Unmasking the Deception: The Interplay between Fake Reviews, Ratings Discrepancy, and Consumer Demand*

Yunhao Huang† J. Miguel Villas-Boas‡ Mingduo Zhao§

September, 2025

*We thank Nano Barahona, Xin Chen, Mark Dean, Celia Gaertig, Benjamin Handel, Sherry He, Brett Hollenbeck, Ganesh Iyer, Przemysław Jeziorski, Kei Kawai, Jonathan Kolstad, Peiyao Li, Sarah Moshary, Olivia Natan, Steven Tadelis, and audiences at California Quantitative Marketing Ph.D. Student Conference, INFORMS Annual Meeting, Berkeley IO Workshop, Rising Scholars Conference, AI in Management Conference, FTC Conference on Marketing and Public Policy, Paris Conference on Digital Economics, Marketplace Innovation Workshop, and American Economic Association Annual Meeting for helpful comments. We thank Sherry He, Brett Hollenbeck, Gijs Overgoor, Davide Proserpio, and Ali Tosyali for sharing their data set with us. We also thank Xiyuan Tao and Yihan Zhou for their excellent research assistance. The experiments are reviewed and determined to be exempt by UC Berkeley IRB under protocol ID 2022-09-15658, 2023-05-16326, and 2024-10-17944. The experiments are pre-registered at AsPredicted (#127216, https://aspredicted.org/8c42-x4qw.pdf; #135597, https://aspredicted.org/74xj-789r.pdf; #162200, https://aspredicted.org/nnr4-v628.pdf); and #203486, https://aspredicted.org/sbwf-cgjf.pdf). This work was partially supported by UC Berkeley Behavioral Lab Mini Grant, UC Berkeley Xlab Research Grant, and UC Berkeley EGAL Research Grant. All errors are our own.

†University of Southern California. Email: yunhao.huang@marshall.usc.edu.

‡University of California, Berkeley. Email: villas@haas.berkeley.edu.

§University of California, Berkeley. Email: mingduo@berkeley.edu.

Unmasking the Deception: The Interplay between Fake Reviews, Ratings Discrepancy, and Consumer Demand

Abstract

In online marketplaces, consumers rely on reviews to make informed purchase decisions, making the presence of fake reviews detrimental. Previous literature implies that products with fake reviews can display some patterns in review distribution, such as a higher discrepancy in ratings. Consumers might take this pattern into account when making their purchase decisions. In this paper, we explore the interplay between fake reviews and ratings discrepancy, and their impact on consumer demand, while controlling for average product ratings. First, using a data set with fake review labels, we find that product ratings discrepancy is positively correlated with the probability that the product has fake reviews, with medium-to-large effect sizes. Second, through an identification strategy exploiting ratings discrepancy changes due to rating distribution rounding, we find evidence consistent with a negative and economically large causal impact of ratings discrepancy on consumer demand. Then, we conduct two experiments to establish and quantify the channel of the impact of ratings discrepancy on consumer demand through consumer suspicion of fake reviews. The first experiment shows that higher ratings discrepancy increases consumer suspicion of fake reviews, and the second experiment shows that heightened suspicion reduces consumer willingness to pay; both effects are economically large. Together, these findings reveal that consumers use ratings discrepancies as a signal of fake reviews, and this suspicion impacts their purchase decisions. The findings highlight the importance of understanding the relationship between fake reviews, ratings discrepancies, and consumer demand in online marketplaces.

Keywords: Fake Reviews, Online Marketplaces, Word of Mouth

JEL Classification: D12, D83, L15, L81

1 Introduction

In e-commerce, consumers often face information asymmetry challenges when evaluating products. To overcome these challenges, they increasingly rely on user-generated content (UGC), particularly product reviews and ratings. Major e-commerce platforms have implemented reputation systems to organize and display this content, helping consumers assess product quality efficiently.

However, the growing presence of fake reviews presents a challenge for these reputation systems, as they can distort the perceived quality of a product. This issue of fake reviews has gained considerable attention in recent years due to increased media coverage, raising public awareness and concern about the authenticity and reliability of online reviews (e.g., Federal Trade Commission, 2022). These reports detail the strategies used by sellers to generate fake reviews, such as offering incentives for positive reviews, using fake accounts, or engaging third-party services that specialize in review manipulation.

As consumers become increasingly aware of the prevalence of fake reviews in online marketplaces, they may look for indicators that help them infer which products may have fake reviews. Based on these indicators, consumers can potentially unmask the deception and make more informed decisions. One potential signal that consumers might use is rating distributions, which are prominently displayed on the product introduction page, frequently depicted as bar charts, representing the percentage of reviews associated with each rating level. In particular, the discrepancy in rating distributions, which represents the degree of variation in user ratings for a particular product, could be a factor to consider.

This paper examines three related research questions. First, we investigate the correlation between ratings discrepancies and fake reviews.[1] Several mechanisms could drive this relationship. When a seller receives negative reviews, it could react strategically by acquiring fake positive reviews (e.g., Luca and Zervas, 2016). This acquisition of fake reviews could potentially result in rating distributions with more discrepant ratings. Another potential reason for this connection could be that new sellers may acquire 5-star ratings when they enter the platform, which leads to a substantial, albeit potentially transient, increase in average ratings (e.g., He et al., 2022b). When sellers stop pur-

[1] A review usually consists of a rating and text, and sometimes with images or videos. The focus of this paper is the rating. In the following, when referring to "reviews," we mostly refer to "ratings."

chasing fake reviews, the proportion of 1-star reviews increases (e.g., Chakraborty et al., 2022). These considerations suggest that, when the average rating is held constant, a larger discrepancy in ratings, one measure being the standard deviation of the rating distribution, is positively associated with the likelihood that a seller's product contains fake reviews. We confirm this hypothesis using a data set with validated fake review labels from He et al. (2022a), showing medium-to-large effect sizes when controlling for the average ratings.

Second, we examine whether ratings discrepancy causally affects consumer demand. One potential mechanism could be that, as a higher ratings discrepancy is correlated with fake reviews, it can raise consumer suspicion and impact consumer demand. To identify the impact of ratings discrepancy on consumer demand, we exploit changes in ratings discrepancy that occur due to percentage rounding in rating distributions. This novel approach allows us to isolate the causal impact of ratings discrepancy while controlling for potential confounding factors. The findings are consistent with a negative effect of ratings discrepancy on consumer demand.

Third, we investigate one potential channel underlying the negative effect of ratings discrepancy on consumer demand: suspicion of fake reviews. Specifically, we examine whether consumers interpret ratings discrepancies as a signal of fake reviews, leading them to reduce their willingness to pay (WTP). We conducted two experiments to establish this causal chain. In the first experiment, participants were shown products with varying rating distributions and asked about their suspicion of fake reviews. We find that a 1.0 increase in the ratings standard deviations raises the perceived probability of fake reviews by 21.8 percentage points. In the second experiment, we manipulate participants' suspicion through platform warning labels and find that a one percentage point increase in the perceived probability of fake reviews reduces WTP by 4.4 cents. Taken together, this implies that a 1.0 increase in the ratings standard deviations reduces WTP by about \$0.96[2] on a \$20 product through the fake review suspicion channel. For context, when using the data of the first experiment alone, we find that a 1.0 increase in the ratings standard deviations reduces WTP by \$0.84. This highlights that consumers use ratings discrepancies as a signal of fake reviews and that this suspicion, as one important channel, impacts their purchase decisions.

[2]That is, $21.8 \times \$0.044 = \$0.096$.

We contribute to research on fake reviews and their market implications. These studies explore the impact of fake reviews on reputation and competition (e.g., Mayzlin et al., 2014; Luca and Zervas, 2016; Li et al., 2020). While prior work has documented various ways fake reviews affect reputation and competition, we identify a new indicator, ratings discrepancy, that consumers use to detect potential review manipulation. We also contribute to the literature examining the factors that drive ratings discrepancy in online reviews (e.g., Schoenmueller et al., 2020; Karaman, 2021) by showing how fake reviews are associated with more discrepant rating distributions. Moreover, we contribute to the literature on fake review detection (e.g., Zhou and Zafarani, 2020; Luo and Ma, 2024) by highlighting rating distribution patterns associated with review manipulation.

We also contribute to the literature that investigates the causal effects of UGC on consumer demand. Prior studies have demonstrated that higher average ratings positively affect consumer demand, and the literature concerning online reviews investigates the influence of these average ratings on consumer choices (e.g., Chen and Xie, 2008; Cabral and Hortacsu, 2010; Moe and Trusov, 2011; De Langhe et al., 2016; Luca, 2016; Park et al., 2021; Pei and Mayzlin, 2022; Reimers and Waldfogel, 2021; Zhong, 2022), and the factors driving consumer engagement in word-of-mouth communication (e.g., Li and Hitt, 2008; Proserpio and Zervas, 2017; Chevalier et al., 2018; Chakraborty et al., 2022; Amaldoss and Shin, 2025; Xu et al., 2025). Regarding ratings discrepancy, the existing literature exhibits mixed evidence of its effect on consumer preferences and demand (e.g., Sun, 2012; Luo et al., 2013; He and Bond, 2015; Wu et al., 2015; Rozenkrants et al., 2017), with both positive and negative impacts depending on factors like self-expression and product polarization. We contribute to this literature by identifying and quantifying a specific mechanism, consumer suspicion of fake reviews, through which ratings discrepancy affects demand.

The remainder of the paper is organized as follows. Section 2 describes two observational data sets employed in this paper. Using the first data set, Section 3 establishes the correlation between ratings discrepancy and the probability of a product having fake reviews. Using the second data set, Section 4 estimates the impact of the ratings discrepancy on consumer demand. To further explore the channel of how ratings discrepancy affects demand through fake reviews, and to quantify this channel, Sections 5 and 6 detail two online lab experiments. Section 5 shows that ratings discrepancy increases consumer

suspicion of fake reviews, and Section 6 shows that heightened suspicion of fake reviews reduces willingness to pay. Lastly, Section 7 discusses managerial implications, and Section 8 reviews the main findings and presents directions for future research.

2 Observational Data

In this paper, we employ two separate observational data sets. The first is a secondary data set from He et al. (2022a). The second is from the "Home and Kitchen" category on Amazon, covering the period from January 2023 to September 2023. In addition to these observational data sets, we conducted two experiments to further investigate the research questions; these will be discussed in detail in Sections 5 and 6.

2.1 Amazon Data with Fake Review Labels

The data set used by He et al. (2022a) includes a crucial variable indicating whether a seller has acquired fake reviews for a product during the observation period. Most sellers identified as having acquired fake reviews compensate reviewers by refunding the product's cost through online transactions after acquiring and posting a 5-star review. These reviewers are incentivized to fabricate seemingly authentic 5-star reviews that can bypass Amazon's detection systems. In contrast to "incentivized reviews," where sellers provide free or discounted products (or future discounts) in exchange for a review that discloses the user's experience and is not necessarily five stars, fake reviews specifically require a five-star rating in order to receive reimbursement.

The data set is collected by tracking Facebook groups that recruited individuals to leave fake reviews from October 2019 to June 2020. This study employs a cross-sectional sample of approximately 1,500 unique products from October 2020, where the identified products with fake reviews are included. In addition to identifying products with manipulated ratings, the data set contains further information about these products on Amazon, such as the number of 1-star and 5-star ratings, average ratings, product ID, the proportion of helpful reviews, and other details about the product and the reviewer's network. The dataset also includes "control" products that share many similar characteristics with the fake review product but did not acquire fake reviews.

Regarding ratings, this data set only records the proportion of 1-star and 5-star rat-

ings, and the average ratings of each product. This means that we lack information on the standard deviation, and we need to impute it. For each product, this imputation method yields an upper bound on the standard deviation by assuming no 3-star ratings (in other words, all other ratings are 2-star or 4-star ratings) and solving for the number of 2-star and 4-star ratings that lead to an average rating consistent with observed ratings. The results are robust to other ways of imputation, such as a lower bound on the standard deviation derived similarly as above.

		Having fake reviews?		
	(1)	(2)	(3)	(4)
Variable	All	Yes	No	p-value
# Reviews	410.23	322.35	477.05	<0.001
Avg. rating	4.16	4.26	4.07	<0.001
Share of helpful reviews	0.22	0.20	0.24	<0.001
N	3408	1472	1936	

Table 1: Summary statistics of products by having fake reviews or not

Notes: In column (4), we test the hypothesis of H_0: the mean of the variable in column (2) = (3) against H_1: the mean of column (2) $\neq$ (3), using a t-test, for each variable respectively.

Table 1 provides a comparison of summary statistics for products in two groups: those having fake reviews (column (2)) and those not (column (3)). The table contrasts various characteristics of products in these groups, such as the number of reviews, average rating, and the proportion of helpful reviews.

In the data set, 1,472 products are flagged as having fake reviews, while 1,936 products are labeled as not engaging in such practices. The latter comprises products that directly compete with the fake review products, serving as a comparison set. To construct the comparison set, He et al. (2022a) identified the two competitors that appeared most frequently on the same search page during the seven days before and the seven days after the initial social post concerning fake reviews. By employing this method, the comparison products share the same sub-category and a similar search ranking as the fake review products. This allows, in a reduced-form way, for better control of confounding factors, making the products more comparable in the analysis.

