Unmasking the Deception: The Interplay between Fake Reviews, Rating Dispersion, and Consumer Demand*

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

October, 2023

*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, and Rising Scholars Conference 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 and 2023-05-16326, and pre-registered at AsPredicted (#127216, https://aspredicted.org/SSR_6WB; and #135597, https://aspredicted.org/R4K_W3R). 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 California, Berkeley. Email: huangyh@haas.berkeley.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, Rating Dispersion, and Consumer Demand

Abstract

In online marketplaces, consumers rely on reviews to make informed purchase decisions, rendering the presence of fake reviews detrimental. Meanwhile, products that have acquired fake reviews may exhibit different characteristics of the review distributions, such as a higher rating dispersion, and consumers may take this into account when making their purchase decisions. In this paper, we explore the interplay between fake reviews and rating dispersion, 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 rating dispersion is positively correlated with the likelihood that the product has acquired fake reviews. Then, by employing rating dispersion changes due to rating distribution rounding as an identification strategy, we isolate the effect of rating dispersion on product sales. Our findings indicate that rating dispersion negatively affects sales. To further investigate the underlying mechanism, we conduct experiments in which participants are shown products with varying rating distributions and are asked about their choices, willingness to pay, and concerns about fake reviews. Additionally, we incorporate an information treatment, which later functions as an instrumental variable, to determine the impact of consumer suspicion of fake reviews on their demand. The experimental results demonstrate that rating dispersion significantly impacts consumer demand, with heightened dispersion in ratings leading to concerns about fake reviews. These findings align with the outcomes from our observational study, underscoring the importance of comprehending the interplay between fake reviews, rating dispersion, and consumer demand in online marketplaces.

Keywords: Fake Reviews, Online Marketplaces, Word of Mouth

JEL Classification: D12, D83, L15, L81

1 Introduction

In the context of e-commerce, consumers often face information asymmetry challenges. To overcome these challenges, they increasingly rely on user-generated content (UGC) to inform their purchasing decisions. To facilitate consumers' access to UGC, major e-commerce platforms have implemented reputation systems. These systems allow online marketplaces to effectively organize and display UGC, enabling consumers to efficiently evaluate product and service offerings.

However, the presence of fake reviews presents a challenge for 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 fraudulent 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 purchased fake reviews. Based on these indicators, consumers can potentially unmask the deception and make more informed decisions. One example could be rating distributions, which are prominently displayed on the product introduction page, frequently depicted as bar charts, representing the proportion of reviews associated with each rating level. The dispersion of rating distributions, which represents the degree of variation in user ratings for a particular product, could be a factor to consider.

Our first research question aims to establish a connection between rating dispersion and fake reviews.[1] Previous research has demonstrated that companies often reimburse buyers for purchasing their products and providing 5-star ratings (e.g., He et al., 2022b). The acquisition of fake reviews leads to a substantial, albeit potentially transient, increase in average ratings and the number of reviews. After a certain period, when firms stop purchasing fake reviews, their average rating declines, and the proportion of 1-star reviews increases (e.g., Chakraborty et al., 2022). This suggests that, when holding the average rating constant, the rating dispersion, characterized by the standard deviation of the rating distribution, will be strongly correlated with the likelihood of a seller having purchased fake reviews for a product. By analyzing a secondary data set from He et al. (2022a), we confirm our hypothesis.

We contribute to the literature on fake reviews, which investigates the existence of fake

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

reviews and the market dynamics driving their demand. 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). A better understanding of the prevalence and implications of fake reviews can be crucial for the integrity of online reputation systems and consumer trust in online marketplaces. We also contribute to the literature that identifies the factors contributing to rating dispersion in online reviews (e.g., Schoenmueller et al., 2020), providing insights into the prevalence and drivers of this phenomenon. Moreover, we contribute to the literature that detects products with fake reviews; we refer readers to Zhou and Zafarani (2020) for a survey of this literature.

Because heightened rating dispersion may be correlated with the inauthenticity of UGC, it can raise consumer concerns and impact consumer demand. Our second research question focuses on exploring the causal effect of rating dispersion on product sales. To address this question, we adopt a novel methodological approach that enables us to isolate the impact of rating dispersion on product sales while controlling for potential confounding factors. Specifically, we analyze changes in the standard deviations of ratings that stem from percentage rounding in the rating distribution. Our findings demonstrate a significant negative effect of rating dispersion on product sales.

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 sales, and the literature concerning online reviews investigates the influence of these average ratings on consumer choices (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 (Li and Hitt, 2008; Proserpio and Zervas, 2017; Chevalier et al., 2018; Chakraborty et al., 2022). Regarding rating dispersion, the existing literature exhibits mixed evidence of its effect on consumer preferences and demand (Sun, 2012; Luo et al., 2013; He and Bond, 2015; Wu et al., 2015; Rozenkrants et al., 2017). These studies reveal both positive and negative effects of dispersion, with factors such as self-expression and product polarization playing key roles.

While our findings align with some of the existing literature, our first research question suggests that additional channels may be at play in the relationship between rating dispersion and consumer behavior: Consumer suspicion about fake reviews could contribute to the observed negative effect of rating dispersion on sales. To further explore the underlying mechanism linking fake reviews, rating dispersion, and consumer demand, our third research question involves conducting two experiments in which participants are presented with products that have varying rating distributions. The other characteristics of these products, such as images and product descriptions, are identical. This experimental design allows us to investigate whether the suspicion of fake reviews drives consumers to prefer products with lower rating dispersion while controlling for confounders. By investigating this causal relationship, we strive to develop a more in-depth understanding of how the presence of fake reviews shapes consumers' subsequent purchasing decisions. This analysis can shed light on the broader implications of fake reviews in the online marketplace.

The two experiments di er in the outcome we measure choices between pairs of products in the rst experiment, and willingness to pay (WTP) for each product in the second experiment but the other parts of the experiments are similar. For each product, we elicit the participants' suspicion about fake reviews. Moreover, in both experiments, participants are divided into treatment and control groups, with the treatment group receiving an information treatment designed to in uence their concerns about fake reviews. The information treatment serves as an instrumental variable (IV) to assess the impact of consumer suspicion of fake reviews on demand. Our ndings reveal that the information treatment signi cantly a ects participants' concerns about fake reviews, which in turn in uences their choices and WTP.

In addition, we contribute to the literature on platform reputation mechanisms. This literature explores their structure, biases, and the signi cance of aggregation in consumer evaluations (Nosko and Tadelis, 2015; Dai et al., 2018; Vellodi, 2018; Timoshenko and Hauser, 2019; Acemoglu et al., 2022; Shi et al., 2022). These investigations underscore the necessity for ideal information disclosure and the mitigation of potential obstacles to entry, in order to guarantee the e cacy and dependability of reputation frameworks. By examining the reactions of study participants to the informational intervention, we can obtain a more comprehensive understanding of the manner in which apprehension regarding fake reviews in uences demand. This may also demonstrate how the platform can utilize supplementary information, such as cautionary messages or evidence indicating patterns of counterfeit review products, to enhance consumer well-being.

