignment). Our aim was to collect data from $N = 428$ participants. With a sample of this size, we would be able

to detect a small-to-medium effect ($d = 0.35$) in an independent two-sample t -test (two-sided) comparing the two experimental conditions based on $(1-\beta) = .95$ and $\alpha = .05$. We aimed to achieve high statistical power due to the null effect of Study 2 regarding Hypothesis 1a. A total of 526 participants took part in the experiment, which was advertised via the university's mailing list. Like Study 2, we preregistered (see <https://aspredicted.org/r8ff5.pdf>) to exclude participants that were not fluent in German, psychology students, and participants older than 35 years. As this study was conducted online, we additionally preregistered that we would exclude participants who used a mobile phone to complete the questionnaire (because the material was designed to fit larger screens). Ninety-six participants were excluded based on these criteria and the results of the statistical outlier analysis. Including these participants did not change the main results. The final sample consisted of $N = 430$ participants (314 female, age: $M = 22.75$ years, $SD = 3.14$, range: 18-35). All participants got the chance to take part in a lottery of 25 vouchers for an online shopping portal worth 10 Euros each.

Procedure

The setup of the study was similar to Study 2. However, this time participants were explicitly told to imagine that they wanted to buy the product for themselves, whereas in Study 2 they were not given a target. Then, the same overview page of six electric toothbrushes used in Study 2 was presented. Afterwards, the manipulation was implemented. In the assignment condition, participants were informed that they were to collect more information about the highest rated product from the pre-test (the one in the top left corner in Figure 2). Thus, this condition resembles the *high-quality* condition from Study 2. In the other condition, participants were *free to choose* the product that they liked the most. As outlined in the introduction, this can be seen as a demonstration of preference (comparable to other selective exposure research, e.g., Greitemeyer & Schulz-Hardt, 2003) and, thus, suggests a positive initial attitude towards the product. Of the participants in the free choice condition, 79.39% selected the toothbrush that was assigned to participants in the other condition.

Next, participants were asked to choose how many reviews they wanted to read. Again, the same seven rating categories used in Study 2 were available. To ensure that participants understood our rating category system, a horizontal bar was displayed next to each form field where participants had to indicate the number of reviews they wanted to read from that category ("How many reviews out of Category X?"). This horizontal bar was filled and colored to represent the respective category (e.g., completely filled deep green bar for Category 7). This time, however, we did not limit the amount of reviews in order to see whether potential differences only occur when the set of reviews is large enough. To make this possible while keeping the study's duration reasonable, we terminated the study after the selections were made. Participants did not actually read the reviews.

After review selection, we asked participants to rate the product and to indicate their certainty regarding this evaluation. Then, exploratory variables were assessed before we told participants that no reviews would be presented. This was followed by a basic demographic questionnaire, debriefing and the possibility to take part in the lottery.

Measures

Review Selection. Again, we assessed the average rating of the chosen reviews. This time, because there was no limit to how many reviews could be selected, the mean evaluation of the selected reviews could range from 1 (only the most negative reviews chosen) to 7 (only the most positive reviews chosen).

Product Evaluation. The subjective rating of the product as well as certainty were assessed like they were in Study 2—with a minor correction of the wording. Here, the slider scale ranged from 0 = *bad* to 100 = *good*.

Results

Hypothesis 1 was tested with a one-sample t -test that revealed a significant deviation from the midpoint of the scale, $d = 0.11$, 95% CI [0.02, 0.21], $t(429) = 2.38$, $p = .018$. Thus, participants selected predominantly positive reviews, but to a very small extent ($M = 4.07$, $SD = 0.63$).

The prediction that the free selection (vs. assignment) of a product would lead to more positive review selection (i.e., Hypothesis 2a) was tested with an independent two-sample *t*-test. There was no significant effect of experimental condition, however, the effect was in the expected direction, $d = 0.18$, 95% CI [-0.02, 0.36], $t(428) = 1.82$, $p = .069$. When freely selecting a product, participants only descriptively chose more positive reviews as compared to when the product was assigned (see Table 1). This tendency should, however, be interpreted with care considering the large sample size.