Table 1 indicates that products with fake reviews have, on average, a smaller total number of reviews (322) compared to their counterparts without fake reviews (477). The lower average number of reviews for products with fake reviews might imply that these

products are less established, and, therefore, sellers resort to acquiring fake reviews to rapidly enhance the products' reputation. This approach may allow them to compete with more established products that have garnered a greater number of authentic reviews over time. By doing so, it may address the "cold-start" problem when launching new products. The mean rating for products in the fake review group is higher (4.26) than those without fake reviews (4.07). A higher average rating for products with fake reviews is anticipated, as one of the primary objectives of acquiring fake reviews is to inflate the product's overall rating, making it more attractive to prospective consumers. This outcome indicates the efficacy of fake reviews in manipulating a product's perceived quality. The proportion of helpful reviews is lower for products having fake reviews (0.20) in comparison to those without (0.24). The decreased proportion of helpful reviews for products with fake reviews may suggest that consumers find these reviews less informative or relevant. This might be attributed to the superficial nature of fake reviews, which often lack the detail and personal experience present in genuine reviews. Consequently, potential consumers may not find the reviews useful in making informed purchasing decisions.

2.2 Amazon Data with Sales and Sales Rankings

The first observational data set has several limitations. First, the available version of the first data set lacks information on sales. Second, the first data set was collected in 2020 and many of the original products are no longer accessible.[3] For these reasons, we obtain a second observational data set through web scraping, which allows us to record sales rankings and infer sales.

This paper's second observational data set focuses on the "Home and Kitchen" category[4] on Amazon during the period from January 2023 to September 2023. This data set comprises a daily panel,[5] incorporating a union of products drawn from the daily best-seller ranking list in the "Home and Kitchen" category and from daily search results on the first three pages for the keywords "Home and Kitchen." The panel encompasses various variables, including average ratings, standard deviations of ratings, prices, number

[3]This is likely due to the fact that numerous products featured in the data set are no longer available under the same link, rendering retrospective tracing infeasible.

[4]The rationale for selecting the "Home and Kitchen" category as the population stems from the findings of He et al. (2022b), which indicate that "Home and Kitchen" is one of the largest categories associated with the purchasing of fake reviews.

[5]The panel is unbalanced because the products appeared in the sales ranking list and the search list could be different every day.

of ratings, and position in the sales ranking or search list. Alongside the variables that we directly observed, another key variable is sales, which is not directly observable. To gather sales quantity data, we leverage the method outlined in Chevalier and Goolsbee (2003) to infer sales based on sales rankings. Specifically, they find a linear relationship between the logarithm of sales and the logarithm of sales rankings, so one can infer sales based on sales rankings if the relationship is known. To estimate such a relationship, we collect product inventory data based on a feature of the Amazon website.[6] If we observe the inventories of a product on two consecutive days, we can calculate the sales as the difference between them.[7] We then utilize such observations to compute daily sales and to estimate the relationship between sales and sales rankings.[8] Finally, we infer sales based on the sales rankings and the estimated relationship.

Other general questions to ask in order to understand the empirical results are (i) which sellers may choose to acquire fake reviews, and (ii) how they might use the fake reviews to obscure any inferences consumers might draw. In this regard, we do not take into account either the selection of which sellers may choose to acquire fake reviews, or how sellers may try to signal-jam inferences of fake reviews by consumers. The results presented in the next sections should then be interpreted as presenting the correlation between the existence of fake reviews and ratings discrepancy, and the effect of ratings discrepancy on consumer demand, given the existing behavior of sellers. Fully accounting for the selection of which sellers choose to acquire fake reviews, and how sellers manage consumers' potential inferences about fake reviews, could be interesting to study in future research, but it is beyond the scope of this paper. We discuss some of these questions further in the Appendix A.

[6]By incrementally adding units of a product to the Amazon shopping cart, we can observe a product's inventory until the seller exhausts their stock. At this juncture, Amazon exhibits an alert denoting the total quantity of available units. For products with inventories below 1000, this approach enables us to observe the number of units in stock.

[7]We discard any observations where the inventory on the first day is lower than that on the second day, assuming that inventory increases indicate the much less frequent restocking behavior. Also, we exclude observations where the inventory is missing or at the upper limit of 999, or if the seller restricts the number of purchasable units.

[8]To increase the accuracy of the relationship, we incorporate the slope coefficient estimate in He et al. (2022b), and we use this data to estimate the intercept coefficient.

3 Ratings Discrepancy and Fake Reviews

In this section, we establish the existence of a positive correlation between fake reviews and ratings discrepancy.[9] This correlation can serve as supporting evidence that consumers may infer that a product with more discrepant ratings is more likely to have fake reviews. To establish such a relationship, we examine the data set containing labels for fake reviews. It is important to recall that the majority of sellers identified as having acquired fake reviews compensate reviewers by reimbursing the product's cost after receiving a 5-star review.

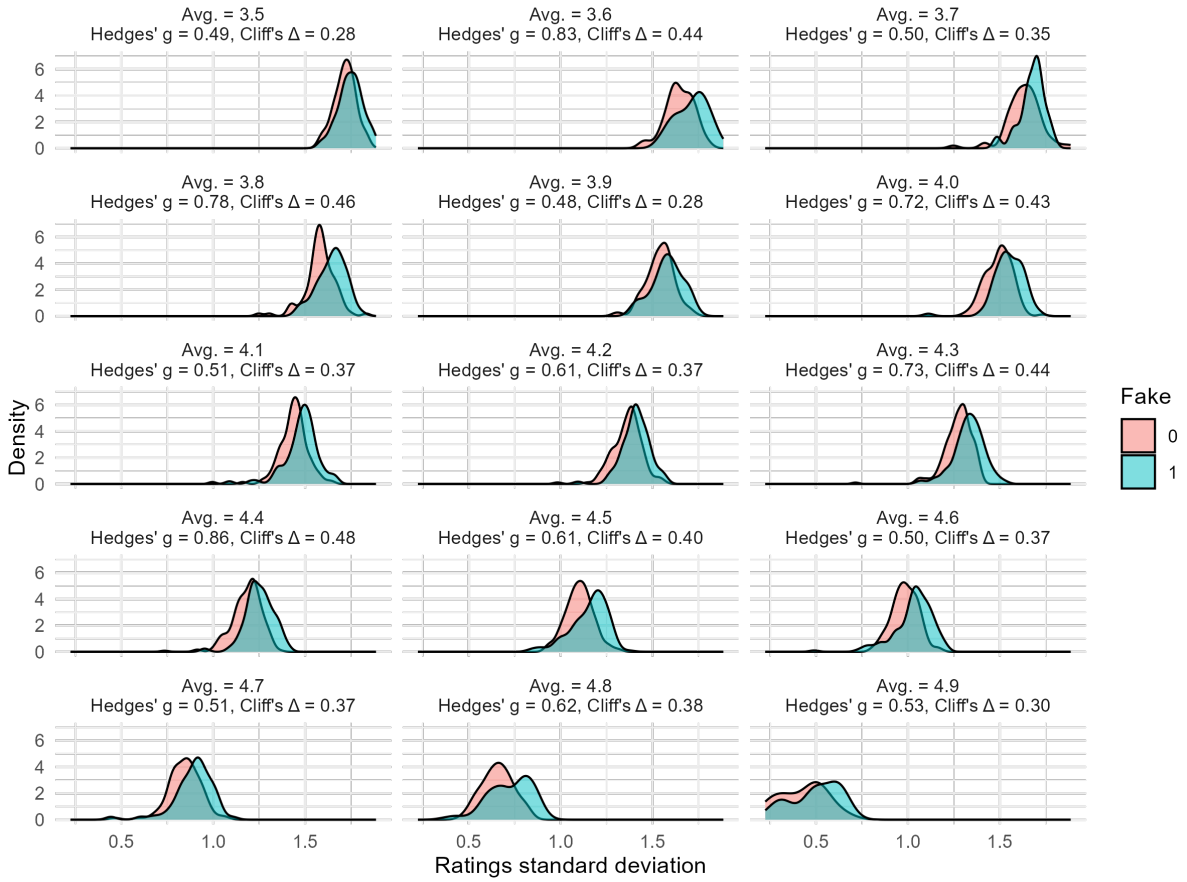


Figure 1: Products with fake reviews have higher ratings discrepancy

Notes: This figure uses the labeled dataset from He et al. (2022a). Products are tagged as "with fake reviews" if they were identified reimbursing buyers via Facebook groups for 5-star reviews; "without" products consist of competitors who did not acquire fake reviews. We restrict the sample to products with at least 10 reviews. In each panel, we further restrict the sample to products with a certain rounded average rating and plot the density distribution of the standard deviations of ratings for products with and without fake reviews. Across all levels of rounded average ratings, the panels suggest higher standard deviations of ratings for products with fake reviews compared to products without fake reviews. To quantify magnitudes, we calculate two effect sizes within each panel: Hedges' g and Cliff's Δ, which compare the standard deviation of ratings for products with versus without fake reviews. The values are displayed in each facet title, suggesting that the effect sizes are medium-to-large.

Each plot in Figure 1 illustrates a variation in the standard deviation of ratings

[9] See the Appendix for a theoretical framework about ratings discrepancy.

when comparing products with fake reviews to those without, across various average rating levels. The horizontal axis represents the standard deviations of ratings, while the vertical axis shows the corresponding density. The average ratings in the plot (ranging from 3.5 to 4.9) include approximately 90% of the products in the data set. Across all observed average rating levels, the plots illustrate that products labeled with fake reviews exhibit a density distribution that leans more to the right. This translates to products with fake reviews showing a higher ratings discrepancy than those without. Moreover, we report effect sizes (Hedges' g and Cliff's Δ) computed within each average-rating bin. Conditional on the average ratings, the difference in ratings discrepancy is medium-to-large across average-rating bins (Hedges' g ranges from 0.49 to 0.86, and Cliff's Δ ranges from 0.28 to 0.48; see Figure 1). Overall, this finding suggests that products with a higher ratings discrepancy are more likely to have fake reviews, holding fixed average ratings.[10] It is important to note that this does not imply whether consumers can perceive such differences and use them as a signal. Our subsequent observational analysis and experiments are designed to test this question, namely, whether consumers identify and act on ratings discrepancies when evaluating products.

Figure 2 displays a heatmap with the standard deviations of ratings on the vertical axis and the average ratings on the horizontal axis. The color denotes the probability of a product having fake reviews within each grid, with lighter colors signifying higher probabilities. This probability is calculated by taking the average of the binary fake review indicator from the data. The figure illustrates how the standard deviation of ratings is related to the probability of a product having fake reviews while keeping the average rating constant. For products with identical average ratings, represented by points on a vertical line, the heatmap is lighter when the standard deviations of ratings are higher, showing its positive correlation with the probability of having fake reviews.

The above robust and model-free observations provide support to the argument that, when the average rating remains constant, a higher standard deviation of ratings is indicative of an increased probability of a product having fake reviews.

We then use regression analysis to further quantify this relation. We first estimate

[10]When comparing sub-plots with varying average ratings (from the top left to the bottom right of Figure 1), those with a higher average rating tend to have a lower standard deviation. This is expected because products with higher average ratings naturally have a reduced upper limit for their standard deviation.

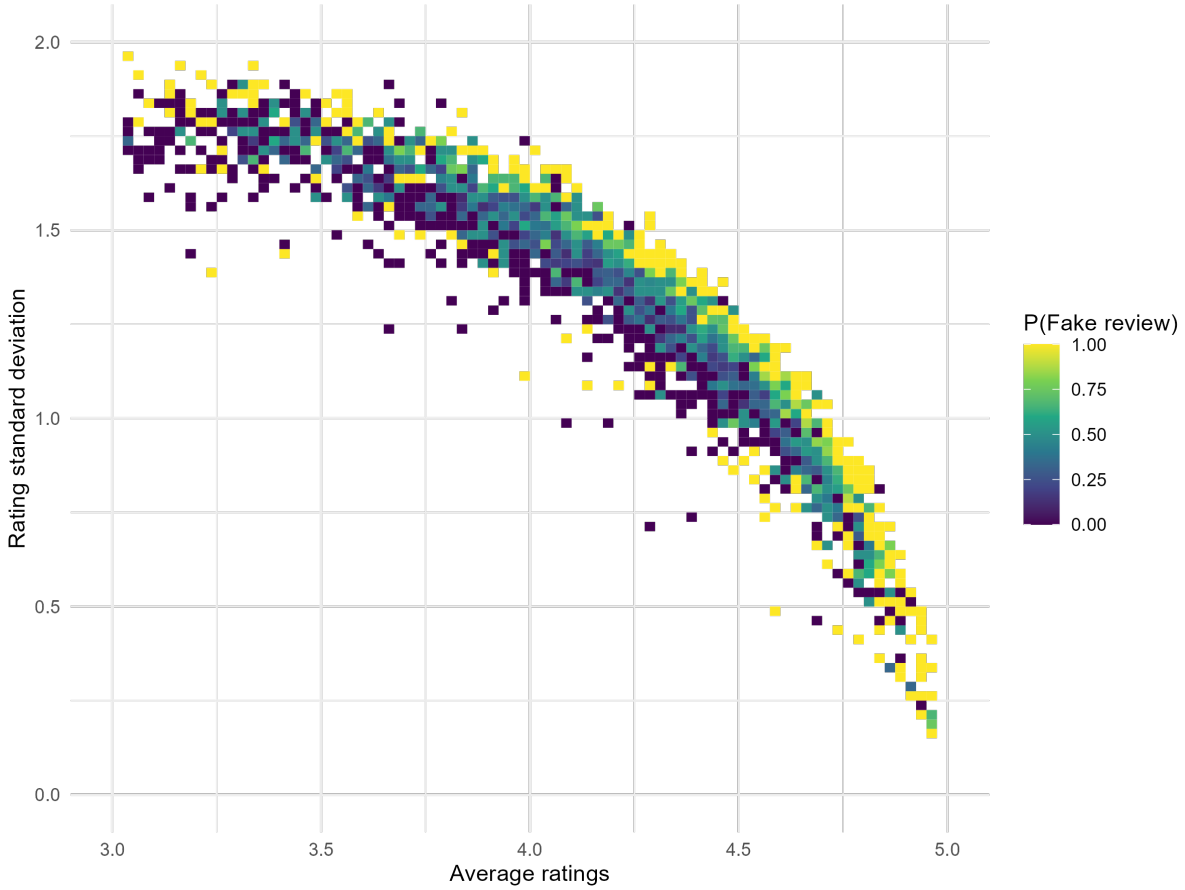


Figure 2: Probability of purchasing fake reviews against average ratings and standard deviations of ratings