The remainder of the paper is organized as follows. Section 2 introduces a framework to elucidate the consumer decision-making process. Section 3 provides a description of two observational data sets employed in this paper. Drawing on the rst data set, Section 4 delves into the correlation between rating dispersion and the likelihood of a product procuring fake reviews. Drawing on the second data set, Section 5 o ers a causal estimation of the rating dispersion's impact on sales. To further explore the mechanism by which rating dispersion a ects demand, and to quantify its in uence, Section 6 details a conjoint experiment that we conducted. Section

7 presents an alternative experimental design, wherein we elicit participants' WTP. Lastly, Section 8 discusses managerial insights, and Section 9 reviews the principal findings and presents directions for future research.

2 Theoretical Framework: Unmasking the Deception

Consider the following framework of the influence of consumers' concerns about fraudulent reviews on demand, specifically by modifying their perceptions of the distribution.

Consumer i can only observe the rating distribution ${}_{j}$ of product j, which we term the observed distribution as represented in the rating histogram. For example, if product j has 10% of 1-star ratings, 5% of 2-star ratings, 5% of 3-star ratings, 5% of 4-star ratings, and 75% of 5-star ratings, then ${}_{j} = (10\%; 5\%; 5\%; 5\%; 75\%)$.

Individuals may harbor suspicions about a ${}_{ij}$ 2 $[0; 1]$ mass of the 5-star ratings being inauthentic. We subsequently employ ${}_{ij}$ to represent the conjectured distribution of product j following the removal of the ${}_{ij}$ fraction of 5-star ratings, which satisfies

$$ {}_{j} = {}_{ij}\ {}_{5} + (1 \quad {}_{ij})\ {}_{ij} : $$

Note that ${}_{5}$ represents a degenerate distribution of 5-star ratings only, that is, ${}_{5} = (0\%; 0\%; 0\%; 0\%; 100\%)$. Consequently, the conjectured distribution, which contributes to consumers' expected utility, is given by

$$ {}_{ij} = \frac{{}_{j} \quad {}_{ij}\ {}_{5}}{1 \quad {}_{ij}} : $$

In the example above, suppose consumer i believes that half (${}_{ij} = 50\%$) of the ratings are fake. Then, consumer i first removes those 5-star ratings, resulting in ${}_{j} \quad {}_{ij}\ {}_{5} = (10\%; 5\%; 5\%; 5\%; 25\%)$. But this is no longer a distribution, as the fractions do not sum up to 100%. Thus, consumer i rescales it with $(1 \quad {}_{ij})$, resulting in ${}_{ij} = (20\%; 10\%; 10\%; 10\%; 50\%)$.

Consumer i's expected utility U_{ij} of purchasing product j can depend on the conjectured distribution ${}_{ij}$, expressed as

$$ U_{ij} = f({}_{ij}) = f\left(\frac{{}_{j} \quad {}_{ij}\ {}_{5}}{1 \quad {}_{ij}}\right) : \qquad (1) $$

One reason that U_{ij} may depend on ${}_{ij}$ is that ${}_{ij}$ is a signal about product quality. Thus, consumer utility depends on expected product quality conditional on this signal, and the function f captures such dependency.

We assume that f exhibits monotonicity in $_{ij}$, such that, if $_{ij}$ demonstrates rst-order stochastic dominance over $\tilde{}_{ij}$, then $f(\ _{ij})\quad f(\tilde{}_{ij})$.[2] Given this assumption, we can establish the following proposition:

Proposition 2.1. Consumer expected utility U_{ij} decreases in the suspicion level $_{ij}$.

The proposition shows that consumer suspicion about fake reviews negatively a ects consumer utility. This raises the question of how consumer suspicion $_{ij}$ is determined, and, specifically, how it depends on the observed rating distribution $_{j}$. Formally, we are interested in the following relationship for each consumer i:

$$_{ij} = g_i(\ _{j}): \tag{2}$$

We hypothesize that consumer suspicion $_{ij}$ increases with rating dispersion when holding the average rating constant, such that, if $_{j}$ is a mean-preserving spread of $\tilde{}_{j}$, then $g_i(\ _{j})\quad g_i(\tilde{}_{j})$.[3] As a special case, this property is satis ed when g_i is an increasing function of the rating standard deviation; we focus on this special case in our empirical speci cations.

To summarize, during the decision-making process for purchases, consumers unmask the deception introduced by fake reviews. That is, they take into account the distribution of ratings showcased on Amazon, infer the proportion of deceptive 5-star reviews, and then subtract them from the showcased distribution. This framework guides our subsequent data analysis.

3 Observational Data

In this paper, we employ two separate observational data sets. The rst is a secondary data set from He et al. (2022a). The second is derived 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 our research questions; these will be discussed in detail in Sections 6 and 7.

3.1 Amazon Data with Fake Review Labels

The data set used by He et al. (2022a) includes a crucial variable indicating whether a seller has purchased fake reviews for a product during the observation period. Most sellers identi ed

[2] For example, suppose $_{ij}$ = (0% ; 0%; 0%; 0%; 100%) denotes a rating distribution with all 5-stars, and $\tilde{}_{ij}$ = (0%; 0%; 0%; 100%; 0%) denotes a rating distribution with all 4-stars; then, $_{ij}$ demonstrates rst-order stochastic dominance over $\tilde{}_{ij}$, and we assume that consumers have higher utility for the former than the latter.

[3] For example, suppose $_{j}$ = (0% ; 0%; 50%; 0%; 50%) denotes a rating distribution with half of 5-stars and 3-stars respectively, and $\tilde{}_{j}$ = (0% ; 0%; 0%; 100%; 0%) denotes a rating distribution with all 4-stars, then $_{j}$ is a mean-preserving spread of $\tilde{}_{j}$. We hypothesize that consumers will have higher suspicion for the former than the latter.

as having bought fake reviews compensate reviewers by refunding the product's cost through online transactions after posting a 5-star review. These reviewers are incentivized to craft authentic-seeming 5-star reviews that can bypass Amazon's detection systems. In contrast to incentivized reviews, where sellers offer free or discounted products or future discounts in exchange for reviews that disclose the transaction and are not necessarily 5-star ratings, fake reviews require a 5-star rating for reimbursement.

The data set is collected by tracking social media groups that recruited individuals to leave fake reviews from October 2019 to June 2020. Our 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.

Regarding ratings, this data set only records the proportion of 1-star and 5-star ratings, 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, our imputation method yields an upper bound on the standard deviation, by assuming no 3-star ratings (in other words, all other ratings are 2- or 4-star ratings) and solving the number of 2-star and 4-star ratings that lead to an average rating consistent with observed ratings. Our results are robust to other ways of imputation, such as a lower bound on the standard deviation derived similarly as above.

		Purchased Fake Review?		
	(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 purchased fake review or not

Note: 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 divided into two groups: those that have acquired fake reviews (Yes) and those that have not (No). The table contrasts various characteristics of these groups, such as the number of reviews, average rating, rating dispersion, and the proportion of helpful reviews.