The average number of reviews selected for each category was as follows: 26.72 from Category 1, 25.50 from Category 2, 25.75 from Category 3, 26.35 from Category 4, 26.42 from Category 5, 27.29 from Category 6, and 7.98 from Category 7. A weak correlation between the valence of selected reviews and the product evaluation emerged, $r(430) = .09$, $p = .053$. Although no reviews were actually presented, participants who selected more positive reviews tended to evaluate the product more positively. No differences for the evaluation of the product nor for the certainty of this rating occurred, $|t|s < 0.91$, $ps > .365$.

Discussion

In Study 3, participants showed a very small but significant preference for positive reviews. Thus, in line with Studies 1a and 1b, the results of Study 3 supported Hypothesis 1 that review selection for a high-quality product leads to selecting more positive reviews. However, this effect did not depend on whether the product was assigned or freely selected. Although pointing in the hypothesized direction, the free choice of a product did not significantly affect information selection. This stands in contrast to Hypothesis 2a derived from selective exposure literature (e.g., Hart et al., 2009).

Additionally, it should be noted that the relationship between selected reviews and product evaluation was non-significant. The fact that the correlation was substantially smaller than in Studies 1a, 1b, and 2 is potentially due to the fact that participants did not actually read reviews in this study, but they did in the earlier studies.

Study 4

In Study 4, we also allowed participants to freely select the product they wanted to explore. In addition, we aimed to test the impact of another predictor of biased information selection that is highly prevalent in the design of most review selection interfaces, namely a sequential selection procedure (Jonas et al., 2001). To test Hypothesis 2b, we compared sequential selection to two versions of blocked selection. In one case, the blocked selection was carried out on one page (i.e., one-step) which is akin to the procedure used in previous research (e.g., Jonas & Frey, 2003; Schulz-Hardt et al., 2000). In the other case, participants selected the total number of reviews to be read, but subsequently selected the reviews for each rating category on separate pages (i.e., step-by-step). By adding this condition, we wanted to make sure that selecting all reviews before reading them and not the presentation of available reviews on the same page drives potential differences to the sequential condition. As before, evidence for Hypothesis 1 would be reflected in a positive deviation of the selected reviews (across all conditions) from the scale mean. Hypothesis 2b would be supported by the selection of more positive reviews in the sequential condition as compared to the two blocked conditions.

Method

Design and Participants

We conducted an online experiment with one between-groups factor with three levels (review selection procedure: sequential vs. blocked with one-step selection vs. blocked with step-by-step selection). Based on the small effect sizes obtained in the previous studies, we aimed to recruit $N = 510$ participants to increase the sensitivity of our analyses. With a sample of this size, we would be able to detect a small-to-medium effect ($f = .175$) in a one-way ANOVA with three experimental groups based on $(1-\beta) = .95$ and $\alpha = .05$. A total of 512 participants took part in our study that was advertised via the SoSci Panel, a large-scale volunteer participant pool (Leiner, 2016). Following our preregistered criteria (not fluent in German, psychology students/psychologists, or mobile phone users; see <https://aspredicted.org/9bs5x.pdf>) and outlier analyses, we excluded 18 participants from further data analysis. As we tested a general-population sample and not only students this time, we did not

exclude participants older than 35 years. Of these participants, 13% were students, 63% were employed, 15% were both students and employed, and 9% were categorized as other (e.g., retired, in training). Our final sample size was $N = 492$ participants (249 female, age: $M = 40.97$ years, $SD = 14.35$, range: 18-84). Including these participants rendered the main pattern of results, if anything, more pronounced. To compensate participants, we had a raffle of 30 vouchers for an online store worth 10 Euros each.

Procedure

At the outset of the study, participants were asked to select one of six different electric toothbrushes that they would like to buy in a situation where someone else was providing the money for it. In doing so, we aimed to establish a positive initial attitude towards the chosen product (like the *free choice* condition of Study 3). The setup of the overview page was identical to Studies 2 and 3 (see Figure 2).