Notes: This figure uses the labeled dataset from He et al. (2022a). Products are tagged as "with fake reviews" if they were identified reimbursing buyers via Facebook groups for 5-star reviews; "without" products consist of competitors who did not acquire fake reviews. We restrict the sample to products with at least 10 reviews. For the products with the same average rating, the probability of a product having fake reviews increases as the standard deviation of ratings increases. The heat map displays the probability of purchasing a product with fake reviews against the average ratings and the standard deviations of ratings, with lighter shades indicating higher probabilities of fake reviews.

the following regression:

$$\text{Fake}_j = \alpha_{R_j} + \alpha_1 \cdot \sigma_j + \alpha_2 \cdot X_j + \epsilon_j, \tag{1}$$

where j indexes the products in this data set, Fake_j denotes whether product j has fake reviews, α_{R_j} denotes a fixed effect based on the rounded average rating R_j of the product, σ_j represents the standard deviation of ratings, which serves as a measure ratings discrepancy for a given product, and X_j denotes potential covariates, including the logarithm of the number of reviews. In addition to the linear regression specification above, we also employ a logistic model and a probit model. Despite potential measurement error that may arise from using the fake review product label as the ground truth outcome variable, the estimates derived from (1) remain consistent under the assumption that the measurement error is orthogonal to the covariates.

	Outcome: Having fake reviews (Indicator)					
	OLS		Logistic		Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
Std. ratings	1.125***	1.233***	6.189***	6.459***	3.118***	3.423***
	(0.078)	(0.078)	(0.443)	(0.439)	(0.252)	(0.250)
Log # reviews		−0.052***		−0.229***		−0.144***
		(0.006)		(0.027)		(0.016)
Avg. rating fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	3,408	3,408	3,408	3,408	3,408	3,408
BIC	4,803	4,724	4,540	4,474	4,569	4,496

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 2: Products that have fake reviews are associated with a higher ratings discrepancy
Notes: This table presents the coefficients from the regression of the indicator of a product having fake reviews on average ratings, standard deviations of ratings, and the logarithm of the number of reviews. Columns (1) and (2) employ ordinary least squares methods, columns (3) and (4) employ logistic regressions, and columns (5) and (6) employ probit regressions. Standard errors are reported in parentheses below the coefficients.

Table 2 displays the coefficients illustrating the conditional correlation of the standard deviation of ratings with the probability of a product having fake reviews. Regardless of whether the covariates are included, the standard deviation is consistently positively correlated across various models with the probability of fake reviews. This relationship can be interpreted in several ways: One potential explanation is that sellers who received a lower initial average rating may acquire fake reviews to increase their average ratings.

This could potentially result in a large proportion of 5-star ratings and a large proportion of negative ratings at the same time, that is, a higher level of ratings discrepancy. Another potential explanation is that acquiring fake reviews early in a product's lifecycle might lead to more discrepant ratings later on. This could occur if customers, influenced by overly positive reviews, receive a product that does not meet their expectations and subsequently leave a 1-star review in reaction to the higher expectations. Moreover, although this relationship may not be causal, the presence of this correlation can potentially serve as a signal for consumers to be wary of products with a higher ratings discrepancy.

Moreover, the degree of ratings discrepancy can be quantified using various methods. In the analysis in this section, we employ the standard deviation of the ratings as a measure, which generally yields higher values with more extreme reviews. In the Online Appendix, we explore two alternative measures that capture a similar notion of discrepancy: (i) the proportion of 1-star ratings, and (ii) the interaction between the proportion of 1-star ratings and the proportion of 5-star ratings, while controlling for each proportion individually. Holding constant the average ratings, these two measures will both be larger when the ratings are more discrepant, i.e., when both the 1-star and 5-star proportions are high. The alternative measures lead to the same substantive results as those presented here. This suggests that the finding of a positive correlation between ratings discrepancy and having fake reviews is robust across different measures of ratings discrepancy.

4 Impact of Ratings Discrepancy on Sales

In Section 3, we show that products with fake reviews, when conditioned on the same average ratings, exhibit higher standard deviations of ratings. In this section, we hypothesize that the discrepancy in the rating distribution influences sales negatively. To test this hypothesis, we use data scraped from Amazon and employ ratings discrepancy changes due to rating distribution rounding as the identification strategy.

The identification approach is based on Amazon's rounding system. We limit the sample to products that have the same rounded average ratings (but potentially different standard deviation of ratings) on adjacent days. Recall that these rating distribution percentages are rounded. We calculate the standard deviation of ratings based on the

rounded percentages. In such cases, the change in σ_{kt} is due to the rounding of rating distributions. For instance, while the average rating of a product may remain the same at 4.4 on adjacent days, the percentage of 5-star ratings could change from 79% to 80%, and the percentage of 1-star ratings could change from 9% to 10%, leading to changes in the ratings discrepancy.

Taking the difference between adjacent observations exploits the variation induced by Amazon's rounding system, leading to the following regression:

$$\Delta \log Y_{jt} = \gamma_t + \gamma_1 \cdot \Delta\sigma_{jt} + \gamma_2 \cdot \Delta \log P_{jt} + \gamma_3 \cdot \Delta X_{jt} + e_{jt}, \tag{2}$$

where j indexes the products in the data, t indexes the date, Δ denotes the variable difference on two adjacent days, Y_{jt} denotes sales or sales rankings, γ_t denotes time fixed effects, including week fixed effects and day-of-the-week fixed effects, σ_{jt} denotes the standard deviation of the ratings, calculated based on the rounded rating percentages,[11] P_{jt} denotes the product price, X_{jt} denotes a vector of covariates, including the logarithm of the number of reviews[12] and the logarithm of the product's position in search results,[13] and e_{jt} is the error term. We cluster the standard errors at the product level.

Intuitively, the identification strategy is to compare a single product on two adjacent days that possess identical rounded average ratings but could potentially exhibit minor variations in discrepancy. Then, by taking the difference between adjacent days, we aim to partially address potential endogeneity concerns that could arise from unobserved factors. Taking the first difference allows us, first, to control for time-invariant unobserved heterogeneity. Second, by taking first differences, we remove time-invariant unobserved product characteristics that may be correlated with both prices and sales. While this does

[11] Using alternative measures of ratings discrepancy, e.g., (i) the proportions of 1-star ratings, or (ii) the proportions of 5-star and 1-star ratings and its interaction, instead of the standard deviation of ratings leads to the same substantive results as those presented here (see the Online Appendix).

[12] It is important to note that in the analysis, the number of reviews is included as an additive control variable, meaning that there is no interaction term between the standard deviation and the number of reviews. In the Online Appendix, we explore the interaction between the standard deviation and the number of reviews, finding that the negative effect of ratings discrepancy on sales diminishes for products with more reviews. In the main analysis, we focus on the average effect of ratings discrepancy while controlling for the number of reviews, as this allows for a more direct interpretation of the main effect.

[13] Including search position as a control accounts for potential algorithmic penalties in Amazon's ranking system. Thus, the estimated effect of ratings discrepancy on sales is not driven by changes in product visibility due to search reordering, but rather reflects consumer responses holding constant the search rankings.

not fully address price endogeneity, it eliminates bias from time-invariant confounders. The short time window between the two adjacent days for which the difference is taken could strengthen the argument for the causality between the sales change and the ratings discrepancy change. By focusing on a short time interval when calculating the difference, we could reduce the likelihood that time-varying confounding factors, such as promotional activities or modifications to algorithms, substantially change within this limited duration and overturn the outcomes we observe.

Moreover, we use the logarithm of sales rankings as an alternative outcome variable to the sales we imputed earlier. Employing sales rankings as a dependent variable ensures that the results are not solely based on the specific imputation method we use for sales. To address the potential measurement error introduced by using imputed sales and sales rankings as outcome variables, it is important to note that the estimates from (2) are consistent, assuming the measurement error is orthogonal to the covariates. Besides, it is important to acknowledge that the sample, which was collected by scraping data from sales rankings, may not be the most representative of all Amazon products. Due to the limitations of the observational data, we interpret the results regarding the impact of standard deviation of ratings on sales as a local average treatment effect (LATE),[14] rather than generalizing the effect to all products. Further investigation of how the sample selection on sales rankings affects the results is available in the Online Appendix.

Table 3 presents the estimated results for the regression. Columns (1) and (2) illustrate the results of the regression models for the logarithm of sales and the logarithm of sales rankings, respectively, as outcome variables. In these columns, the standard deviation of ratings reveals a substantial impact on both the logarithm of sales and the logarithm of sales rankings. The estimate indicates that a 0.01 increase in the standard deviation of ratings results in a 6.5% decrease in sales and a 4.2% decline[15] in sales rankings, implying reduced sales. To contextualize this magnitude, in the regression sample, the standard deviation of ratings has an interquartile range of 0.78 to 0.98 (see additional descriptive statistics in the Online Appendix). This indicates that the sales of a product at the 25th percentile of ratings discrepancy are approximately 3.7[16] times the sales of a product at

[14]Here the "local average" means the average among those products in the sample, whose rounded average ratings are the same in adjacent days but whose ratings discrepancy could potentially be different. The sample consists of top products from the "Home and Kitchen" category.

[15]Note that a higher ranking number means a lower ranking.

[16]That is, $e^{-6.498\times(0.98-0.78)} = 3.7$.

	Outcomes	
	(1) Log sales	(2) Log sales rankings
Std. ratings	-6.498^{***}	4.211^{***}
	(1.577)	(1.022)
Log price	-1.158^{**}	0.750^{**}
	(0.557)	(0.361)
Covariates	Yes	Yes
N	254	254
R^2	0.194	0.194
Adjusted R^2	0.097	0.097
BIC	465.4	245.0

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 3: Ratings discrepancy harms sales and sales rankings

Notes: This table presents the impact of ratings discrepancy on sales and sales rankings. The outcome variables are (1) the logarithm of sales and (2) the logarithm of sales rankings. The independent variable of interest is the standard deviation of ratings. We control for the logarithm of price, the logarithm of the number of ratings, and search rankings, and we include time fixed effects. Standard errors are clustered at the product level.

the 75th percentile, when holding other covariates the same. Additionally, the coefficients for the price variables are as expected. With a 1% increase in prices, sales would decrease by 1.2%, and sales rankings would decline by 0.8%, consistent with a downward-sloping demand curve.

The results in Table 3 provide evidence that could be seen as consistent with a negative causal relationship between sales and ratings discrepancy. However, one has to be careful to interpret this result. The identification strategy leverages variation from Amazon's rounding system to mitigate some endogeneity concerns, but it does not fully eliminate the possibility that unobserved, time-varying factors, such as the strategic timing of fake reviews acquired by sellers, affect both sales and review distributions. In this sense, the current evidence can only be seen as showing a correlation. We will formally examine the effect of ratings discrepancy on demand in the next sections.