It is essential to note that 1,472 products in the data set are flagged as having fake reviews, while only 1,936 products are labeled as not engaging in such practices. The latter comprise products that directly compete with the fake review products, serving as a comparison set. To create this comparison set, He et al. (2022a) select the two competitors that appeared most frequently on the same search page during the seven days before and seven days after the first social post related to the fake reviews for each fake review product. 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 purchased fake reviews have, on average, a smaller total number of reviews (322) compared to their counterparts without fake reviews (477), with a p-value less than 0.001. The lower average number of reviews for products with fake reviews might imply that these products are either new or less established, and, therefore, sellers resort to buying 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, potentially addressing 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), demonstrating a statistically significant difference ($p < 0.001$). The elevated average rating for products with fake reviews is anticipated, as one of the primary objectives of buying fake reviews is to inflate the product's overall rating, rendering it more attractive to prospective buyers. This outcome emphasizes the efficacy of fake reviews in manipulating a product's perceived quality.

The proportion of helpful reviews is lower for products containing fake reviews (0.20) in comparison to those without (0.24), and this difference also bears statistical significance ($p < 0.001$). The decreased proportion of helpful reviews for products with fake reviews may suggest that consumers find these reviews less informative or pertinent. 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 buyers may not find the reviews useful in making informed purchasing decisions.

3.2 Amazon Data with Sales and Sales Rankings

The first observational data set has several limitations, so we obtain another data set through web scraping. First, the available version of the first data set lacks information pertaining to sales. Second, the first data set was assembled in 2020 and many of the original products are no longer accessible[4]. For these reasons, we obtain a second observational data set through web scraping, which allows us to record sales rankings and infer sales.

Our second observational data set focuses on the Home and Kitchen category[5] on Amazon during the period from January 2023 to September 2023. This data set comprises a daily panel of 721 products[6], 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 rst three pages for the keywords Home and Kitchen.

The panel encompasses various variables, including average ratings, rating standard deviations, prices, number 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. Speci cally, they nd 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[7]. If we observe the inventories of a product on two consecutive days, we can calculate the sales as the di erence between them[8]. We then utilize such observations to compute daily sales and to estimate the relationship between sales and sales rankings[9]. 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 can manage the fake reviews to signal-jam any inferences by consumers. 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 here (and in the next sections) should then be interpreted as presenting the correlation between the existence of fake reviews

[4]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.

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

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

[7]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.

[8]We discard any observations where the inventory on the rst day is lower than that on the second day, assuming that inventory increases signify much less frequent restocking. 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.

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

and ratings' dispersion, and the effect of ratings' dispersion on sales, 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.

4 Rating Dispersion and Fake Reviews

In this section, we establish a positive correlation between fake reviews and rating dispersion. This correlation can serve as supporting evidence for the framework in Section 2 that consumers may infer that a product with more dispersed ratings is more likely to have purchased 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 procured fake reviews compensate reviewers by reimbursing the product's cost after receiving a 5-star review. Within the data set, 1,472 products are flagged as having purchased fake reviews, while 1,936 products are identified as not engaging in such practices.

Each plot in Figure 1 illustrates a variation in the standard deviation of ratings when comparing products with fake reviews to those without, across various average rating levels. The average ratings in the plot (ranging from 3.5 to 4.9) include approximately 90% of the products in our data set. Across all observed average rating levels, the plots illustrate that products flagged with fake reviews exhibit a density distribution that leans more to the right. This translates to products with fake reviews showing a higher rating dispersion than those without. Overall, this finding suggests that products with a higher rating dispersion are more likely to have fake reviews, holding fixed the average ratings.[10]

Figure 2 displays a model-free heatmap with the rating standard deviation on the y-axis and the average ratings on the x-axis. The color denotes the estimated probability of a product having purchased fake reviews within each grid, calculated by taking the average of the binary fake review indicator from the raw data, with lighter colors signifying higher probabilities.

The figure illustrates how the standard deviation of ratings is related to the likelihood 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 rating standard deviation is higher, showing its positive correlation with the likelihood of having fake reviews.

The above robust and model-free observations provide support to the argument that, when

[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 1: Products that purchased fake reviews have higher rating dispersion

Note: We restrict our sample to products with at least 10 reviews. In each panel, we restrict the sample to products with a certain rounded average rating and plot the density distribution of rating standard deviations for products with and without fake reviews. For all rounded average ratings, the panels suggest higher rating standard deviations for products with fake reviews compared to products without fake reviews.

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

Note: We restrict our sample to products with at least 10 reviews. For the products with the same average ratings, the likelihood of a product having fake reviews increases as the rating standard deviation increases. The heat map displays the probability of purchasing a product with fake reviews against the average ratings and rating standard deviations, with lighter shades indicating higher probabilities of fake reviews.

the average rating remains constant, a higher standard deviation of ratings is indicative of an increased likelihood of a seller having purchased fake reviews.

We then use regression analysis to further quantify this relation. We rst estimate the following regression:

$$\text{Fake}_j = {}_{R_j} + {}_1 \quad {}_j + {}_2 \; X_j + {}_j; \tag{3}$$

where j indexes the products in this data set, Fake_j denotes whether product j has fake reviews before, ${}_{R_j}$ denotes a xed e ect based on the rounded average rating R_j of the product, and X_j denotes potential covariates, including the logarithm of the number of reviews. In addition to the ordinary least squares (OLS) speci cation above, we also run a logistic model and a probit model.

	Outcome: Purchased fake review before					
	OLS		Logistic		Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
Std. rating	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 xed e ects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,408	3,408	3,408	3,408	3,408	3,408

$p < 0.1$; $p < 0.05$; $p < 0.01$

Table 2: Products that have fake reviews are associated with a higher rating dispersion

Note: This table presents the coe cients from the regression of the likelihood of a product having fake reviews on average ratings, rating standard deviations, 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 coe cients.

Table 2 displays the coe cients illustrating the conditional correlation of rating standard deviation with the likelihood 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 procuring fake reviews early in a product's lifecycle might lead to more polarized ratings later on. This could occur if customers, in uenced 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. The negative coe cient on the number of reviews variable further supports this narrative. This suggests that the cycle of purchasing five-star reviews and subsequent retaliation is more prevalent for products at an earlier stage. Moreover, independent from the interpretation of the causal relationship, the mere presence of this correlation is significant in itself. It can potentially serve as a warning for potential buyers who are wary of products with a higher rating dispersion, especially when compared to others with similar average ratings.

5 Impact of Rating Dispersion on Sales

In Section 4, we show that products with fake reviews, when conditioned on the same average, exhibit a higher standard deviation. Motivated by the theoretical discussions in Section 2, in this section, we hypothesize that the dispersion of the rating distribution influences sales. To test this hypothesis, we use data scraped from Amazon and employ rating dispersion changes due to rating distribution rounding as our identification strategy.