After the initial product selection, participants could select between eight and sixteen reviews based on the same rating category system used in Studies 2 and 3 (i.e., seven categories represented by vertical bars). The order of review selection and review presentation differed between experimental conditions. In the *blocked with one-step selection* condition, participants had to indicate on one page how many reviews of each category they wanted to read, summing up to a total of at least eight but not exceeding sixteen reviews. Thus, the selection of categories was completed before any review was presented. In a second condition, the procedure was also *blocked* but *with step-by-step selection*. Here, participants indicated on separate pages from which category they wanted to read the next review. From the ninth selection onwards, participants could decide not to select any more reviews. Here, in the same way as the first condition, the selection was completed before any review was presented. In both conditions, the selected number of reviews from the respective categories were presented subsequently with one review per page. In the *sequential* condition, participants had to select one category at a time which was immediately followed by the presentation of a review from that category. This alternating selection and reading procedure was carried out between eight and sixteen times depending on whether participants wanted to stop or continue.

After the selection and reading phase, participants had to indicate how they personally evaluated the product and how certain they were about this rating. Finally, we asked some exploratory questions and retrieved demographic information, before participants were thanked, debriefed, and guided to the lottery.

Measures

Review Selection. Like previous studies, our main dependent measure was the average rating of the selected reviews. Again, the potential range of these ratings lay between 1 and 7.

Product Evaluation. Product evaluation and certainty were assessed in the same way they were in Study 3.

Results

To test whether there was a positive bias in the selection of reviews across conditions (i.e., Hypothesis 1), we calculated a one-sample t -test against the midpoint of the scale. In line with Hypothesis 1, this analysis revealed a small but significant effect, $d = 0.21$, 95% CI [0.12, 0.30], $t(491) = 4.58$ $p < .001$. On average, participants selected predominantly positive reviews ($M = 4.14$, $SD = 0.67$).

We computed a one-way ANOVA to test whether there are differences in review selection as a function of the selection procedure (Hypothesis 2b). Specifically, we were interested in the hypothesized pattern that sequential selection might foster a positive bias compared to the blocked selection procedures. This would result in a significant focal contrast (+2 sequential, -1 blocked one-step, -1 blocked step-by-step). The ANOVA revealed no significant effect of the experimental condition, $F(2, 489) = 2.79$, $p = .062$, $\eta^2 = .011$, 90% CI [.000, .029]⁴. The focal contrast was likewise non-significant and, if anything, pointed in the opposite direction to the one expected, $d = -0.18$, 95% CI [-0.37, 0.005], $t(489) = -1.92$, $p = .056$ (see Table 1).

Participants selected on average 1.90 reviews from Category 1, 1.13 from Category 2, 0.95 from Category 3, 1.19 from Category 4, 1.25 from Category 5, 1.34 from Category 6, and 2.12 from Category 7. Consistent with the

previous studies, the valence of selected reviews predicted the evaluation of the product, $r(492) = .27, p < .001$. The more positive the reviews selected were, the more positively participants evaluated the product after reading those reviews. Product evaluation and certainty did not differ between conditions, $F_s < 0.40, p_s > .646$.

Discussion

In Study 4, we obtained a small effect supporting Hypothesis 1: participants preferably selected positive reviews. However, we did not find evidence for Hypothesis 2b: sequential selection of reviews did not foster biased selection compared to blocked selection which we expected based on previous research (e.g., Jonas et al., 2001). If at all, the results pointed in the opposite direction. Thus, Study 4 completes the picture provided by our previous studies. Review selection is slightly biased towards positive reviews. However, potential promoting factors of biased information selection, which we derived from the literature, did not have a substantial effect in this context (see Table 1). To gain a comprehensive overview of the results and to be able to evaluate the practical relevance of the observed bias, we carried out an internal meta-analysis across all studies.

Table 1. Means and SDs for Selected Reviews as a Function of Experimental Condition Across All Studies.

Study	M (SD)
Study 1a (N = 156)	
Overall rating presented before selection	3.11 (0.32)
Overall rating presented after selection	3.11 (0.38)
Study 1b (N = 106)	
Overall rating presented before selection	3.18 (0.35)
Overall rating presented after selection	3.25 (0.42)
Study 2 (N = 192)	
High-quality product	3.92 (0.72)
Low-quality product	3.90 (0.69)
Study 3 (N = 430)	
Free choice of product	4.12 (0.63)
Assignment of product	4.01 (0.63)
Study 4 (N = 492)	
Sequential selection	4.06 (0.66)
Blocked (one-step) selection	4.23 (0.78)
Blocked (step-by-step) selection	4.13 (0.56)