5 Ratings Discrepancy and Consumer Suspicion of Fake Reviews

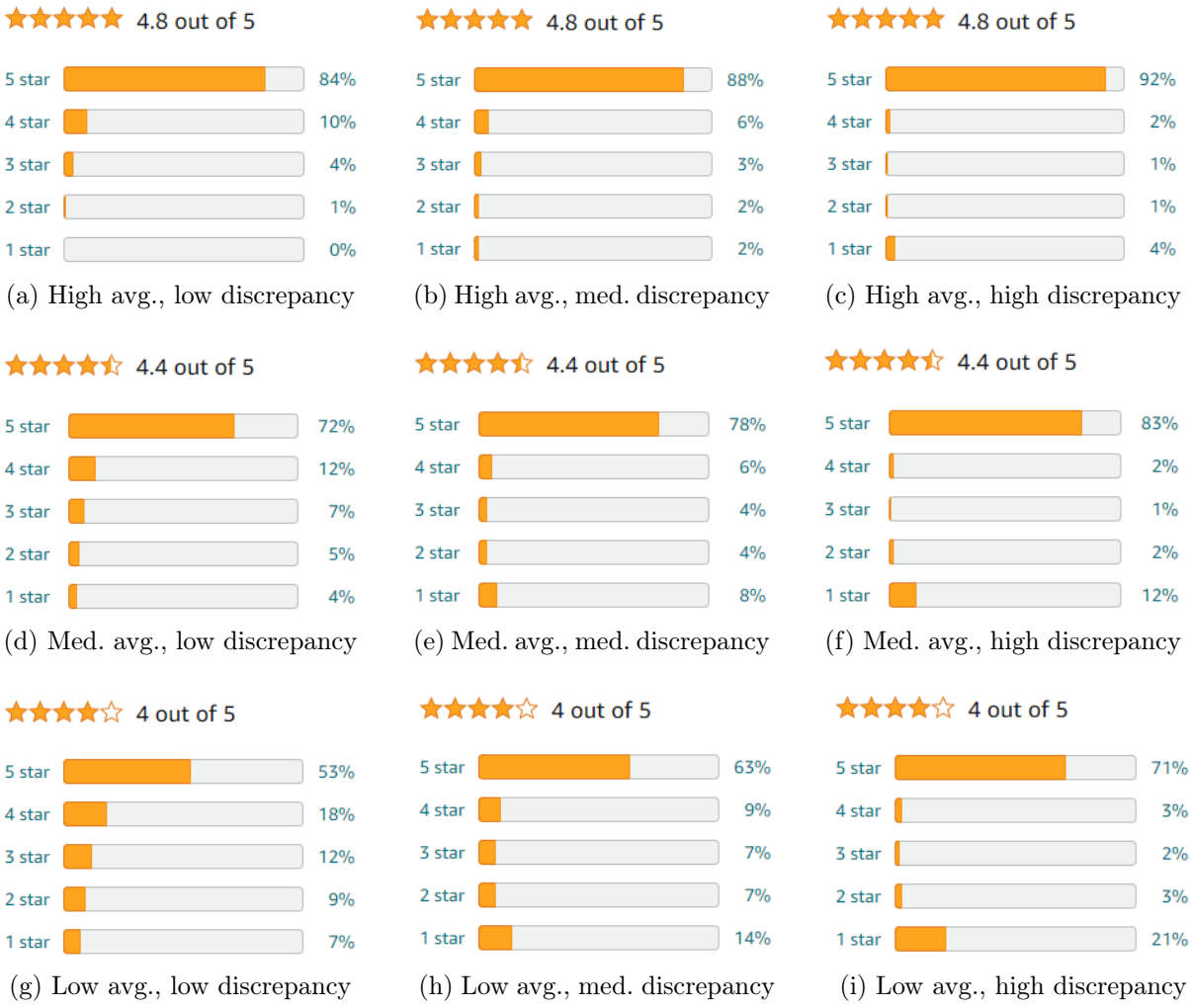
In Section 3, we presented correlational evidence that products with higher ratings discrepancies are more likely to have fake reviews. In Section 4, we showed that higher ratings discrepancies are negatively related to consumer demand. We hypothesize that one mechanism driving this relationship is that ratings discrepancy influences consumer suspicion of fake reviews, which in turn affects demand.

To test this hypothesis, we conduct two experiments. The first experiment, detailed in this section, examines whether ratings discrepancy causally affects consumer suspicion of fake reviews. The second experiment, detailed in the next section, tests whether consumer suspicion of fake reviews causally affects WTP. Together, these studies allow us to establish and quantify the proposed mechanism.

We recruited $N = 812$ participants from Prolific Academic for the first experiment.[17] The experiment mimics closely the online shopping environment where participants viewed product pages of a blender bottle, selected from Amazon's best-selling products. Each participant viewed nine variations of the product page that were identical in every aspect, including appearance, description, and other attributes, except for their rating distributions. The rating distributions, shown in Figure 3, varied along two dimensions: average rating and ratings discrepancy. We selected three levels of average ratings (4.8, 4.4, and 4.0) and for each average rating level, we created three versions with different levels of ratings discrepancy (low, medium, and high), resulting in nine distinct distributions. The product pages are displayed to the participants in a randomized order.

After viewing each product page, participants answered a set of questions measuring their suspicion of fake reviews. First, they completed a multiple-selection question (i.e., they are allowed to select multiple options) on the primary factors in their product evaluation. The reasons presented were drawn from a separate survey we conducted on Prolific and are listed in Appendix C.4. To mitigate ordering effects, the options were presented in randomized order. We then coded a binary variable indicating whether "fake reviews" was among the factors selected. After completing the multiple-selection question, participants reported the perceived probability that the product had fake reviews

[17] We restricted that participants had to have an Amazon account to be eligible for the study.



(a) High avg., low discrepancy (b) High avg., med. discrepancy (c) High avg., high discrepancy

(d) Med. avg., low discrepancy (e) Med. avg., med. discrepancy (f) Med. avg., high discrepancy

(g) Low avg., low discrepancy (h) Low avg., med. discrepancy (i) Low avg., high discrepancy

Figure 3: Product rating distributions in the experiment

Notes: Products on the same row have the same average ratings (4.8, 4.4, and 4.0, respectively), but the levels of discrepancy are different. The standard deviations of the rating distributions are (a) 0.555, (b) 0.779, (c) 0.859, (d) 1.079, (e) 1.234, (f) 1.350, (g) 1.287, (h) 1.497, and (i) 1.655, respectively. These distributions are the probability mass functions of a family of beta-binomial distributions $\text{BetaBin}(4, \alpha\gamma, \beta\gamma)$. Here, the ratio of α to β influences the mean of the distribution, and γ influences the shape of the distribution while holding the mean fixed. We set (α, β) to be $(16, 1)$ for the distributions in the first row, $(6, 1)$ for the second row, and $(6, 2)$ for the third row. The γ's are the same for the distributions in each column, $2^{-2.5}$, 2^{-4}, and 2^{-6}, for the three columns, respectively.

(on a 100-point scale) and the perceived proportion of reviews they believed were fake (on a 100-point scale).

We use the perceived probability of fake reviews as the primary measure of consumer suspicion, with the other two measures serving as robustness checks. While we also collected participants' WTP using the Becker–DeGroot–Marschak (BDM) mechanism, we do not use this data to establish the relationship between suspicion and WTP, as suspicion in this study could be endogenous.[18] Instead, we address this relationship through a separate experiment in Section 6.

After participants evaluated all nine products, there was a second phase of the study, which involved an information treatment that is not the main focus of the main text (see Appendix B). We therefore restrict the data set to only the first phase of the study, and present the second phase only in the appendix for completeness. Additionally, participants allocated 100 points across the different star rating levels to indicate how much attention they paid to each level when evaluating products. Analysis of this attention allocation, presented in the Online Appendix, provides supporting evidence that consumers actively consider ratings discrepancy in their decision-making process.

	Variable			
Product	(1) Perceived fake review probability (%)	(2) Perceived fake review proportion (%)	(3) List fake review as a factor	(4) WTP
High average, low discrepancy	36.207	24.477	0.128	15.342
High average, med. discrepancy	34.821	24.384	0.128	15.342
High average, high discrepancy	39.840	28.448	0.175	15.389
Med. average, low discrepancy	31.650	21.151	0.100	14.521
Med. average, med. discrepancy	35.037	23.615	0.154	14.321
Med. average, high discrepancy	41.841	29.680	0.219	13.606
Low average, low discrepancy	30.296	21.367	0.138	12.788
Low average, med. discrepancy	36.607	25.985	0.188	12.845
Low average, high discrepancy	44.674	33.067	0.261	12.324

Table 4: Summary statistics by rating distributions

Notes: This table presents average responses across products with different rating distributions. For each product, we collected: (1) participants' perceived probability that the product has fake reviews (0–100 scale), (2) perceived proportion of fake reviews (0–100 scale), (3) whether fake reviews were mentioned as a factor in evaluation (binary), and (4) WTP elicited using the BDM mechanism. Products are grouped by their average rating (high= 4.8, medium= 4.4, low= 4.0) and level of ratings discrepancy.

Table 4 presents model-free evidence that higher ratings discrepancy is associated with increased suspicion of fake reviews. For products with the same average rating, higher ratings discrepancy consistently corresponds to (i) a higher perceived probability of fake reviews (column (1)), (ii) a higher perceived proportion of fake reviews (column (2)), and

[18] See the Online Appendix for an analysis of this relationship and a mediation analysis with the three measures of consumer suspicion as moderators, assuming that consumer suspicion is exogenous.

(iii) a higher probability of listing fake reviews as a factor (column (3)). This pattern is particularly pronounced for products with low average ratings. For example, when the average rating is low (4.0), increasing ratings discrepancy from low to high raises the mean perceived probability of fake reviews from 30.3% to 44.7%. The data also shows that higher ratings discrepancy corresponds to lower WTP (column (4)), consistent with the proposed mechanism.

	Outcomes		
	(1) Perceived fake review probability (%)	(2) Perceived fake review proportion (%)	(3) List fake review as a factor
Avg. ratings	20.808***	15.511***	0.141***
	(1.726)	(1.443)	(0.027)
Std. ratings	21.750***	17.461***	0.216***
	(1.593)	(1.332)	(0.025)
Individual fixed effects	Yes	Yes	Yes
N	7,308	7,308	7,308
R^2	0.527	0.474	0.403
Adjusted R^2	0.468	0.408	0.328
BIC	70,484	67,860	9,759

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 5: Ratings discrepancy increases consumer suspicion of fake reviews

Notes: This table presents regression estimates of the relationship between ratings discrepancy and three measures of consumer suspicion of fake reviews: (1) participants' perceived probability that the product has fake reviews (0–100 scale), (2) perceived proportion of fake reviews (0–100 scale), (3) whether fake reviews were mentioned as a factor in evaluation (binary). The independent variable is ratings discrepancy, measured by the standard deviation of ratings. All specifications include individual fixed effects and control for average ratings. Standard errors are reported in parentheses.

To formally quantify these relationships, we estimate regression models with the three measures of consumer suspicion as dependent variables (Table 5).[19] Each specification includes individual fixed effects to control for participant-level heterogeneity and controls for average ratings to isolate the effect of ratings discrepancy. The results show that ratings discrepancy, measured by the standard deviation of ratings,[20] has a positive and economically significant relationship with all three measures of consumer suspicion.[21]

[19]We estimate a similar regression model with the WTP as the dependent variable and conduct a mediation analysis with the three measures of consumer suspicion as moderators. The results are presented in the Online Appendix.

[20]Using alternative measures of ratings discrepancy, e.g., (i) the proportions of 1-star ratings, or (ii) the proportions of 5-star and 1-star ratings and its interaction, instead of the standard deviation of ratings leads to the same substantive results as those presented here (see the Online Appendix).

[21]See the Online Appendix for estimates allowing the effect of ratings discrepancy to vary with average

Focusing on the primary measure, we find that a 1.0 increase in the ratings standard deviations leads to a 21.8 percentage point increase in the perceived probability of fake reviews, holding average ratings constant (column (1) of Table 5). This result is robust to using alternative measures of consumer suspicion (columns (2) and (3) of Table 5).[22]

To provide context for the magnitude of these effects, the standard deviations of the nine rating distributions shown to participants ranged from 0.555 to 1.655. Within each average rating level, the range (highest minus lowest) of the standard deviations for products is approximately 0.3. This implies that for each average rating level, the product with the most discrepant ratings can lead to a 6.5 percentage point higher perceived probability of fake reviews than the product with the least discrepant ratings, representing an economically significant effect.

6 Consumer Suspicion and Willingness to Pay

In Section 5, we showed that ratings discrepancy affects consumer suspicion of fake reviews. However, that experiment does not cleanly identify how this suspicion impacts WTP, as consumer suspicions could be endogenous. For example, consumers' prior experience with certain product categories might simultaneously affect both their suspicion of fake reviews and their WTP. To establish the causal impact of suspicion on WTP, we conducted a second experiment that directly manipulates participants' beliefs about the presence of fake reviews.

We recruited $N = 792$ participants from Prolific Academic for this experiment. To maintain consistency with the first experiment, we used the same product, a blender bottle, and a similar online shopping interface as in the experiment detailed in Section 5. However, we omitted the rating distribution display to avoid confounding the manipulation of fake review suspicion with participants' reactions to the distribution directly.

The study started with a baseline measurement phase, in which all participants were

ratings.

[22]There could be concerns that the experimental design overestimates consumer suspicion of fake reviews. Specifically, after participants assess their first product, being asked repeatedly about fake reviews could heighten their attention to this issue and inflate their reported suspicion in later evaluations. To address this concern, we conduct a robustness check using only responses to the first product each participant evaluated (see the Online Appendix). The results from this analysis, while showing more conservative effect sizes, remain directionally consistent with the main findings, suggesting that the relationship between ratings discrepancy and consumer suspicion is robust.

familiarized with the product page and were asked to report their WTP using the BDM mechanism and their perceived probability of fake reviews. This initial measurement served as a benchmark for consumer suspicion of fake reviews before any experimental manipulation. Following this, in the treatment phase, participants were randomly assigned to one of three treatment conditions, varying in the indicated probability of fake reviews: high, medium, or low probability. The participants were then shown the same product page as in the baseline measurement phase, but with a platform warning label displaying the assigned probability level.[23] Participants then reported their WTP and perceived probability of fake reviews again.

	Variable	
Condition	(1) WTP	(2) Perceived fake review probability (%)
High probability	15.266	50.663
Medium probability	18.139	40.608
Low probability	20.008	29.234

Table 6: Experimental manipulation on consumer suspicion and WTP

Notes: This table presents average outcomes by treatment condition after the experimental manipulation. Treatment conditions varied in the displayed probability of fake reviews (high, medium, and low). Column (1) shows WTP elicited using the BDM mechanism. Column (2) shows participants' perceived probability that the product had fake reviews (0–100 scale).

Table 6 presents evidence that the experimental manipulation successfully affected both participants' beliefs and their WTP. When the label provided to the participants changes from "Low Probability" to "High Probability," the perceived probability of fake reviews increased by 21.4 percentage points (from 29.2% to 50.6%), and the average WTP decreased by \$4.7 (from \$20.0 to \$15.3). These differences are economically significant compared to pre-manipulation levels, where participants reported an average perceived probability of 44.3% and WTP of \$19.9.