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

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

$$\Delta \log Y_{jt} = \alpha_t + \beta_1 \Delta \sigma_{jt} + \beta_2 \Delta \log P_{jt} + \beta_3 \Delta X_{jt} + e_{jt}; \qquad (4)$$

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, α_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 and the logarithm of the product's position in search results, and e_{jt} is the error term. We cluster the standard errors at the product level.

Intuitively, our identification strategy is to compare a single product on two adjacent days

[11] Using the sum of the proportions of 5-star and 1-star ratings instead of the standard deviation of ratings leads to the same substantive results as those presented here.

that possess identical rounded average ratings but exhibit minor variations in dispersion. Then, by taking the rst di erence, we aim to partially address potential endogeneity concerns that could arise from unobserved factors. Taking the rst di erence allows us, rst, to control for time-invariant unobserved heterogeneity. Second, it can also partially address some endogeneity issues in prices, under the assumption that price variations across time are more exogenous than the price itself.

It is worth mentioning that 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 our results are not solely based on the speci c imputation method we use for sales and that they are applicable to our raw data.

	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
Observations	254	254
R^2	0.194	0.194
Adjusted R^2	0.097	0.097

$p < 0.1$; $p < 0.05$; $p < 0.01$

Table 3: Rating dispersion harms sales and sales rankings

Note: This table presents the impact of rating dispersion 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 xed e ects. Standard errors are clustered at the product level.

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 rating standard deviation results in a 6.5% decrease in sales and a 4.2% decline[12] in sales rankings, implying reduced sales. The coe cients for the

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

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. Both results consistently exhibit the in uence of rating dispersion on sales. This substantiates our hypothesis that the dispersion of the rating distribution negatively impacts sales.

6 Choice Experiment

In Section 4, we present correlational evidence highlighting the relationship between rating dispersion and fake reviews. In Section 5, we nd that the rating dispersion negatively in uences sales. Our objective in Sections 6 and 7 is to test the hypothesis, suggested in Section 2, that rating dispersion reduces demand by making consumers suspicious of fake reviews. We use two experiments to verify this mechanism and quantify its e ects. The main di erence between the two experiments is the outcome we measure. The experiment in this section entails a conjoint analysis, while the experiment in the next section involves the elicitation of WTP.

6.1 Experimental Design

(a) Product 1 (b) Product 2 (c) Product 3

Figure 3: Product rating distributions in the choice experiment

Note: The products have the same average ratings (4.4/5.0), but the distribution is di erent. The standard deviations of the products are 1.350, 1.234, and 1.079 respectively.

During the experiment, participants were presented with three products with the same average ratings but varying rating distributions. Products are displayed in pairs. They are identical in every aspect, including average rating, price, appearance, and other product attributes, with the sole di erence being the rating dispersion. For each product pair, participants chose between products with the two rating distributions. Figure 3 displays the three distributions, resulting in three groups of comparisons. We henceforth refer to them as products 1, 2, and 3. Note that product 1 has the most dispersed ratings, while product 3 has the least dispersed ratings. We emphasized to the participants that all the products have the same attributes. Participants also identi ed the primary reasons for their choice, which are then converted to a binary variable indicating whether concern about fake reviews is one of the reasons.[13] Moreover, they assessed the extent of differing opinions among reviewers, the likelihood of the seller obtaining fake reviews, and the estimated percentage of fake reviews using a 100-point scale.

The participants were randomly allocated to a treatment (information) group and a control (placebo) group. The information group is 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 purchased fake reviews. The treatment consisted of reading a summary of a news article about fake reviews (Economist, 2020). (See details in Figure 6 in Appendix A.) 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 4: Balance tests of the experiment

Note: 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.

We recruited the participants on Prolific, with the restriction that participants had to have an Amazon account to be eligible for the study. The data set comprises $N = 712$ participants,[14] with 350 in the treatment group and 362 in the control group. The balance table (Table 4) provides a demographic comparison between the treated and control groups, with no statistically significant imbalances between the two groups.

6.2 Average Treatment Effects

Table 5 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 rating dispersion. The

[13] A detailed list of the reasons can be found in Appendix A.

[14] We set up three attention checks in the experiment, and exclude all those who fail at least one attention check.

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

Table 5: Summary statistics of main variables of interest in the experiment

Note: Participants were randomly assigned into 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 1, 2, and 3 and provide the reasons; (ii) rate each product's probability of having fake reviews, the proportion of fake reviews, and the perceived dispersion of ratings. In column (4), we calculate the difference between the variable 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.

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 rating dispersion is elicited. Focusing on the pre-treatment period, the treatment and control groups are fairly balanced when considering a 5% signi cance level threshold.

The summary statistics table shows that more than 50% of the participants favored products with lower rating dispersions, with apprehensions regarding fake reviews being a major factor in their decision-making. This aligns with the ndings from our observational study. For the outcome variable and the measures of fake review concerns, the di erence between the treatment and control groups is statistically signi cant in the post-treatment period, while the di erence is null at the 5% level for the pre-treatment period. Given the balance between the two groups in the pre-treatment period, we can interpret this di erence as an estimate of the average treatment e ect (ATE) of the information treatment.

We observe that 65% (1 v. 2: 64%, 2 v. 3: 67%, and 1 v. 3: 66%) of participants chose the product with a lower rating dispersion even in the absence of the information treatment. Also, 15% (1 v. 2: 18%, 2 v. 3: 18%, and 1 v. 3: 12%) of respondents cited fake reviews as the rationale behind their choice. These results are fairly consistent and robust, regardless of the speci c comparison being examined, which lends external validity to this setup. The perceived probability that a seller purchased fake reviews and the proportion of reviews deemed fake follow a decreasing order from product 1 to 3, as expected, since the dispersion also diminishes. This observation further supports the assertion that higher dispersion increases consumer suspicion about products purchasing fake reviews.

Following the implementation of the information treatment, we observe that the treatment increases the proportion of participants perceiving fake reviews by 11% (1 v. 2: 13%, 2 v. 3: 12%, and 1 v. 3: 10%). The information treatment also increases the number of individuals citing fake reviews as the reason for their choice by 30% (1 v. 2: 33%, 2 v. 3: 31%, and 1 v. 3: 27%). Both of these results remain robust, irrespective of the speci c comparisons being made. The treatment e ect on the perceived probability of products purchasing fake reviews and the proportion of fake reviews is relatively heterogeneous, although the positive signs and statistically signi cant changes persist. This suggests that the impact of the information treatment can depend on the speci c shape of the distribution.

6.3 Fake Review Suspicion

The last three rows of table 5 show that the perceived rating dispersion did not change significantly after the information treatment. This suggests that the information treatment did not impact the perceived rating dispersion; instead, it may have influenced how the participants made inferences based on the perceived rating dispersion, ultimately affecting the perceived propensity of fake reviews, as well as consumer choice. This empirical result is consistent with the framework in Section 2 that the observed rating distribution is relatively objective.