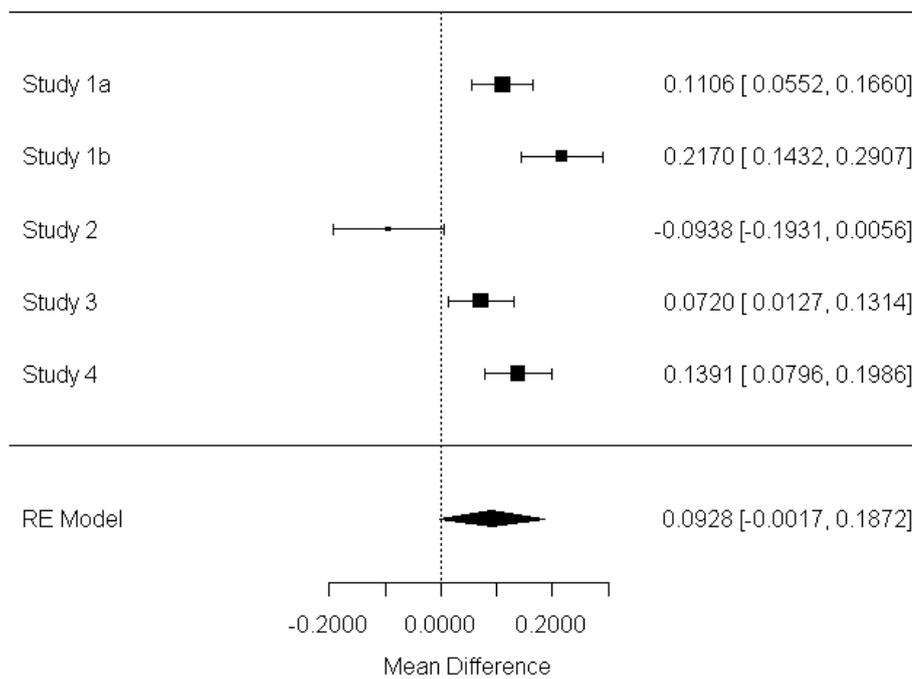
Note. In Studies 1a and 1b, participants selected reviews from five categories (midpoint = 3), whereas in Studies 2, 3, and 4, participants selected reviews from seven categories (midpoint = 4). Differences between experimental conditions in the single studies were non-significant at $\alpha = .05$.

Meta-Analysis Across All Studies

This internal meta-analysis sought to test our hypothesis that visitors of product evaluation portals are biased towards selecting positive customer reviews (compared to an average selection equal to the midpoint of the scale). In Studies 1a, 1b, and 2, Hypothesis 1 predicted this selection bias primarily in the experimental (i.e., overall rating presented before selection and high-quality product) and not in the control conditions (i.e., overall rating presented after selection and low-quality product). Given that no differences between the conditions emerged in these studies, we deemed it appropriate to enter effect sizes for the whole respective samples into the meta-analysis which increases statistical power.

We used the R package "metafor" (Viechtbauer, 2010) to calculate an aggregated effect size, weighted by the sample sizes of each study (random-effects model). This package does not allow Cohen's d values from one-sample t -tests to be entered. Thus, to calculate the internal meta-analysis's effect size, we converted the effect sizes reported above into mean differences (i.e., the deviation from the scale mean).

Figure 3. Mean Differences and 95% CIs From an Internal Meta-Analysis Across All Studies.



The internal meta-analysis revealed no significant effect in favor of Hypothesis 1 although pointing in the expected direction, $M_{diff} = 0.09$, $SE = 0.05$, 95% CI [-0.0017, 0.1872], $p = .054$, (see Figure 3). Transferring this aggregated score back to Cohen's d results in $d = 0.10$. Thus, across all studies the results did not support Hypothesis 1; participants showed no substantial preference for positive reviews.

General Discussion

The aim of the current research was twofold. First, we wanted to test whether the selection of customer reviews in the context of product evaluation portals is done in a biased fashion. Second, we wanted to investigate whether such a tendency is enhanced by certain interface properties that according to previous research should modulate selective exposure.

In four out of the five studies, our data suggests that visitors of product evaluation portals are biased towards positive review selection which is in line with Hypothesis 1. However, one needs to view this in light of the non-significant internal meta-analysis that provides a highly powered estimate of the true effect size. Thus, although there seems to be evidence for selective exposure in some of our studies, Hypothesis 1 was not substantially supported across the whole set of studies (which were all studies we conducted to address this research question). This stands against the majority of previous studies that propose a confirmation bias when selecting information (Hart et al., 2009). This means that, according to the current data, biased information selection is not a serious issue in the context of product evaluation portals, which is good news overall.