To quantify the causal impact of fake review suspicion on WTP, we estimate a linear regression using the experimental variation (Table 7). The results show that each percentage point increase in the perceived probability of fake reviews reduces WTP by 4.4 cents. This implies that shifting from complete trust (0% probability) to complete distrust (100% probability) would reduce WTP by \$4.4, representing a 22% decrease

[23]This setup replicates the functionality of some browser extensions that estimate and display the likelihood of fake reviews to consumers, which Amazon could potentially integrate in the future.

	Outcome: WTP
Perceived fake review probability (%)	-0.044^{***} (0.014)
N	792
R^2	0.013
Adjusted R^2	0.011

$^{*}p$ <0.1; $^{**}p$ <0.05; $^{***}p$ <0.01

Table 7: Perceived probability of fake reviews negatively affects consumer WTP

Notes: This table presents the estimated causal effect of consumer suspicion of fake reviews on WTP. The independent variable is the perceived probability of fake reviews (0–100 scale), which was experimentally manipulated across treatment conditions. WTP was elicited using the BDM mechanism. Standard errors are reported in parentheses.

relative to the mean pre-manipulation WTP of \$19.90.

By integrating the findings of this experiment with those presented in Section 5, we find that larger ratings discrepancies increase consumer suspicion of fake reviews, and this increased suspicion, in turn, leads to a reduction in WTP. As a back-of-the-envelope calculation of the effect size of this channel, recall that Section 5 shows that when controlling for average ratings, a 1.0 increase in the ratings standard deviations increases the perceived probability of fake reviews by 21.8 percentage points. Combined with the estimate that each percentage point of perceived probability of fake reviews reduces WTP by 4.4 cents, this suggests that a 1.0 increase in the ratings standard deviations could reduce WTP by approximately \$0.96[24] through the fake review suspicion channel. To put this number into context, when using the data of the first experiment alone, we find that when controlling for average ratings, a 1.0 increase in the ratings standard deviations reduces WTP by \$0.84 (see the Online Appendix).[25] This suggests that the channel of fake reviews is substantial.

We should also acknowledge that this evidence just shows that suspicion of fake reviews is a mechanism by which ratings discrepancy may affect sales, and there may be other mechanisms by which ratings discrepancy may affect sales. We provide further discussion on this issue below.

[24]That is, $21.8 \times \$0.044 = \$0.096$.

[25]The indirect effect of 0.96 exceeds the total effect of 0.84. One explanation could be that ratings discrepancies may influence WTP through additional channels beyond fake review suspicion, such as the perceived heterogeneity in buyers' preferences. See the Online Appendix for a formal mediation analysis.

7 Managerial Implications

The implications of the findings primarily concern two key stakeholders: sellers considering the acquisition of fake reviews and online platforms seeking to regulate their marketplace.

For sellers, the findings reveal an important trade-off in the decision to acquire fake reviews. When a product first launches, while fake reviews can boost its average rating, they may also increase the overall discrepancy in ratings. This effect is especially pronounced if they trigger retaliatory reviews from consumers who feel misled. Although a higher average rating usually stimulates consumer demand, a resulting increase in discrepancy in ratings can diminish it. Moreover, sellers should be particularly cautious about acquiring fake reviews when their products already have some negative reviews, as the resulting increase in ratings discrepancy could outweigh any benefits from a higher average rating.

For online platforms such as Amazon, the process of identifying and removing fake reviews could take about six months. This delay could be due to limited detection capabilities, or it might involve strategic considerations that explain why Amazon deliberately takes this approach. Amazon also competes with other platforms. If consumers favor products with more reviews on competing platforms like eBay, Amazon might permit fake reviews to exist for some time, particularly for new products. Consequently, policy implications may not always suggest that stricter policies are universally beneficial for the platform.

Furthermore, Amazon could emphasize reviews authored by users with more extensive and credible histories. While assigning different weights to reviews from various users may be counterproductive if not carefully executed, a well-designed algorithm could alleviate the fake review issue by calculating weighted ratings without requiring the complete removal of reviews. Considering that consumers presently submit 1-star ratings in response to being deceived by fake reviews, Amazon could develop a more user-friendly mechanism for customers to report suspicious reviews and express their dissatisfaction in a more constructive manner rather than through retaliatory actions. Amazon could then investigate these reports or assign less weight to the reviews or users implicated in fraudulent behavior.

Moreover, Amazon can educate consumers in a manner akin to an information treat-

ment (considered in the Appendix). By equipping consumers with the knowledge and tools to identify fake reviews, Amazon can enable them to make better-informed purchasing decisions. This educational strategy could involve incorporating tutorials, guidelines, or interactive features within the platform, helping users become more adept at detecting fake reviews. As consumers become more discerning, sellers might be less tempted to acquire fake reviews, which consequently reduces the prevalence of such practices. In the long term, this could lead to increased consumer trust in the platform and a healthier online marketplace.

In conclusion, the research findings provide insights for both online platforms and sellers, emphasizing the necessity of a multi-faceted approach to addressing the fake review problem. Platforms like Amazon could consider enacting stricter policies regarding identified fake reviews, refining their algorithms to give precedence to reviews from credible users, and investing in consumer education. Meanwhile, sellers should be mindful of the potential risks and adverse effects of purchasing fake reviews, as consumer perceptions of their products may suffer once they are identified as having fake reviews. By considering these factors, both platforms and sellers can contribute to the development of a more trustworthy and transparent online shopping environment.

8 Future Directions

The present research offers insights into the relationship between ratings discrepancy, consumer suspicion of fake reviews, and consumer demand. However, there are several aspects that need further investigation and potential avenues for future research.

Firstly, we find it challenging to establish a causal relationship between the acquisition of fake reviews and the increase in ratings discrepancy, as the decision to acquire fake reviews is inherently endogenous. The experiment focuses on quantifying people's suspicion of fake reviews after observing rating distributions with higher discrepancies, which is a distinct issue. Future research that addresses the endogeneity problem using various approaches to identify the impact of fake review acquisition on ratings discrepancy would be valuable, but beyond the scope of this paper.

Additionally, it is also important to acknowledge that due to the limitations of the data, the results from the observational study primarily focus on products with higher

sales compared to the overall population of Amazon products. The impact of changes in the standard deviation on consumer demand is more linked to a local average treatment effect. It is unlikely that consumers directly observe and make decisions based on the small changes in the rating distribution between adjacent days. Instead, we hypothesize that heterogeneous consumers enter the market in each period, and changes in the ratings discrepancy influence the decisions of some marginal consumers who arrive in the market during that specific period when ratings discrepancies change while the average rating stays the same. Compared with the counterfactual scenario, these marginal consumers may adjust their purchase decisions as the changes take effect. Further studies are needed to explore whether the size of these effects would be comparable for products with lower popularity, such as those that rarely attract new customers. Another limitation of the data set is that for the data containing fake review labels, we do not have access to panel data. On the other hand, for the data from which we infer sales, although we have sales information over time, we lack the corresponding fake review labels. Due to the differences in the time spans covered by the sales data and the fake review label data, we are unable to observe the dynamics of sales patterns for products with fake reviews. It would be valuable if future research could analyze these dynamics in more detail, as this could provide additional evidence on how consumers might identify products with fake reviews.

It is important to acknowledge that ratings discrepancies could affect consumer demand through multiple channels, in addition to the suspicion of fake reviews. For instance, ratings discrepancies could signal underlying product quality uncertainty, and risk-averse consumers may reduce their demand in response to this uncertainty. While the experiments establish consumer suspicion of fake reviews as one of the mechanisms linking ratings discrepancy to reduced demand, this may not be the only channel by which ratings discrepancy affects demand. Future research using natural experiments or field interventions that specifically target fake review concerns versus general uncertainty could help disentangle these different channels.

Moreover, while the experiments demonstrate that consumers use ratings discrepancies as a signal of fake reviews, gathering direct field evidence of these concerns would be valuable. Future research could analyze review text and social media discussions to identify cases where consumers explicitly express suspicion of fake reviews, particularly

in relation to rating patterns. Researchers could also conduct broader surveys of Amazon customers about their awareness and perceptions of fake reviews in natural shopping contexts. Such field evidence would complement the experimental findings by showing how consumers express and act on these concerns in real marketplace settings.

While the experiments establish that consumer suspicion of fake reviews is one mechanism linking ratings discrepancy to reduced demand, finding direct field evidence of this mechanism remains challenging. Future research could strengthen this link by identifying exogenous shocks to consumer concerns about fake reviews, such as through media coverage or platform policy changes. Additionally, researchers could examine whether the relationship between ratings discrepancy and sales varies with factors that specifically affect fake review acquisition, such as product prices or category-specific review monitoring policies. Such field evidence would complement the experimental findings by demonstrating how this mechanism operates in natural marketplace settings.

Additionally, the main focus of this paper is on fake positive reviews posted by sellers about their own products. Considering fake negative reviews as part of a broader strategic interaction among sellers is an interesting direction for future research.

Lastly, while this research focuses on the implications for platforms and sellers, it is important to consider the broader societal impact of fake reviews. Future research could investigate the consequences of fake reviews on consumer trust in online marketplaces and the potential spillover effects on other industries or sectors. Additionally, future research might investigate how government regulations and industry standards tackle fake reviews, or more broadly, fake information such as fake news, while comparing the effectiveness of various approaches across different countries or regions.

References

Amaldoss, W. and Shin, W. (2025). Content provision on UGC platforms. *Management Science*.

Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78.

Chakraborty, I., Deb, J., and Oery, A. (2022). When do consumers talk? *Available at SSRN 4155523*.

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

Chevalier, J. and Goolsbee, A. (2003). Measuring prices and price competition online: Amazon.com and barnesandnoble.com. *Quantitative Marketing and Economics*, 1:203–222.

Chevalier, J. A., Dover, Y., and Mayzlin, D. (2018). Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science*, 37(5):688–709.

De Langhe, B., Fernbach, P. M., and Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, 42(6):817–833.

Federal Trade Commission (2022). FTC to explore rulemaking to combat fake reviews and other deceptive endorsements.

He, S., Hollenbeck, B., Overgoor, G., Proserpio, D., and Tosyali, A. (2022a). Detecting fake-review buyers using network structure: Direct evidence from amazon. *Proceedings of the National Academy of Sciences*, 119(47):e2211932119.

He, S., Hollenbeck, B., and Proserpio, D. (2022b). The market for fake reviews. *Marketing Science*, 41(5):896–921.

He, S. X. and Bond, S. D. (2015). Why is the crowd divided? Attribution for dispersion in online word of mouth. *Journal of Consumer Research*, 41(6):1509–1527.

Karaman, H. (2021). Online review solicitations reduce extremity bias in online review distributions and increase their representativeness. *Management Science*, 67(7):4420–4445.

Li, L., Tadelis, S., and Zhou, X. (2020). Buying reputation as a signal of quality: Evidence from an online marketplace. *The RAND Journal of Economics*, 51(4):965–988.

Li, X. and Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4):456–474.

Luca, M. (2016). Reviews, reputation, and revenue: The case of yelp.com. *Harvard Business School NOM Unit Working Paper*, (12-016).

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427.

Luo, L. and Ma, L. (2024). Wisdom of the AI crowd? can we detect AI-generated product reviews? *Available at SSRN 4511025*.

Luo, X., Raithel, S., and Wiles, M. A. (2013). The impact of brand rating dispersion on firm value. *Journal of Marketing Research*, 50(3):399–415.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–2455.

Moe, W. W. and Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3):444–456.

Park, S., Shin, W., and Xie, J. (2021). The fateful first consumer review. *Marketing Science*, 40(3):481–507.

Pei, A. and Mayzlin, D. (2022). Influencing social media influencers through affiliation. *Marketing Science*, 41(3):593–615.

Proserpio, D. and Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5):645–665.

Reimers, I. and Waldfogel, J. (2021). Digitization and pre-purchase information: The causal and welfare impacts of reviews and crowd ratings. *American Economic Review*, 111(6):1944–71.

Rozenkrants, B., Wheeler, S. C., and Shiv, B. (2017). Self-expression cues in product rating distributions: When people prefer polarizing products. *Journal of Consumer Research*, 44(4):759–777.

Schoenmueller, V., Netzer, O., and Stahl, F. (2020). The polarity of online reviews: Prevalence, drivers and implications. *Journal of Marketing Research*, 57(5):853–877.

Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58(4):696–707.

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

Xu, L., Wei, Y. and Zhu, Y. (2025). Customer voice on two-sided platforms: The effect of surge pricing on customer complaints. *Management Science*, 71(8):6729–6745.

Zhong, Z. (2022). Chasing diamonds and crowns: Consumer limited attention and seller response. *Management Science*, 68(6):4380–4397.

Zhou, X. and Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.

A A Theoretical Framework about Ratings Discrepancy

Consider the following model, which characterizes sellers' equilibrium fake review acquisition decisions and consumers' decision-making process. This is just an illustration of some effects at work.

There are two periods. At the beginning of period 1, a seller enters the model with no ratings. The seller's product quality can be either high or low, with probability p for the high type and $1-p$ for the low type. Consumers in both periods do not observe the product quality prior to purchase.

In period 1, some short-lived consumers enter the model. Some of them purchase the product from the seller and leave ratings. We normalize the mass of (a continuum of) ratings to be one.[26] The ratings have three levels, denoted as 1-star, 2-star, and 3-star, respectively. A high-type seller will receive α 3-star ratings, β 2-star ratings, and $1-\alpha-\beta$ 1-star ratings. Denote this rating distribution as $(1-\alpha-\beta, \beta, \alpha)$. We assume the proportion of 3-star ratings, α, is sufficiently high.[27] Moreover, a low-type seller's rating distribution at the end of period 1 is $(1-\alpha'-\beta', \beta', \alpha')$, where $\alpha' < \alpha$ and $\beta' = \beta$.