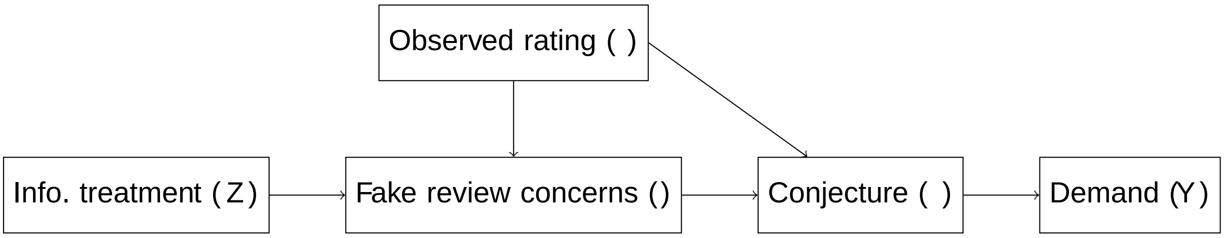

Figure 4: Causal diagram

Note: This causal diagram illustrates the relationships among rating dispersion, concerns about fake reviews, and demand. The information treatment aimed to raise consumers' awareness of fake reviews, which may have influenced their purchase decisions through conjectured rating distributions. Interestingly, the observed rating dispersion did not change significantly when the information treatment was provided. Instead, the treatment influenced conjectured rating distributions, ultimately affecting consumer choice.

Drawing from the framework in Section 2 and the discussions above, we construct the causal diagram depicted in Figure 4. This diagram implies that the exogenous information treatment amplified participants' concerns about fake reviews, which subsequently influenced demand. The rating dispersion affects demand by impacting concerns about fake reviews, as suggested in previous sections. Importantly, the discussions above imply that the information treatment Z has no impact on the observed ratings that is, there is no arrow from Z to .

This experiment provides an ideal context for using the information treatment as an instrumental variable to investigate the impact of concerns about fake reviews on demand. We run an IV regression where the outcome variable is whether the participant chose the product with a lower rating dispersion post-treatment. The independent variable is based on the concern about fake review measures, which are detailed shortly. We instrument the independent variable by whether or not the participant received the information treatment.[15]

Table 6 displays the causal evidence of the impact of concerns regarding fake reviews on product demand. The table is divided into three subtables, each focusing on a distinct measure

[15] We did not pre-register this analysis, so it should be regarded as exploratory instead of confirmatory. However, in the other experiment, a similar analysis is pre-registered.

	Outcome: Decide to purchase the low-rating-dispersion product			
	(1)	(2)	(3)	(4)
	All	Product 1 v. 2	Product 1 v. 3	Product 2 v. 3
List fake review as a reason	0.373	0.381	0.367	0.370
	(0.058)	(0.092)	(0.095)	(0.117)
Constant	0.624	0.617	0.634	0.621
	(0.019)	(0.032)	(0.034)	(0.033)
Observations	2,136	712	712	712
First-stage F	41.412	17.120	15.031	9.942

$p < 0.1$; $p < 0.05$; $p < 0.01$

(a) Measure 1: Listing fake reviews as one of the reasons for making their purchase decision

	Outcome: Decide to purchase the low-rating-dispersion product			
	(1)	(2)	(3)	(4)
	All	Product 1 v. 2	Product 1 v. 3	Product 2 v. 3
Di erence in perceived fake review probability (%)	0.015	0.013	0.010	0.054
	(0.003)	(0.004)	(0.003)	(0.033)
Constant	0.589	0.591	0.607	0.517
	(0.028)	(0.043)	(0.044)	(0.125)
Observations	2,136	712	712	712
First-stage F	30.237	13.162	12.514	2.580

$p < 0.1$; $p < 0.05$; $p < 0.01$

(b) Measure 2: The di erence between the two products in the elicited probability of fake reviews

	Outcome: Decide to purchase the low-rating-dispersion product			
	(1)	(2)	(3)	(4)
	All	Product 1 v. 2	Product 1 v. 3	Product 2 v. 3
Di erence in perceived fake review proportion (%)	0.031	0.027	0.021	0.115
	(0.006)	(0.008)	(0.006)	(0.104)
Constant	0.559	0.563	0.577	0.467
	(0.037)	(0.055)	(0.057)	(0.226)
Observations	2,136	712	712	712
First-stage F	23.815	11.072	10.671	1.221

$p < 0.1$; $p < 0.05$; $p < 0.01$

(c) Measure 3: The di erence between the two products in the elicited proportion of fake reviews

Table 6: Concerns about fake reviews negatively a ect the purchase decision

Note: This table presents the impact of concerns regarding fake reviews on product demand using three di erent measures of fake review concerns. We regress whether the participant chose to purchase the product with a lower rating dispersion on each measure, instrumented by the information treatment dummy. Each sub-table corresponds to one of the measures: (a) listing fake reviews as one of the reasons for making the purchase decision, (b) the di erence between the two products in the elicited probability of fake reviews, and (c) the di erence between the two products in the elicited proportion of fake reviews.

of fake review concerns: Listing fake reviews as one of the reasons for making their purchase decision (Table 6a), the difference between the two products in the elicited probability of fake reviews (Table 6b), and the difference between the two products in the elicited proportion of fake reviews (Table 6c). By incorporating various measures, we gain a more robust understanding of how concerns about fake reviews affect purchase decisions.

The first measure unveils a positive and statistically significant causal relationship between the indicator of citing fake reviews as a reason and the probability of selecting the product with a lower rating dispersion. This relationship remains consistent across all comparisons in terms of sign, magnitude, and significance level, suggesting that participants expressing heightened concern about fake reviews tend to favor products with a lower rating dispersion. Notably, when participants identified fake reviews as their primary concern, the likelihood of selecting the product with a lower rating dispersion within the pair increases by 37% compared to those who did not identify fake reviews as their primary concern.

The second measure also demonstrates a positive and statistically significant causal relationship between the difference in perceived fake review probability and the decision to purchase a product with a lower rating dispersion. This relationship remains stable throughout the pooled case as well as the first and second comparisons (columns 1, 2, and 3). Thus, the instrument is valid, further supporting the notion that concerns about fake reviews adversely affect purchase decisions. Based on the result using pooled data (column 1), a 1% increase in the difference between the perceived probability of fake reviews leads to a 1.5% increase in sales of the product with a lower rating dispersion.

In the context of the third metric, the demonstrated patterns are similar, signifying a positive and statistically significant causal link between the perceived fake review proportion and the purchasing decision favoring a product with a lower rating dispersion. As per the analysis utilizing all the data, a 1% increase in the difference between the fake review proportions contributes to a 3.1% sales increment for the product with a lower rating dispersion. This correlation persists throughout all comparative instances except for the third, where the instrument is weak. The instrument's persistent weakness for the third comparison (product 2 v. 3) across three measures suggests that the difference in distribution between the second and third products did not significantly affect in consumers through any of the three channels. This lack of impact implies that their beliefs about fake reviews preclude substantial change, whether the informational treatment is presented or not.