In addition, typical interface design properties that resemble factors which fostered biased information selection in earlier research, proved ineffective in our studies⁵. Whether participants were allowed to freely choose the product to further scrutinize (or not) did not substantially impact their preference for reviews of a certain valence (Study 3). This is unexpected given that earlier theorizing suggests that freely choosing a product should increase senses of ownership and commitment to the initial preference (Abelson, 1988; Hart et al., 2009). In our case, the free choice of a product did not foster biased review selection—though the means pointed in the expected direction. However, given the large sample size of that study we refrain from interpreting marginal differences as evidence against the null hypothesis. Therefore, Hypothesis 2a was not supported.

The second interface property under investigation concerned the procedure of selecting and reading reviews. Based on previous literature, choosing and reading pieces of information in an alternating (i.e., sequential) way should increase biased (i.e., confirmatory) selection as compared to a blocked procedure, where first all

information is selected and only read after the selection has been completed (Jonas et al., 2001). However, we did not find such a difference (Study 4). The strength of biased selection did not differ between the sequential procedure and the two types (i.e., one-step and step-by-step) of blocked selection, which is counter to Hypothesis 2b.

Taken together, none of the aspects of the interface that we varied, affected review selection. Based on our data, we can only speculate whether this is due to the small bias that occurred in the first place, which might have rendered it difficult at a statistical level, to find a moderation of this effect. Another potential issue is that previous research relied on small sample sizes by current standards and was thus underpowered. Alternatively, the selective exposure effect might not generalize to the current context due to specific features of product review selection.

Therefore, the present research ties in with previous work on advice taking. Most of the experimental research in this domain has relied on artificial paradigms like the judge-advisor system (JAS; Snizek & Buckley, 1995). The corresponding findings suggest that people behave egocentrically in these paradigms and adapt their initial judgements only to a small extent in response to the available advice (Harvey & Fischer, 1997). However, when enriching these paradigms, for instance, by letting participants freely choose how much advice to sample, they are actually willing to take advice in order to correct their initial judgments (Hütter & Ache, 2016). Likewise, research on information selection has found a strong confirmation bias under artificial conditions (e.g., Schulz-Hardt et al., 2000). The current research indicates that when transferring these paradigms to a more realistic context (i.e., product evaluation portals), information selection becomes way more balanced and rational (i.e., less egocentric).

Limitations and Future Directions

One might ask whether participants, indeed, obtained a positive initial attitude towards the product in our studies. In fact, we never explicitly measured this initial attitude as is done in the classical selective exposure paradigm (e.g., Jonas et al., 2005; Schulz-Hardt et al., 2000). As we outlined in the introduction, we wanted to keep the external validity of our studies as high as possible and initial product ratings are not provided by visitors of product evaluation portals. Still, we tried to capture the initial preference of participants by other means. First, we built on the assumption that the presentation of a high-quality product should be enough to form an at least mildly positive attitude. This corresponds with findings from consumer psychology (e.g., Jacoby et al., 1971; van Ooijen et al., 2017; Szybillo & Jacoby, 1974). Second, we pre-tested our materials in order to make sure that our high-quality product was, indeed, positively evaluated by participants. In line with this, the manipulation check in Study 2 revealed a more positive evaluation of the high-quality (vs. low-quality) product. This indicates that our manipulation of participants' (initial) attitudes towards the product was successful. Admittedly, stronger initial attitudes could have produced larger effects. Thus, testing the same effects in a less externally valid setting – where the initial attitude is explicitly assessed – or taking into account a priori interindividual differences (e.g., previous purchasing behavior, goal to buy a certain product) could be potential avenues for future research.

Related to this, one could argue that we implemented conditions of high accuracy motivation in our studies. This means that when exploring a product on the Internet, users might be inclined to search for the best possible outcome and not so much to defend their own standpoint (Jonas et al., 2005). It is debatable whether our study design or product evaluation portals in general instigate accuracy motivation. However, it should be noted that the role of accuracy motivation for selective exposure is somewhat inconclusive itself. Indeed, it has been argued and found that accuracy motivation reduces selective exposure (Hart et al., 2009). At the same time, there are situations in which accuracy motivation can even increase confirmatory information search (Fischer & Greitemeyer, 2010). Thus, to shed some light on the role of accuracy motivation in the context of product evaluation portals, more research is clearly needed.