Between period 1 and period 2, the seller decides whether to acquire fake reviews. If so, it may acquire fake reviews of any star level. We focus on the scenarios where a low-type seller could potentially mimic a high-type seller by acquiring fake reviews. Each unit of fake reviews costs c for the seller. The cost c is stochastic, realized before the seller's fake review acquisition, and can only be observed by the seller. The cumulative distribution function of c is F.

In period 2, two types of consumers, naïve consumers and sophisticated consumers, enter the model. Naïve consumers only observe the average ratings and infer the seller's type based on it, while sophisticated consumers observe the full rating distribution. The fraction of naïve consumers is γ, and the fraction of sophisticated consumers is $1-\gamma$.[28] If

[26] We assume that the mass of original ratings is the same for both the high-type and low-type seller, and that the mass of ratings is not used by consumers to assess the type of the seller. This could be justified with sufficient uncertainty about the mass of ratings, but formally considering this is beyond the scope of this illustrative framework.

[27] Specifically, we assume $\alpha \geq 1 - \frac{2-\gamma}{2(1-\gamma)}\beta$, where γ is the fraction of naïve consumer specified below.

[28] An alternative interpretation of the consumer types could be that, even the same consumer may choose to only look at the average ratings at some time, and look at the full rating distribution at other times. Then, γ can be interpreted as the frequency consumers only look at the average ratings.

a consumer has a posterior belief μ of the seller being a high-type, the consumer's demand is $D(\mu)$ and the seller's profit from this consumer is $\Pi(\mu)$. Both D and Π are monotone in μ. Denote the naïve consumers' posterior belief of the seller being a high-type as μ^{N} and the sophisticated consumers' posterior belief as μ^{S}, the seller's payoff can then be expressed as $\gamma\Pi(\mu^{\mathrm{N}}) + (1-\gamma)\Pi(\mu^{\mathrm{S}})$. If the mean of ratings is different than the one of a high-type seller, both types of consumers believe that it is a low-type seller with probability one. If the distribution of ratings is different than the one of a high-type seller, the sophisticated consumers believe that it is a low-type seller with probability one.

We first characterize how the sellers could potentially acquire fake reviews, specifically, which star levels and how many fake reviews they acquire.

Intermediate Result: A low-type seller's equilibrium strategy can only be one of the following:

1. **Match Exact Distribution**: It acquires a mass of κ_2^{D} 2-star ratings and κ_3^{D} 3-star ratings (presented in the Online Appendix) so that its rating distribution is the same as the high-type seller.

2. **Match Mean**: It acquires a mass of κ^{M} 3-star ratings (presented in the Online Appendix) so that its average rating is the same as the high-type seller. In this case, its rating distribution becomes a mean-preserving spread of $(1-\alpha-\beta, \beta, \alpha)$, which we denote as $(1-\widetilde{\alpha}-\widetilde{\beta}, \widetilde{\beta}, \widetilde{\alpha})$.

3. **No Fake Reviews**: It chooses not to acquire any fake reviews.

Moreover, $\kappa^{\mathrm{M}} < \kappa_2^{\mathrm{D}} + \kappa_3^{\mathrm{D}}$, which means that the fake review acquisition cost for "Match Mean" is lower than that of "Match Exact Distribution."

The proof is presented in the Online Appendix. To see the intuition that if a low-type seller chooses "Match Mean," its rating distribution becomes a mean-preserving spread of the high-type seller's rating distribution, note, first, that a low-type seller receives more 1-star negative ratings in period 1. Second, if it chooses "Match Mean," it acquires only 3-star ratings. Both of these factors make its rating distribution more discrepant than the high-type seller's rating distribution.

We now turn to a characterization of the equilibrium seller strategy and consumer decisions. Note that a high-type seller's rating distribution is always $(1-\alpha-\beta, \beta, \alpha)$, so we omit this in the following result.

Result 1: There exists a unique equilibrium of the game. The seller's equilibrium fake review acquisition decisions can be characterized as below. There exist two cutoffs, $\underline{c} \leq \overline{c}$, for the realization of the cost c, such that:

1. When $c > \overline{c}$, a low-type seller chooses "No Fake Reviews," and its rating distribution remains $(1-\alpha'-\beta', \beta', \alpha')$.

2. When $\underline{c} < c \leq \overline{c}$, a low-type seller chooses "Match Mean," and its rating distribution becomes $(1-\widetilde{\alpha}-\widetilde{\beta}, \widetilde{\beta}, \widetilde{\alpha})$.

3. When $c \leq \underline{c}$, a low-type seller chooses "Match Exact Distribution," and its rating distribution becomes $(1-\alpha-\beta, \beta, \alpha)$.

The proof is presented in the Online Appendix. The intuition for the result is that the cost of acquiring fake reviews could play an important role in the seller's fake review acquisition strategy. If the cost is very high, it will not be profitable for a low-type seller to acquire fake reviews. On the other extreme, if the cost is very low, it will be not very costly for a low-type seller to match a higher-type seller's exact rating distribution. If the cost is intermediate, it can be costly for a low-type seller to match a higher-type seller's exact rating distribution, but matching a higher-type seller's average ratings could still be profitable.

This theoretical framework illustrates the empirical analyses. Specifically, it can generate hypotheses regarding the relationship between the rating distribution and consumer demand. There are three possible rating distributions on the equilibrium path, corresponding to those specified in Result 1. The demand for a product with rating distribution $(1-\alpha'-\beta', \beta', \alpha')$ would be the lowest since both types of consumers infer that the seller is a low-type. Moreover, comparing products with rating distribution $(1-\widetilde{\alpha}-\widetilde{\beta}, \widetilde{\beta}, \widetilde{\alpha})$ and $(1-\alpha-\beta, \beta, \alpha)$, the naïve consumers have the same demand for the two products, and the sophisticated consumers have a higher demand for the latter product, so the latter product has the highest demand. We summarize this in the following result.

Result 2: On the equilibrium path, the total consumer demand for a product has the following relationship with the rating distribution:

1. The total consumer demand for a product increases with the product's average rating.

2. Holding the average ratings constant, the total consumer demand decreases in the level of discrepancy, in terms of second-order stochastic dominance. In addition, the mechanism is that sophisticated consumers' suspicion of fake reviews increases with the level of discrepancy.

In the main empirical analyses, we employ the standard deviation of the ratings as a measure of ratings discrepancy, which is consistent with second-order stochastic dominance. In the Online Appendix, we consider additional measures of ratings discrepancy, which are also consistent with second-order stochastic dominance.

B Information Treatment

This section details the second phase of the study in Section 5, which involves a randomized information treatment.

B.1 Experimental Design

In the second phase of the study, participants were randomly assigned to a treatment ("information") group and a control ("placebo") group. The information group was provided with an informational briefing highlighting the findings that products with more discrepant ratings tend to have a higher probability of having fake reviews. This information treatment is presented via a format akin to a browser-extension-style textbox (see details in Figure 4).[29]

After being exposed to this additional information, the same set of questions as in the first phase is posed to the treatment group. In contrast, the control group will also be asked to respond to the same set of questions once more without being exposed to the additional information. The differences in responses between the treatment and control

[29]However, tailoring the message with specific product attributes and customer characteristics could potentially result in a more impactful change in welfare, which we leave for future research.

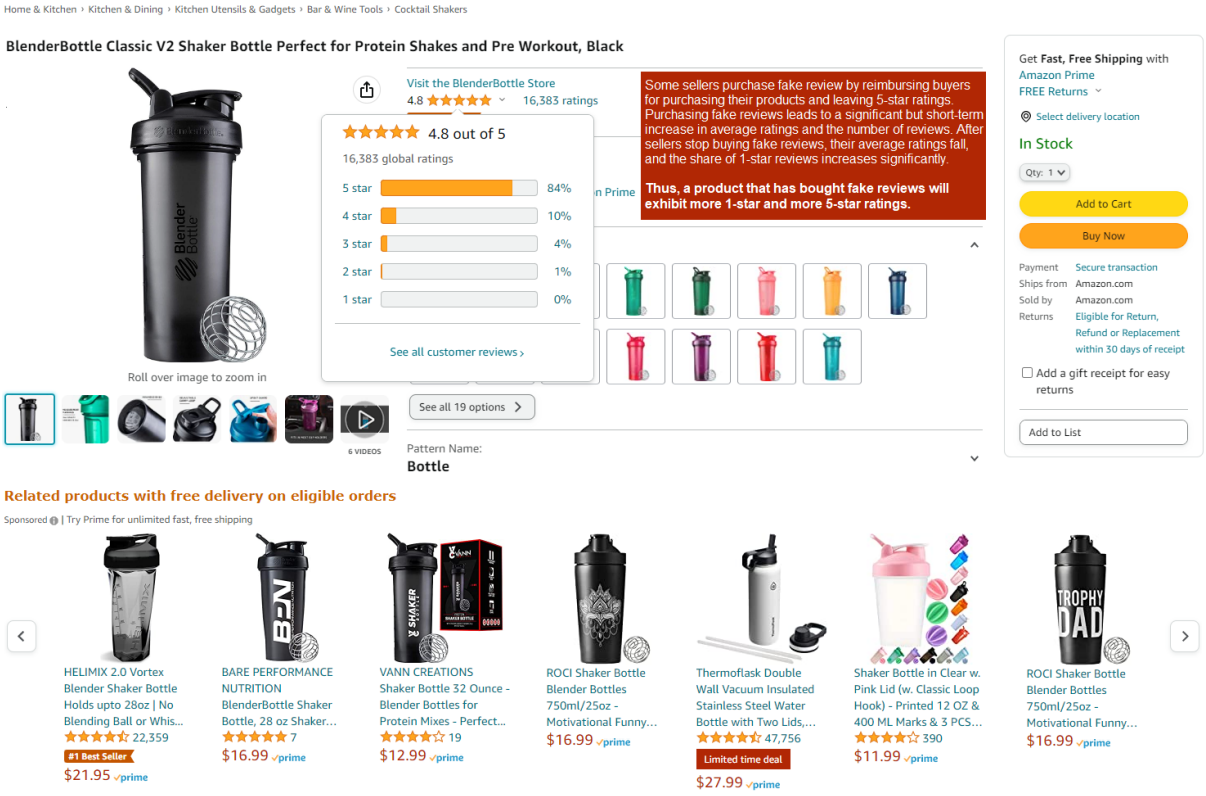


Figure 4: Information treatment in the experiment

groups can then be compared, enabling us to understand the information treatment's impact on the participants' decision-making process and their WTP.

		Treatment condition		
Variable	(1) Both	(2) Information	(3) Placebo	(4) p-value
Female	0.47	0.46	0.48	0.513
Age	40.30	40.22	40.38	0.870
Black	0.09	0.09	0.09	0.707
Education $\geq$ bachelor	0.58	0.58	0.59	0.823
Income $\geq$ \$75,000/year	0.43	0.42	0.45	0.494
# Review websites	2.19	2.11	2.27	0.095
# Years on Amazon $\geq$ 10	0.46	0.47	0.44	0.290
Use Amazon $\geq$ 1/week	0.42	0.42	0.42	0.898
N	812	401	411	

Table 8: Balance tests of the experiment

Notes: In column (4), we test the hypothesis of H_0: the mean of the variable in column (2) = (3) against H_1: the mean of column (2) $\neq$ (3), using a t-test, for each variable respectively.

Table 8 provides the results of balance tests between the treatment and control groups. The variables considered as observables in this study include gender, age, ethnicity, ed-

ucation level, income, and experience with both Amazon and other review platforms. When examining education, income, years on Amazon, and Amazon usage frequency, these variables are compared to the median value found within the data set. The analysis of demographic observables provides supportive evidence that the treatment group and control group are similar, especially considering that multiple testing could lead to some small p-values.

B.2 Average Treatment Effects

		Treatment condition			
	(1)	(2)	(3)	(4)	(5)
Product	Both	Information	Placebo	(2)−(3)	p-value
High average, low discrepancy	15.076	15.237	14.920	0.317	0.577
High average, med. discrepancy	14.756	14.825	14.689	0.137	0.805
High average, high discrepancy	14.685	14.449	14.915	−0.466	0.412
Med. average, low discrepancy	13.900	14.177	13.630	0.547	0.291
Med. average, med. discrepancy	13.560	13.459	13.659	−0.201	0.697
Med. average, high discrepancy	12.344	11.574	13.095	−1.521	0.003
Low average, low discrepancy	12.395	12.870	11.932	0.938	0.072
Low average, med. discrepancy	11.818	11.576	12.054	−0.477	0.331
Low average, high discrepancy	11.089	10.663	11.504	−0.840	0.093
N	812	401	411		

Table 9: Summary statistics of willingness-to-pay (WTP) in the experiment

Notes: Participants were randomly assigned to an "information" condition that received an information treatment and a "placebo" condition that received a placebo treatment. In column (4), we calculate the difference between the variables in columns (2) and (3). In column (5), we test the hypothesis of H_0: the mean of the variable in column (2) = (3) against H_1: the mean of column (2) $\neq$ (3), using a t-test, for each variable respectively.

Table 9 presents the summary statistics of the experiment for different products in the second phase. Recall that products are categorized by three levels of average ratings (high, medium, and low) and by three levels of ratings discrepancy (low, medium, and high). The table underscores a prevailing preference among participants for products with high average ratings and low ratings discrepancy.