In summary, the results from Table 6 indicate that concerns about fake reviews adversely affect purchase decisions, with participants more likely to choose products with lower rating dispersions when they are apprehensive about the presence of fake reviews.

6.4 Heterogeneous Treatment E ects

In the experiment, participants were asked about their demographics. We leverage this additional data to estimate conditional average treatment e ects (CATEs) of the information treatment for each demographic group. As a summary, we nd that the CATEs for socially disadvantaged groups are larger, suggesting the equity-promoting potential of providing additional information.

Table 7 illustrates the heterogeneous treatment e ects of the information treatment concerning fake reviews on the probability of selecting products with lower rating dispersions across various demographic strata. In the case of gender (Panel A), providing information about fake reviews increases the likelihood of choosing products with lower rating dispersions 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 e ective for male customers in choosing products with lower dispersion, as their suspicions regarding fake reviews are now elevated.

In terms of age (Panel B), no substantial di erences 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 our 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 our research, it is possible that this speci c 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 e ect exhibits variations among di erent ethnic groups. The probability of opting for products with lower rating dispersions increases by 24% for Black individuals, 21% for Asian individuals, 8% for White individuals, and 20% for those belonging to other or unspeci ed ethnic backgrounds. These ndings suggest that the in uence of information about fake reviews on purchasing decisions is more signi cant 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. Di erences in income and education levels (potentially due to di erent 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.,

Dimension	(1) # Obs.	Treatment condition (2) Information	 (3) Placebo	(4) CATE	p-value (5) (2) = (3)?	 (6) 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 7: Heterogeneous treatment effects on the probability of choosing low-rating-dispersion products

Note: Participants were randomly assigned into 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.

our 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 rating dispersions (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 rating dispersions 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 likelihood of opting for products with lower rating dispersions 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 more significant 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.

7 Willingness to Pay Experiment

The second experiment has a similar setup as the first experiment. The main difference is that, unlike the previous experiment where we use conjoint analysis and ask participants to choose between products, in the second experiment, we elicit the participants' WTP. Besides, in the experiment, we investigate whether the effect of rating dispersion is heterogeneous regarding different levels of average ratings. We choose three levels of average ratings, each having three levels of rating dispersion, resulting in a combination of nine rating distributions.

7.1 Experimental Design

During the experiment, participants were presented with nine products with varying rating distributions, as shown in Figure 5. Similarly to the first experiment, products were identical in every other aspect, including price, appearance, and other product attributes, and we emphasized this to the participants. Participants in the experiment were asked a series of questions after viewing each product. These questions included determining their WTP for each product using the Becker DeGroot Marschak (BDM) mechanism to ensure honest responses (Becker et al., 1964). As in the first experiment, they were asked about the factors influencing their purchasing decisions, their perception of the likelihood of fake reviews, and their estimation of the percentage of fraudulent reviews.

We recruited $N = 812$ participants from Prolific. The participants were randomly assigned to one of two groups: an information group and a placebo group. After responding to the questions, participants in the treatment group were additionally provided with information about review authenticity for all products. This information was the same as in the first experiment but was presented via a format akin to a browser-extension-style textbox (See details in Figure 7 in

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

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

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

Figure 5: Product rating distributions in the willingness to pay experiment

Note: Products on the same row have the same average ratings (4.8, 4.4, and 4.0, respectively), but the level of dispersion is di erent.

Appendix A).[16] After being exposed to this additional information, the same set of questions 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 di erences in responses between the treatment and control 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 bachelor	0.58	0.58	0.59	0.823
Income $75,000/year	0.43	0.42	0.45	0.494
# Review websites	2.19	2.11	2.27	0.095
# Years on Amazon 10	0.46	0.47	0.44	0.290
Use Amazon 1/week	0.42	0.42	0.42	0.898
N	812	401	411	

Table 8: Balance tests of the experiment

Note: 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 insight into the validation between the treatment and control groups. The variables considered as observables in this study encompass gender, age, ethnicity, education 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.

Our analytical framework reveals how rating dispersion a ects consumer suspicion of fraudulent reviews, and how this subsequently impacts their WTP. Initially, we will use regression analysis to investigate the relationship between rating dispersion and consumer WTP. Subsequently, we will utilize the additional information treatment provided to the treatment group as an instrumental variable, enabling us to identify the e ect of consumer suspicion on WTP.

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

7.2 Average Treatment E ects

Product	(1) Both	(2) Information	(3) Placebo	(4) (2) (3)	(5) p-value
		Treatment condition			
Panel (a): Pre-treatment WTP					
High average, low dispersion	15.342	15.596	15.095	0.501	0.384
High average, med. dispersion	15.342	15.401	15.285	0.117	0.832
High average, high dispersion	15.389	15.618	15.165	0.453	0.426
Med. average, low dispersion	14.521	14.703	14.343	0.360	0.501
Med. average, med. dispersion	14.321	14.349	14.294	0.055	0.916
Med. average, high dispersion	13.606	13.576	13.635	0.059	0.911
Low average, low dispersion	12.788	12.963	12.618	0.345	0.497
Low average, med. dispersion	12.845	12.925	12.766	0.159	0.754
Low average, high dispersion	12.324	12.259	12.387	0.128	0.805
Panel (b): Post-treatment WTP					
High average, low dispersion	15.076	15.237	14.920	0.317	0.577
High average, med. dispersion	14.756	14.825	14.689	0.137	0.805
High average, high dispersion	14.685	14.449	14.915	0.466	0.412
Med. average, low dispersion	13.900	14.177	13.630	0.547	0.291
Med. average, med. dispersion	13.560	13.459	13.659	0.201	0.697
Med. average, high dispersion	12.344	11.574	13.095	1.521	0.003
Low average, low dispersion	12.395	12.870	11.932	0.938	0.072
Low average, med. dispersion	11.818	11.576	12.054	0.477	0.331
Low average, high dispersion	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

Note: Participants were randomly assigned into an information condition that received an information treatment and a placebo condition that received a placebo treatment. In column (4), we calculate the di erence 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 di erent products in both pre-treatment and post-treatment conditions. Products are categorized by three levels of average ratings (high, medium, and low) and by three levels of rating dispersion (low, medium, and high). The table underscores a prevailing preference among participants for products with high average ratings and low rating dispersions. As demonstrated, their WTP across nine product categories exhibits a general descending order, aligning well with ndings from our prior observational study.

There are noticeable di erences between the treatment and control groups in the post-treatment period for certain rating distributions. Conversely, in the pre-treatment phase, these differences are not statistically significant at the 5% level. Given the balance between the groups during the pre-treatment phase, this variation in the post-treatment phase can be interpreted as an estimate of the ATE of the information treatment. It is noteworthy that the ATE is larger in magnitude for products with lower average ratings. This is, attributable to the inherently minor differences in dispersion 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 dispersion of the products. In cases where the products have higher dispersion, the information treatment leads to a reduction in participants' WTP. For instance, for the low average, high dispersion 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 dispersion, the information treatment led to an increase in WTP. This is particularly evident in the low average, low dispersion case, where the inherent uncertainty associated with a low average is alleviated by the information treatment, resulting in a significant boost in WTP. In fact, for the low average, low dispersion 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 dispersion increases significantly, while the WTP for products with high dispersion experiences a substantial decrease. The magnitude of these effects is negatively correlated with the average rating.