All in all, the current research offers a solid starting point for future endeavors that address issues regarding biased review selection and product evaluation portals more generally. To build on our research on biased selection, one might also take into account biased information processing or contribution to such platforms (i.e., writing reviews). Given the prevalence of online purchasing in our everyday lives, these possibilities should be considered in the future. Also, it would be interesting to expand the scope of application beyond the product evaluation portals. For instance, selective exposure to customer reviews might also be an issue on service

evaluation portals. Based on the current findings, one would assume that people's selection of reviews on hotels or restaurants is also unbiased. However, this needs to be tested empirically.

Practical Implications

Although in some single studies the overall selection was slightly biased in the direction of positive reviews, the non-significant internal meta-analysis renders these findings less meaningful. In terms of practical significance, however, one might prefer to look at the aggregated effect size. Averaged across all studies, the mean difference from the midpoint of the scale (i.e., absolutely unbiased selection) was 0.9. This means that out of ten selected reviews, participants on average only chose one review that positively deviated by one point (i.e., four stars on a five stars scale) from a perfectly unbiased selection. Even when taking into account the upper limit of the effect's confidence interval, this number remains below two positively deviating reviews. Given that the internal meta-analysis provides a highly powered estimate of the true effect size, this effect cannot be seen as practically meaningful. Thus, the current research offers a clear message for visitors and designers of product evaluation portals: selective exposure only plays a minor role in this context. Rather Internet users seem to select customer reviews in an unbiased fashion which offers a good pre-condition for hopefully well-informed purchasing decisions.

Conclusion

The current work investigated the occurrence of biases in the context of customer review selection on product evaluation portals. Given that online customers often base their purchasing decisions on these reviews, biased selection could have detrimental effects. A meta-analytic inspection of our data revealed that across all studies there was no preference for a certain type of reviews, although descriptively more positive reviews were selected. Established moderators of selective exposure did not affect participants' choices when it comes to selecting customer reviews. It thus seems justified to conclude that the selection of these reviews does not show the biases that have consistently been found for information selection in other (less applied) contexts. Thus, the good news conveyed by the current research is that there seems to be little reason to worry that online customers select product reviews in a biased way.

Footnotes

1. In all studies in which a second product was presented, the results for the second product replicate those for the first one, so we mainly focused on the first product. Analyses of memory performance did not provide evidence for better memory regarding positive or negative information.
2. Differences in degrees of freedom are based on missing data for this item.
3. The experimental manipulation of the preceding study, which was about perceptions of situations in the working life, did not influence the results of this study.
4. For η^2 , we report 90% instead of 95% confidence intervals (see Lakens, 2014; Steiger, 2004).
5. This is also bolstered by comparisons of the distribution of the selected reviews (based on a non-parametric indicator), which likewise did not differ between conditions.

Acknowledgement

We would like to thank Nicole Russell Pascual for proofreading this manuscript. The research reported here was supported by a grant from the ScienceCampus Tübingen awarded to Mandy Hütter and Kai Sassenberg. The data and analysis code of all studies are openly accessible (data: <http://dx.doi.org/10.23668/psycharchives.4549>, analysis code: <http://dx.doi.org/10.23668/psycharchives.4547>).