There are noticeable differences between the treatment and control groups for certain rating distributions. Given the balance between the groups, this variation in the post-treatment phase can be interpreted as an estimate of the average treatment effect (ATE) of the information treatment. It is noteworthy that the ATE is larger in magnitude for products with lower average ratings. This can be attributable to the inherently minor

differences in discrepancy for products with high average ratings.

When holding the average constant, the presence of the information treatment has varying effects on individuals' WTP depending on the ratings discrepancy in the products. In cases where the products have higher ratings discrepancies, the information treatment leads to a reduction in participants' WTP. For instance, for the "low average, high discrepancy" product, compared to individuals who did not receive the information treatment, those who received the information treatment experienced a decrease of \$0.8 in their WTP, which represents around 8% of the WTP. Conversely, for products with low ratings discrepancy, the information treatment could lead to an increase in WTP. This is particularly evident in the "low average, low discrepancy" case, where the inherent uncertainty associated with a low average is alleviated by the information treatment, resulting in an economically significant boost in WTP. In fact, for the "low average, low discrepancy" scenario, the presence of the information treatment increased WTP by \$0.9, representing around 8% of the total value. The interpretation of these findings suggests that, when a product has a lower average rating, individuals find themselves in a "pooling" state, where uncertainty exists regarding whether the product has been subjected to fake reviews. Consequently, their WTP for such products is lower. However, once this uncertainty is addressed by the information treatment, individuals transition to a "separating" state where the WTP for products with low ratings discrepancy increases, while the WTP for products with high ratings discrepancy experiences a substantial decrease. The magnitude of these effects is negatively correlated with the average rating.

C Choice Experiment

The main difference between the experiment in this section and in the last section is the outcome we measure. The experiment in this section entails a conjoint analysis.

C.1 Experimental Design

During the experiment, participants were presented with three products with the same average ratings but varying rating distributions. Products were displayed in pairs. They are identical in every aspect, including average rating, price, appearance, and other product attributes, with the sole difference being the rating distribution. For each product



(a) Product L (b) Product M (c) Product H

Figure 5: Product rating distributions in the choice experiment

Notes: The products have the same average ratings (4.4/5.0), but the rating distributions are different. The standard deviations of the product ratings are 1.079, 1.234, and 1.350, respectively. These distributions are the probability mass functions of a family of beta-binomial distributions $\text{BetaBin}(4, 6\gamma, \gamma)$. Here, γ is a parameter that influences the shape of the distribution while holding the mean fixed. The γ's for the three distributions are $2^{-2.5}$, 2^{-4}, and 2^{-6}, respectively.

Here is a summary of an *Economists* ariticle. The article is based on a research paper (He, Hollenbeck, & Proserpio, 2022):

Title: A new study analyses the murky world of fake Amazon reviews

Summary: Companies can temporarily boost sales by paying for fake reviews on Amazon. Researchers identified 1,500 products with manipulated ratings through Facebook groups, where companies promised to reimburse buyers for purchasing their products and leaving **five-star** ratings. Purchasing fake reviews leads to a significant but short-term increase in average ratings and the number of reviews. After firms stop buying fake reviews, their average ratings fall, and the share of **one-star** reviews increases significantly.

Figure 6: Information treatment in the choice experiment

pair, participants chose between products with the two rating distributions. Figure 5 displays the three distributions, resulting in three groups of comparisons. We henceforth refer to them as products L, M, and H. Note that product H has the most discrepant ratings, while product L has the least discrepant ratings. Note that these rating distributions are also different in other aspects, e.g., the proportion of 1-star ratings. In Appendix D, we present the results of an experiment, where the effect of the proportion of 1-star ratings is investigated.

The participants were randomly allocated to a treatment ("information") group and a control ("placebo") group. The information group was provided with an informational briefing halfway through the experiment after making their initial set of choices, highlighting the findings that products with more dispersed ratings tend to have a higher likelihood of containing acquired fake reviews (see details in Figure 6). After receiving this information, participants made a second set of choices. In contrast, the placebo group made both sets of choices without receiving the information treatment.

		Treatment condition		
Variable	(1) Both	(2) Information	(3) Placebo	(4) p-value
Female	0.50	0.49	0.50	0.848
Age	39.18	39.13	39.22	0.926
Black	0.07	0.07	0.07	0.925
Student	0.11	0.12	0.11	0.903
Full-time employed	0.46	0.47	0.45	0.644
N	712	350	362	

Table 10: Balance tests of the experiment

Notes: In column (4), we test the hypothesis of H_0: the mean of the variable in column (2) = (3) against H_1: the mean of column (2) $\neq$ (3), using a t-test, for each variable respectively.

The data set comprises $N = 712$ participants,[30] with 350 in the treatment group and 362 in the control group. The balance table (Table 10) provides a demographic comparison between the treated and control groups, with no substantial imbalances between the two groups.

C.2 Average Treatment Effects

Table 11 displays the summary statistics of the main variables in the experiment. The outcome of interest is whether the participant chose the product with a lower ratings discrepancy. The next variables are the measures of consumer suspicion of fake reviews, including (i) listing fake reviews as reasons for making the choice, (ii) the perceived probability of the product purchasing fake reviews, and (iii) the proportion of fake reviews in the product. Moreover, the perceived ratings discrepancy is elicited. Focusing on the pre-treatment period, the treatment and control groups are fairly balanced.

The summary statistics table shows that more than 50% of the participants favored products with lower ratings discrepancy, with suspicion of fake reviews being a major factor in their decision-making. This aligns with the findings from the observational study. For the outcome variable and the measures of fake review suspicion, the differences between the treatment and control groups are positive in the post-treatment period. Given the balance between the two groups in the pre-treatment period, we can interpret this difference as an estimate of the ATE of the information treatment.

[30] We set up three attention checks in the experiment and excluded all subjects who failed at least one attention check.

Variable	(1) Both	Treatment condition (2) Information	(3) Placebo	(4) (2)−(3)	(5) p-value
Panel (a): Pre-treatment variables					
Choose low discrepancy product (H v. M)	0.638	0.649	0.627	0.021	0.551
Choose low discrepancy product (H v. L)	0.671	0.700	0.644	0.056	0.110
Choose low discrepancy product (M v. L)	0.662	0.657	0.666	−0.009	0.809
List fake reviews as a reason (H v. M)	0.181	0.189	0.174	0.015	0.615
List fake reviews as a reason (H v. L)	0.178	0.186	0.171	0.014	0.615
List fake reviews as a reason (M v. L)	0.117	0.106	0.127	−0.021	0.375
Probability of fake reviews (H)	40.709	41.271	40.166	1.106	0.572
Probability of fake reviews (M)	32.952	33.097	32.812	0.285	0.874
Probability of fake reviews (L)	31.490	32.500	30.514	1.986	0.267
Proportion of fake reviews (H)	22.143	22.309	21.983	0.325	0.817
Proportion of fake reviews (M)	18.074	18.546	17.619	0.927	0.464
Proportion of fake reviews (L)	16.361	16.903	15.837	1.066	0.382
Perceived ratings discrepancy (H)	39.881	38.180	41.525	−3.345	0.060
Perceived ratings discrepancy (M)	35.014	34.071	35.925	−1.854	0.279
Perceived ratings discrepancy (L)	35.548	34.071	36.975	−2.904	0.090
Panel (b): Post-treatment variables					
Choose low discrepancy product (H v. M)	0.733	0.797	0.671	0.126	0.000
Choose low discrepancy product (H v. L)	0.753	0.811	0.696	0.115	0.000
Choose low discrepancy product (M v. L)	0.712	0.763	0.663	0.100	0.003
List fake reviews as a reason (H v. M)	0.306	0.474	0.144	0.331	0.000
List fake reviews as a reason (H v. L)	0.323	0.483	0.169	0.314	0.000
List fake reviews as a reason (M v. L)	0.246	0.383	0.113	0.270	0.000
Probability of fake reviews (H)	39.146	45.957	32.561	13.396	0.000
Probability of fake reviews (M)	28.379	30.331	26.492	3.840	0.026
Probability of fake reviews (L)	24.753	25.760	23.779	1.981	0.242
Proportion of fake reviews (H)	24.381	28.160	20.727	7.433	0.000
Proportion of fake reviews (M)	18.093	19.509	16.724	2.785	0.030
Proportion of fake reviews (L)	15.961	16.934	15.019	1.915	0.119
Perceived ratings discrepancy (H)	37.294	36.957	37.619	−0.662	0.707
Perceived ratings discrepancy (M)	32.955	32.091	33.790	−1.699	0.283
Perceived ratings discrepancy (L)	32.235	32.206	32.262	−0.057	0.972
N	712	350	362		

Table 11: Summary statistics of main variables of interest in the experiment

Notes: Participants were randomly assigned to an "information" condition that received an information treatment and a "placebo" condition that received a placebo treatment. Participants were asked to do the following both before and after the treatment: (i) choose between each pair of products H, M, and L and provide the reasons; (ii) rate each product's probability of having fake reviews, the proportion of fake reviews, and the perceived discrepancy in ratings. In column (4), we calculate the difference between the variables in columns (2) and (3). In column (5), we test the hypothesis of H_0: the mean of the variable in column (2) = (3) against H_1: the mean of column (2) $\neq$ (3), using a t-test, for each variable.

We observe that 65% (H v. M: 64%, M v. L: 67%, and H v. L: 66%) of participants chose the product with a lower ratings discrepancy before the information treatment. Also, 15% (H v. M: 18%, M v. L: 18%, and H v. L: 12%) of respondents cited fake reviews as the rationale behind their choice.[31] These results are fairly consistent and robust, regardless of the specific comparison being examined. The perceived probability that a seller acquired fake reviews and the proportion of reviews deemed fake follow a decreasing order from product H to L, as expected, since the discrepancy also diminishes. This observation further supports the assertion that higher discrepancy increases consumer suspicion of products purchasing fake reviews.

Following the implementation of the information treatment, we observe that the treatment increases the proportion of participants who chose the product with a lower ratings discrepancy by 11% (H v. M: 13%, M v. L: 12%, and H v. L: 10%). The information treatment also increases the number of individuals citing fake reviews as the reason for their choice by 30% (H v. M: 33%, M v. L: 31%, and H v. L: 27%). Both of these results remain robust, irrespective of the specific comparisons being made. The treatment effect on the perceived probability of products purchasing fake reviews and the proportion of fake reviews is relatively heterogeneous, although the positive signs persist. This suggests that the impact of the information treatment can depend on the specific shape of the distribution.

C.3 Heterogeneous Treatment Effects

In the experiment, participants were asked about their demographics. We leverage this additional data to estimate conditional average treatment effects (CATEs) of the information treatment for each demographic group. To summarize, we find that the CATEs for socially disadvantaged groups are larger, suggesting the equity-promoting potential of providing additional information. This analysis is more exploratory in nature and may be of interest to some readers.

Table 12 illustrates the heterogeneous treatment effects of the information treatment concerning fake reviews on the probability of selecting products with lower ratings discrepancy across various demographic strata. In the case of gender (Panel A), providing

[31] See Appendix C.4 for a detailed analysis of the relative importance of fake review suspicion as a factor behind participants' choices.

		Treatment condition			p-value	
	(1)	(2)	(3)	(4)	(5)	(6)
Dimension	# Obs.	Information	Placebo	CATE	(2) = (3)?	Hetero.?
Panel A: Gender						
Female	1041	0.771	0.695	0.076	0.006	0.051
Male	1059	0.814	0.665	0.150	0.000	0.060
Other/unspecified	36	0.583	0.542	0.042	0.819	0.648
Panel B: Age group						
< 30	522	0.782	0.692	0.090	0.025	0.527
30 – 39	771	0.805	0.675	0.130	0.000	0.497
40 – 49	453	0.787	0.714	0.074	0.081	0.316
50 – 59	294	0.778	0.627	0.151	0.007	0.451
≥ 60	204	0.821	0.644	0.178	0.009	0.306
Other/unspecified	57	0.667	0.833	−0.167	0.368	0.116
Panel C: Ethnicity						
Black	150	0.747	0.507	0.240	0.002	0.068
Asian	210	0.812	0.602	0.210	0.001	0.099
White	1527	0.785	0.709	0.076	0.001	0.002
Other/unspecified	249	0.838	0.639	0.199	0.000	0.107
Panel D: Education						
High school	273	0.833	0.610	0.223	0.000	0.028
College	1488	0.784	0.686	0.098	0.000	0.199
Graduate school	348	0.804	0.694	0.109	0.019	0.944
Other/unspecified	27	0.600	0.750	−0.150	0.431	0.118
Panel E: Income						
$<$ \$25,000	306	0.778	0.623	0.154	0.003	0.377
\$25,000 – \$49,999	483	0.799	0.606	0.193	0.000	0.020
\$50,000 – \$74,999	429	0.819	0.770	0.049	0.205	0.093
\$75,000 – \$99,999	333	0.741	0.708	0.033	0.501	0.069
\$100,000 – \$149,999	330	0.797	0.768	0.029	0.532	0.069
Other/unspecified	255	0.795	0.602	0.194	0.001	0.123
Panel F: # Consumer review websites						
1	981	0.807	0.665	0.142	0.000	0.153
2	528	0.812	0.636	0.176	0.000	0.062
3	363	0.742	0.731	0.011	0.819	0.022
4+	363	0.776	0.750	0.026	0.582	0.053
All	2136	0.790	0.677	0.114	0.000	/

Table 12: Heterogeneous treatment effects on the probability of choosing low-ratings-discrepancy products

Notes: Participants were randomly assigned to an "information" condition that received an information treatment and a "placebo" condition that received a placebo treatment. In column (5), we estimate the conditional average treatment effect (CATE) by taking the difference between the values in columns (2) and (3). In column (6), we test the hypothesis of H_0: the CATE of this subset of participants $= 0$ against H_1: the CATE of this subset of participants $\neq 0$, using a t-test. In column (7), we test the hypothesis of H_0: the CATE of this subset of participants is equal to the CATE of other participants, against H_1: the CATE of this subset of participants is not equal to the CATE of other participants.

information about fake reviews increases the probability of choosing products with lower ratings discrepancy by 8% for females and 15% for males, with males generally exhibiting a lower choice probability in the placebo. One potential reason is that females initially may have had a heightened awareness of fake reviews. However, supplying information about fake reviews proves more effective for male customers in choosing products with lower discrepancies, as their suspicion regarding fake reviews is now elevated.