7.3 Fake Review Concerns

Now we quantify the impact of the concerns about fake reviews on WTP, using a benchmark OLS specification and a series of two-stage least-squares (2SLS) specifications. We treat the exposure to the information treatment as one of the instruments for participants' concerns about fake reviews. Similarly to the first experiment, such concerns are measured by one of the following: (i) identifying fake reviews as a significant factor when expressing WTP; (ii) the perceived probability of fake reviews; (iii) the perceived proportion of fake reviews.

Considering that products have varying levels of rating dispersion, the impact of the information treatment on concerns about fake reviews could differ. Specifically, as suggested by the last subsection, the impact of the information treatment could be larger for products with higher standard deviations. Therefore, in the rst 2SLS speci cation, we instrumented the fake review concerns using the information treatment dummy, the standard deviation of the product ratings, and their interaction terms. Moreover, we include the average rating in the model as a control variable. The rst-stage regression, thus, served as a predictive model for fake review concerns based on the information treatment, the standard deviation of the product ratings, and the control variables.

To ensure the robustness of our ndings, we tested three additional speci cations. In the second 2SLS speci cation, we augmented the instruments by incorporating the average ratings and their interaction with the information treatment dummy. This addition was motivated by the potential predictive power of average ratings in relation to fake review concerns. In the third 2SLS model, we employed the information treatment dummy, dummies for each product, and interaction terms between the information treatment dummy and each product's dummy as instruments. This design allowed for di erential predictions of fake review concerns for each product, o ering a more exible approach. Finally, to enhance the robustness of our results, the last column includes the product xed e ects.

In summary, Table 10 suggests that the provision of product rating information can indeed have a signi cant impact on consumers' WTP for products.

The analysis regarding the rst measure (Table 10a) reveals a signi cant negative relationship, supported by the OLS estimation, between the variable indicating participants citing fake reviews as a reason and their WTP. This negative relationship persists across all 2SLS speci cations, indicating consistency in terms of sign, magnitude, and signi cance level. As indicated in the table, the signs of the control variables, namely average ratings and standard deviation of ratings, align with what one would expect from the results in the previous sections. Additionally, the presence of concerns regarding fake reviews may act as a mediating factor in the relationship between product rating dispersion and WTP. These ndings suggest that participants who express heightened concerns about fake reviews tend to exhibit lower WTP. Notably, when participants speci cally identi ed fake reviews as their primary concern, as compared to those who did not, their WTP decreased by $1.6 (and up to $3.0 in column 4). This decrease corresponds to a 14% (up to 25%) reduction in WTP, indicating a substantial e ect. In summary, the results consistently demonstrate a negative and statistically signi cant relationship between participants citing fake reviews as a reason and their WTP. The magnitude of this e ect can be seen as substantial, with participants expressing concerns about fake reviews exhibiting a decrease in their WTP.

	Outcome: WTP				
	(1) OLS	(2) $2SLS_1$	(3) $2SLS_2$	(4) $2SLS_3$	(5) $2SLS_4$
Avg. ratings	1.839	1.870	2.036	2.262	
	(0.669)	(0.737)	(0.735)	(0.719)	
Std. ratings	2.022	1.978	1.733	1.401	
	(0.622)	(0.770)	(0.765)	(0.730)	
List fake reviews as a reason	1.614	1.710	2.239	2.957	2.524
	(0.204)	(1.004)	(0.985)	(0.845)	(0.965)
Constant	7.925	7.766	6.888	5.699	15.521
	(3.596)	(3.948)	(3.936)	(3.854)	(0.313)
Product xed e ects	/	/	/	/	Yes
Observations	7,308	7,308	7,308	7,308	7,308
R^2	0.038	0.038	0.037	0.032	0.036
Adjusted R^2	0.038	0.038	0.036	0.032	0.035
First-stage F	/	143.180	117.336	42.195	42.195

$p < 0.1$; $p < 0.05$; $p < 0.01$

(a) Measure 1: Listing fake reviews as a reason

	Outcome: WTP				
	(1) OLS	(2) $2SLS_1$	(3) $2SLS_2$	(4) $2SLS_3$	(5) $2SLS_4$
Avg. ratings	3.085	2.496	2.945	3.017	
	(0.665)	(0.871)	(0.842)	(0.758)	
Std. ratings	0.260	1.103	0.460	0.357	
	(0.624)	(1.017)	(0.968)	(0.813)	
Fake review probability (%)	0.053	0.035	0.048	0.051	0.052
	(0.003)	(0.017)	(0.016)	(0.011)	(0.015)
Constant	2.059	4.929	2.740	2.386	16.611
	(3.566)	(4.498)	(4.366)	(3.984)	(0.505)
Product xed e ects	/	/	/	/	Yes
Observations	7,308	7,308	7,308	7,308	7,308
R^2	0.065	0.061	0.065	0.065	0.066
Adjusted R^2	0.065	0.061	0.065	0.065	0.065
First-stage F	/	225.971	190.393	75.446	75.446

$p < 0.1$; $p < 0.05$; $p < 0.01$

(b) Measure 2: Perceived probability of fake reviews

	Outcome: WTP				
	(1) OLS	(2) $2SLS_1$	(3) $2SLS_2$	(4) $2SLS_3$	(5) $2SLS_4$
Avg. ratings	2.829	2.395	2.802	2.818	
	(0.664)	(0.840)	(0.817)	(0.739)	
Std. ratings	0.561	1.201	0.602	0.578	
	(0.622)	(0.980)	(0.938)	(0.785)	
Fake review proportion (%)	0.061	0.043	0.059	0.060	0.064
	(0.004)	(0.021)	(0.020)	(0.014)	(0.018)
Constant	3.232	5.369	3.368	3.288	16.480
	(3.563)	(4.373)	(4.264)	(3.906)	(0.483)
Product xed e ects	/	/	/	/	Yes
Observations	7,308	7,308	7,308	7,308	7,308
R^2	0.063	0.060	0.063	0.063	0.063
Adjusted R^2	0.063	0.060	0.063	0.063	0.062
First-stage F	/	199.142	167.617	69.325	69.325

$p < 0.1$; $p < 0.05$; $p < 0.01$

(c) Measure 3: Perceived proportion of fake reviews

Table 10: Concerns about fake reviews negatively a ect the willingness-to-pay (WTP)

The second measure also demonstrates a negative and statistically signi cant causal relationship between the perceived fake review probability and the WTP (Table 10b). This relationship remains stable throughout all the speci cations. The signs of the coe cients on average ratings and standard deviation of ratings align with our expectations. Based on the result in column (1), a 1% increase in the di erence between the perceived probability of fake reviews leads to a $0.05 decrease in the WTP.