References

- Abelson, R. P. (1988). Conviction. *American Psychologist*, 43(4), 267–275. <https://doi.org/10.1037/0003-066X.43.4.267>
- Becker, D., Grapendorf, J., Greving, H., & Sassenberg, K. (2018). Perceived threat and internet use predict intentions to get bowel cancer screening (colonoscopy): Longitudinal questionnaire study. *Journal of Medical Internet Research*, 20(2), Article e46. <https://doi.org/10.2196/jmir.9144>
- Buder, J., & Schwind, C. (2012). Learning with personalized recommender systems: A psychological view. *Computers in Human Behavior*, 28(1), 207–216. <https://doi.org/10.1016/j.chb.2011.09.002>
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). The Guilford Press.
- Chen, Y.-F. (2008). Herd behavior in purchasing books online. *Computers in Human Behavior*, 24(5), 1977–1992. <https://doi.org/10.1016/j.chb.2007.08.004>
- Clement, J. (2019). *Worldwide e-commerce share of retail sales 2015-2023*. Statista – The Statistics Portal. <https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/>
- Ditrich, L., Lüders, A., Jonas, E., & Sassenberg, K. (2019). Leader's group-norm violations elicit intentions to leave the group—if the group-norm is not affirmed. *Journal of Experimental Social Psychology*, 84, 103798. <http://dx.doi.org/10.1016/j.jesp.2019.03.009>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fischer, P., & Greitemeyer, T. (2010). A new look at selective-exposure effects: An integrative model. *Current Directions in Psychological Science*, 19(6), 384–389. <https://doi.org/10.1177/0963721410391246>
- Fretwell, L., Stine, J., Sethi, H., & Noronha, A. (2013). *'Catch and keep' digital shoppers: How to deliver retail their way*. Cisco IBSG. http://www.cisco.com/web/about/ac79/docs/retail/Catch-and-Keep-the-Digital-Shopper_PoV.pdf
- Greitemeyer, T., & Schulz-Hardt, S. (2003). Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology*, 84(2), 322–339. <https://doi.org/10.1037/0022-3514.84.2.322>
- Greving, H., & Sassenberg, K. (2015). Counter-regulation online: Threat biases retrieval of information during Internet search. *Computers in Human Behavior*, 50, 291–298. <https://doi.org/10.1016/j.chb.2015.03.077>
- Greving, H., & Sassenberg, K. (2018). Threatened individuals prefer positive information during Internet search: An experimental laboratory study. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(1), Article 6. <https://doi.org/10.5817/CP2018-1-6>
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>
- Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, 70(2), 117–133. <https://doi.org/10.1006/obhd.1997.2697>

- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, 57, 42–53. <https://doi.org/10.1016/j.dss.2013.07.009>
- Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9(3), 201–214. <https://doi.org/10.1007/s10799-008-0041-2>
- Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment and Decision Making*, 11(4), 401–415. <http://journal.sjdm.org/15/151110a/jdm151110a.html>
- Jacoby, J., Olson, J. C., & Haddock, R. A. (1971). Price, brand name, and product composition characteristics as determinants of perceived quality. *Journal of Applied Psychology*, 55(6), 570–579. <https://doi.org/10.1037/h0032045>
- Jonas, E., & Frey, D. (2003). Information search and presentation in advisor-client interactions. *Organizational Behavior and Human Decision Processes*, 91(2), 154–168. [https://doi.org/10.1016/S0749-5978\(03\)00059-1](https://doi.org/10.1016/S0749-5978(03)00059-1)
- Jonas, E., Schulz-Hardt, S., & Frey, D. (2005). Giving advice or making decisions in someone else's place: The influence of impression, defense, and accuracy motivation on the search for new information. *Personality and Social Psychology Bulletin*, 31(7), 977–990. <https://doi.org/10.1177/0146167204274095>
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571. <https://doi.org/10.1037/0022-3514.80.4.557>
- Jonas, E., Traut-Mattausch, E., Frey, D., & Greenberg, J. (2008). The path or the goal? Decision vs. information focus in biased information seeking after preliminary decisions. *Journal of Experimental Social Psychology*, 44(4), 1180–1186. <https://doi.org/10.1016/j.jesp.2008.02.009>
- Knobloch-Westerwick, S., & Kleinman, S. B. (2012). Preelection selective exposure: Confirmation bias versus informational utility. *Communication Research*, 39(2), 170–193. <https://doi.org/10.1177/0093650211400597>
- Kray, L. J., & Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, 91(1), 69–81. [https://doi.org/10.1016/S0749-5978\(02\)00534-4](https://doi.org/10.1016/S0749-5978(02)00534-4)
- Lakens, D. (2014, June 7). Calculating confidence intervals for Cohen's d and eta-squared using SPSS, R, and Stata. *The 20% Statistician*. <http://daniellakens.blogspot.de/2014/06/calculating-confidence-intervals-for.html>
- Leiner, D. J. (2016). Our research's breadth lives on convenience samples: A case study of the online respondent pool "SoSci Panel". *SCM Studies in Communication and Media*, 5(4), 367–396. <https://doi.org/10.5771/2192-4007-2016-4-367>
- Lueders, A., Prentice, M., & Jonas, E. (2019). Refugees in the media: Exploring a vicious cycle of frustrated psychological needs, selective exposure, and hostile intergroup attitudes. *European Journal of Social Psychology*, 49(7), 1471–1479. <https://doi.org/10.1002/ejsp.2580>
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (5th ed.). Irwin.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, 66(1), 102–110. <https://doi.org/10.1006/obhd.1996.0041>