In terms of age (Panel B), no substantial differences are detected. When comparing with other age groups, it is important to note that, although the overall awareness of fake reviews might be lower among individuals in particular age groups, the participants in this study may not be a representative sample of their entire age cohort. Recall that we require them to have an Amazon account. Given that they participated in this research, it is possible that this specific group of participants possesses a higher baseline awareness of fake reviews compared to their counterparts who did not take part in the study.

Regarding ethnicity (Panel C), the information treatment effect exhibits variations among different ethnic groups. The probability of opting for products with lower ratings discrepancy increases by 24% for Black individuals, 21% for Asian individuals, 8% for White individuals, and 20% for those belonging to other or unspecified ethnic backgrounds. These findings suggest that the influence of information about fake reviews on purchasing decisions is larger for Black, Asian, and other minority groups compared to White individuals. This can be attributed to their lower baseline levels of awareness and similar post-treatment choice probability. Differences in income and education levels (potentially due to different access to information or resources) may contribute to disparities among ethnic groups in prior experiences and access to information about online shopping. Consequently, these groups may possess less knowledge about fake reviews in the absence of targeted interventions. However, providing education (i.e., the information treatment) to these groups yields the largest change, as measured by the CATE, highlighting the different effects of supplementing information across different ethnicities.

Panel D, which focuses on education levels, demonstrates that the CATEs vary among the groups. Participants with high school education display the most substantial increase in the probability of selecting products with lower ratings discrepancy (22%), followed closely by those with graduate school (11%) and college education (10%). In contrast, the probability for individuals with other or unspecified education levels shows a slight

decrease (−15%). These results suggest that providing information about fake reviews may potentially help bridge information gaps. The impact is particularly pronounced for high school graduates, who exhibit a comparatively lower baseline awareness of fake reviews.

In relation to income levels, the information treatment effect varies across different income brackets (Panel E). The probability of choosing products with lower ratings discrepancy rises by 15% for individuals earning less than $25,000, 19% for those whose earnings are between $25,000 and $49,999, 5% for those whose earnings are between $50,000 and $74,999, 3% for those whose earnings are between $75,000 and $99,999, and 3% for those whose earnings are between $100,000 and $149,999. The results indicate that disseminating information about fake reviews has a more substantial influence on the purchasing decisions of lower-income individuals. This can be attributed to the fact that the baseline awareness of fake reviews is lower for these individuals, and the behavior of treated individuals is similar across income groups. The findings and interpretation here echo the insights gleaned from the analysis on other dimensions. One possible reason for this discrepancy is that lower-income individuals may have limited access to information or resources, rendering them more vulnerable to the effects of fake reviews. As a result, providing these individuals with information about fake reviews could promote more informed purchasing decisions.

Lastly, when examining the number of consumer review websites utilized, the information treatment's effect demonstrates heterogeneity as well (Panel F). The probability of opting for products with lower ratings discrepancy increases by 14% for individuals using one website, 18% for those using two websites, 1% for those using three websites, and 3% for those using four or more websites. This finding implies that disseminating information about fake reviews has a larger impact on the decision-making of individuals who rely on fewer consumer review websites. This observation corroborates the results in other cases of heterogeneous treatment effects. Generally, individuals with less prior experience or information tend to have a lower baseline. However, after the treatment, their behavior becomes more similar to others. This leads to individuals with limited information having a lower pre-treatment level, but a comparable post-treatment level once the information is provided.

The results derived from the analysis of heterogeneous treatment effects indicate that

the provision of information concerning fake reviews serves as an equalizing force. This may lead to a more equitable distribution of outcomes among various consumer groups. These insights hold significant implications for policymakers and businesses aiming to enhance market efficiency and fairness of consumers ending up with more similar information. By understanding the differential impacts of information treatments across diverse consumer segments, targeted interventions can be designed to effectively address information disparities and promote a more balanced market environment.

C.4 Reasons for Choice

Before launching the experiments, we ran a survey on Prolific Academic and collected reasons why participants chose products with a high or low ratings discrepancy. We then used an online natural language processing tool to summarize the reasons for choosing products with a high or low ratings discrepancy, respectively. In the experiment, after the participant made the choice, depending on whether they chose products with a lower or higher ratings discrepancy, one list of these summarized reasons was shown to the participants in a random order. They were then asked which of the reasons resonated with them the most.

For participants who chose products with a lower ratings discrepancy, the list of candidate reasons is:

1. The other product may have purchased many fake 5-star reviews, leading some customers to leave 1-star reviews in retaliation.

2. I prefer products with less 1-star ratings since I don't have the time to analyze the distribution of reviews in detail.

3. This product does not have extremely negative reviews.

4. The other product may have a certain fail rate, meaning that it sometimes works and sometimes does not.

5. The other product may have inaccurate descriptions, causing people's opinions to be more diverse when receiving the product.

6. The other product may attract customers with more varied tastes.

For participants who chose products with a higher ratings discrepancy, the list of candidate reasons is:

7. I prefer products with more 5-star ratings since I don't have the time to analyze the distribution of reviews in detail.

8. This product has more haters, without which the average rating should be higher.

9. If the product were less complicated and more people knew how to use it, there would be fewer 1-star ratings, resulting in a higher average rating.

10. Despite more 1-star negative reviews, I am confident in the effectiveness of this product as an expert in this field.

11. I am willing to take a risk on a product with more extreme ratings because of Amazon's excellent return policy. Even if I encounter a 1-star rating, I can easily return the product, and with a high probability, the product will be excellent.

To determine which factors influence a participant's choice, we compare the importance of these factors by regressing the binary variable indicating whether a participant chose the product with a lower ratings discrepancy on the listed reasons. We use the pre-treatment data, which includes people's choices before the information intervention. Candidate reasons 1-6 are treated as explanatory variables, while candidate reasons 7-11 serve as controls. This distinction is made because reasons 1-6 are the options presented to participants when they choose the product with lower ratings discrepancies. Additionally, the regression model includes individual fixed effects. Since all explanatory variables are binary, the magnitude of the coefficients can be compared directly.

The results presented in Table 13 indicate that the suspicion of fake reviews is one of the primary contributing factors to why participants chose products with low ratings discrepancies before receiving the information treatment. Among the main reasons (1-6), "Fake reviews (1)" has the largest coefficient (0.152). Other important reasons include "No time to analyze (2)," "Fewer negative reviews (3)," and "Varied taste (6)." In summary, the results in Table 13 provide supporting evidence that consumer suspicion of fake reviews could be one of the primary reasons why participants chose products with lower ratings discrepancies before receiving the information treatment.

	Outcome: Choose the lower-ratings-discrepancy product
Fake reviews (1)	0.152***
	(0.017)
No time to analyze (2)	0.149***
	(0.014)
Fewer negative reviews (3)	0.149***
	(0.014)
Fail rate (4)	0.099***
	(0.014)
Inaccurate descriptions (5)	0.116***
	(0.018)
Varied taste (6)	0.124***
	(0.021)
Covariates	Yes
N	2,136
R^2	0.929
Adjusted R^2	0.893

Table 13: Reasons for choosing products with lower ratings discrepancies

Notes: This table presents the results of a regression analysis examining the reasons why participants chose products with lower ratings discrepancies before receiving the information treatment.

D Experiment Controlling for 1-Star Ratings

We conducted an experiment to investigate the potential alternative mechanism that participants' choices in the previous study might have been driven by their dislike of 1-star ratings, and that the impact of ratings discrepancy might not exist after controlling for this factor. To investigate this, we designed an experiment similar to the one described in Appendix C, but with most of the comparisons having the same proportion of 1-star ratings while differing in the levels of ratings discrepancy, and the other having the same level of ratings discrepancy but different proportions of 1-star ratings.

The rating distributions used in the experiment are illustrated in Figure 7. The experiment involved pairwise comparisons between the following adjacent products: products 1 v. 2, 2 v. 3, 3 v. 4, and 4 v. 5. Among the pairwise comparisons, products 1 v. 2, 3 v. 4, and 4 v. 5 are designed to have the same proportion of 1-star ratings but different



(a) Product 1 (b) Product 2 (c) Product 3 (d) Product 4 (e) Product 5

Figure 7: Product rating distributions in the experiment controlling for 1-Star ratings

Notes: The products have the same average ratings (4.4/5.0), but the distribution is different. The rating distributions vary mainly in two aspects: the proportion of 1-star ratings and the standard deviation of the ratings. Products 1 and 2 both have 4% of 1-star ratings, and products 3, 4, and 5 all have 8% of 1-star ratings. The standard deviations of the rating distributions are (a) 1.079, (b) 1.168, (c) 1.168, (d) 1.234, and (e) 1.310, respectively. That is, products 2 and 3 have the same standard deviation of ratings. These distributions are the probability mass functions of a family of zero-inflated beta-binomial distributions. That is, we first fix the proportion of 1-star ratings and set the proportions of 2-star to 5-star ratings to be proportional to the probability mass functions of BetaBin$(3, \alpha\gamma, \beta\gamma)$. Here, the ratio of α to β influences the mean of the distribution, and γ influences the shape of the distribution while holding the mean fixed. We set (α, β) to be $(6, 1)$ for products 1 and 2, and $(9, 1)$ for products 3-5. The γ's are $2^{-2.5}$, 2^{-4}, 2^{-2}, 2^{-4}, and 2^{-6}, for the five products, respectively.

levels of ratings discrepancy. These pairs allow us to isolate the effect of ratings discrepancy on consumer preferences while controlling for the influence of 1-star ratings. On the other hand, products 2 v. 3 are constructed to have different proportions of 1-star ratings but the same level of ratings discrepancy. This comparison enables us to examine the impact of 1-star ratings on consumer choice while holding the ratings discrepancy constant. Moreover, the rating distribution of product 1 is the same as that of product L in the choice experiment described in Appendix C, and the rating distribution of product 4 is the same as that of product M in the choice experiment. Thus, products 2 and 3 can be seen as "interpolations" between products L and M in the choice experiment, holding either the proportion of 1-star ratings or the level of ratings discrepancy constant between adjacent products.[32] This research design allows us to distinguish the impact of the proportion of 1-star ratings and ratings discrepancy on consumer choice.

We now quantify the impact of ratings discrepancy and the proportion of 1-star ratings on consumer choice (Table 14). For each pair of products, we calculate the difference in the standard deviations of ratings and the difference in the proportions of 1-star ratings. We then estimate regressions using these two variables as explanatory variables and an indicator of choosing the first product in the pair as the dependent variable. Moreover, the regressions are based on data collected prior to the information treatment, and we pool the data over all four pairs of comparisons. The analysis reveals that both larger discrepancies

[32] Additionally, product 5 can be seen as an "extrapolation" beyond product 4. We can not perform a similar interpolation between products M and H in the choice experiment. This is because almost all ratings of product H are 1-star and 5-star ratings, and this does not allow for variation in ratings discrepancy when holding fixed the average rating and the proportion of 1-star ratings.

	Outcome: Choose products with lower ratings discrepancies or proportions of 1-star ratings					
	OLS		Logistic		Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
Difference in std. ratings	−3.129**	−3.281**	−17.478**	−18.317**	−11.175***	−11.666***
	(1.470)	(1.478)	(7.132)	(7.172)	(4.274)	(4.299)
Difference in proportions of 1-star	−6.620**	−6.266**	−36.972***	−34.991**	−23.324***	−22.035***
	(2.898)	(2.921)	(14.057)	(14.168)	(8.424)	(8.491)
Difference in proportions of 5-star		0.387		2.158		1.340
		(0.400)		(1.934)		(1.160)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	1,852	1,852	1,852	1,852	1,852	1,852
BIC	5,574	5,580	5,442	5,448	5,442	5,448

*p <0.1; **p <0.05; ***p <0.01

Table 14: Impact of ratings discrepancy and the proportion of 1-star ratings on consumer choice

Notes: This table presents the impact of ratings discrepancy and the proportion of 1-star ratings on consumers' probability of choosing a product. For each pair of products, we calculate the difference in standard deviations of ratings and the difference in the proportions of 1-star ratings and 5-star ratings. We use them as explanatory variables in regressions with an indicator of choosing the first product in the pair as the dependent variable.

in ratings and larger proportions of 1-star ratings negatively influence people's probability of choosing a product. This finding is robust to various specifications, including a linear probability model (columns (1) and (2)), a logistic model (columns (3) and (4)), and a probit model (columns (5) and (6)). These results indicate that while people care about the proportion of 1-star ratings, they also consider the discrepancy in ratings even after controlling for the proportion of 1-star ratings. In other words, the impact of ratings discrepancy on consumer choice persists when accounting for the influence of 1-star ratings.