In the context of the third metric (Table 10c), the demonstrated patterns are similar, showing a negative and statistically signi cant causal link between the perceived fake review proportion and the WTP for the product. The directions of the control variables, namely average ratings and standard deviation of ratings, are consistent with our anticipated expectations. In line with the ndings in the initial column of the analysis, a 1% increment in the discrepancy between the perceived proportion of fake reviews results in a $0.05 reduction in WTP. This pattern persists throughout all speci cations.

8 Managerial Implications

The implications of our ndings primarily concern one key stakeholder: online platforms, which may have an interest in regulating the market.

In the context of online platforms such as Amazon, the process of identifying and removing fake reviews takes about six months. This delay could be due to limited detection capabilities, or it might involve a strategic component that explains why Amazon deliberately takes this approach. If consumers favor products with more reviews on competing platforms like eBay, Amazon might permit fake reviews to persist, particularly for new products. Consequently, policy implications may not always suggest that stricter policies are universally bene cial for the platform.

Furthermore, Amazon could emphasize reviews authored by users with more extensive and credible histories. While assigning di erent 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 the information treatment used in our study. 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 grow more discerning, sellers might be less tempted to purchase fake reviews, consequently reducing the prevalence of such practices. In the long term, this could lead to heightened consumer trust in the platform and a healthier online marketplace.

In conclusion, our research findings provide insights for both online platforms and merchants, emphasizing the necessity of a multi-faceted approach to addressing the fake review problem. Platforms like Amazon should 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, merchants 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 merchants can contribute to the development of a more trustworthy and transparent online shopping environment.

9 Future Directions

The present research offers insights into the relationship between concerns about fake reviews and consumer demand. However, there are several aspects that warrant further investigation and potential avenues for future research.

First, while our study focuses on the effects of information treatment on consumer behavior, it is crucial to recognize that different types of information and their presentation may yield varying results. Future research could explore the impact of different information formats (e.g., text, videos, infographics) and the level of detail provided in the information treatment. This would help platforms and merchants understand the most effective ways to communicate with consumers and educate them about the possibility of fake reviews.

Second, our research primarily investigates the short-term consequences of fake reviews on consumer behavior. Future studies could examine the long-term effects of fake reviews on brand reputation, customer loyalty, and customer lifetime value. Understanding these long-term implications would provide valuable insights for merchants, helping them weigh the potential benefits and costs of engaging in fake review practices.

Third, our findings are based on an experimental setting, which may not perfectly capture the nuances of real-world online shopping experiences. Future research could leverage observational data from online platforms, tracking the changes in consumer behavior and market dynamics in response to fake review detection and deletion. Such studies could o er a more comprehensive understanding of how the online marketplace reacts and adapts to the ongoing market interaction with fake reviews.

Lastly, while this research focuses on the implications for platforms and merchants, it is essential to consider the broader societal impact of fake reviews. Future research could investigate the repercussions of fake reviews on consumer trust in online marketplaces and the potential spillover e ects on other industries or sectors. Additionally, studies could explore the role of government regulations and industry standards in managing the existence of fake reviews, comparing the e ectiveness of di erent approaches across countries or regions.

Declarations

Funding and Competing Interests

All authors certify that they have no a liations with or involvement in any organization or entity with any nancial interest or non- nancial interest in the subject matter or materials discussed in this manuscript.

The authors received funding from Haas Behavioral Lab, Haas Center For Equity, Gender & Leadership, and Xlab at the University of California, Berkeley.

References

Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Learning from reviews: The selection e ect and the speed of learning. *Econometrica*, 90(6):2857 2899. 5

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral science*, 9(3):226 232. 27

Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54 78. 4

Chakraborty, I., Deb, J., and Oery, A. (2022). When do consumers talk? *Available at SSRN 4155523*. 3, 4

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. 4

Chevalier, J. and Goolsbee, A. (2003). Measuring prices and price competition online: Amazon. com and barnesandnoble. com. *Quantitative marketing and Economics*, 1:203 222. 10

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. 4

Dai, W., Jin, G., Lee, J., and Luca, M. (2018). Aggregation of consumer ratings: an application to yelp. com. *Quantitative Marketing and Economics*, 16:289 339. 5

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. 4

Economist (2020). A new study analyzes the murky world of fake amazon reviews. 18

Federal Trade Commission (2022). Ftc to explore rulemaking to combat fake reviews and other deceptive endorsements. 3

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. 3, 7, 9

He, S., Hollenbeck, B., and Proserpio, D. (2022b). The market for fake reviews. *Marketing Science*, 41(5):896 921. 3, 10

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. 4

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. 4

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

Luca, M. (2016). Reviews, reputation, and revenue: The case of yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, (12-016). 4

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427. 4

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. 4

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–2455. 4

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. 4

Nosko, C. and Tadelis, S. (2015). The limits of reputation in platform markets: An empirical analysis and field experiment. Technical report, National Bureau of Economic Research. 5

Park, S., Shin, W., and Xie, J. (2021). The fateful first consumer review. *Marketing Science*, 40(3):481–507. 4

Pei, A. and Mayzlin, D. (2022). Influencing social media influencers through affiliation. *Marketing Science*, 41(3):593–615. 4

Proserpio, D. and Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5):645–665. 4

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. 4

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. 4

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. 4

Shi, Z., Srinivasan, K., and Zhang, K. (2022). Design of platform reputation systems: Optimal information disclosure. *Marketing Science*. 5

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

Timoshenko, A. and Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, 38(1):1–20. 5

Vellodi, N. (2018). Ratings design and barriers to entry. *Available at SSRN 3267061*. 5

Wu, C., Che, H., Chan, T. Y., and Lu, X. (2015). The economic value of online reviews. *Marketing Science*, 34(5):739–754. 4

Zhong, Z. (2022). Chasing diamonds and crowns: Consumer limited attention and seller response. *Management Science*, 68(6):4380–4397. 4

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. 4

A Experiment Details

A.1 Reason of Choice

Before launching the experiments, we run a survey on Prolific and collect reasons why participants choose products with a high or low rating dispersion. We then use an online natural language processing tool to summarize the reasons for choosing products with a high or low rating dispersion, respectively. In the experiment, each list of reasons is shown to the participants in a random order.

For those who choose products with a lower rating dispersion, the reasons are:

- The other product may have purchased many fake 5-star reviews, leading some customers to leave 1-star reviews in retaliation.
- I prefer products with less 1-star ratings since I don't have the time to analyze the distribution of reviews in detail.
- This product does not have extremely negative reviews.
- The other product may have a certain fail rate, meaning that it sometimes works and sometimes does not.
- The other product may have inaccurate descriptions, causing people's opinions to be more diverse when receiving the product.
- The other product may attract customers with more varied tastes.

For those who choose products with a higher rating dispersion, the reasons are:

- I prefer products with more 5-star ratings since I don't have the time to analyze the distribution of reviews in detail.
- This product has more haters, without which the average rating should be higher.
- 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.
- Despite more 1-star negative reviews, I am confident in the effectiveness of this product as an expert in this field.
- 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.