Scholl, A., & Sassenberg, K. (2014). Where could we stand if I had . . .? How social power impacts counterfactual thinking after failure. *Journal of Experimental Social Psychology, 53*, 51–61.

<http://dx.doi.org/10.1016/j.jesp.2014.02.005>

Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology, 78*(4), 655–669. <https://doi.org/10.1037/0022-3514.78.4.655>

Schwind, C., & Buder, J. (2012). Reducing confirmation bias and evaluation bias: When are preference-inconsistent recommendations effective – and when not? *Computers in Human Behavior, 28*(6), 2280–2290.

<https://doi.org/10.1016/j.chb.2012.06.035>

Schwind, C., Buder, J., Cress, U., & Hesse, F. W. (2012). Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education, 58*(2), 787–796. <https://doi.org/10.1016/j.compedu.2011.10.003>

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing, 80*(2), 159–169. <https://doi.org/10.1016/j.jretai.2004.04.001>

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes, 62*(2), 159–174. <https://doi.org/10.1006/obhd.1995.1040>

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*(2), 164–182. <https://doi.org/10.1037/1082-989X.9.2.164>

Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre- and post-decision processes. *Acta Psychologica, 80*(1–3), 143–168. [https://doi.org/10.1016/0001-6918\(92\)90044-E](https://doi.org/10.1016/0001-6918(92)90044-E)

Szybillo, G. J., & Jacoby, J. (1974). Intrinsic versus extrinsic cues as determinants of perceived product quality. *Journal of Applied Psychology, 59*(1), 74–78. <https://doi.org/10.1037/h0035796>

van Ooijen, I., Fransen, M. L., Verlegh, P. W. J., & Smit, E. G. (2017). Packaging design as an implicit communicator: Effects on product quality inferences in the presence of explicit quality cues. *Food Quality and Preference, 62*, 71–79. <https://doi.org/10.1016/j.foodqual.2017.06.007>

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3). <https://doi.org/10.18637/jss.v036.i03>

Zhang, Y., Sun, Y., & Kim, Y. (2017). The influence of individual differences on consumer's selection of online sources for health information. *Computers in Human Behavior, 67*, 303–312. <https://doi.org/10.1016/j.chb.2016.11.008>

Correspondence to:

Kevin Winter
Leibniz-Institut für Wissensmedien
Schleichstr. 6
72076 Tübingen
Germany
Email: [k.winter\(at\)iwm-tuebingen.de](mailto:k.winter(at)iwm-tuebingen.de)

Editorial record: First submission received on July 9, 2020. Revision received on November 19, 2020. Accepted for publication on January 4, 2021.

Editor in charge: Alexander Schouten

About Authors

Kevin Winter is a post-doctoral researcher at the Leibniz-Institut für Wissensmedien, Tübingen, Germany. He researches the possibilities of reducing biased judgements in various settings (e.g., intergroup relations, online environments).

Birka Zapf is a research assistant at the Leibniz-Institut für Wissensmedien, Tübingen, and PhD candidate at the Eberhard Karls University of Tübingen, Germany. Her research centers around opinion expression in different contexts (e.g., on product evaluation portals or in conversations).

Mandy Hütter is full professor for Social Cognition and Decision Sciences at Eberhard Karls University of Tübingen, Germany. Her research focuses on evaluative learning and judgment as well as information sharing, sampling, and integration.

Nicolas Tichy is currently a PhD student at LMU Munich, Germany. His research focuses on information processing and human decision making at work.

Kai Sassenberg is head of the Social Processes Lab at the Leibniz-Institut für Wissensmedien, Tübingen, and full professor at the Eberhard Karls University of Tübingen, Germany. His research focuses on self- and emotion-regulation in the context of Internet use and social relationships (e.g., group membership, social power, leadership).