NBER WORKING PAPER SERIES

MISINFORMATION AND MISTRUST: THE EQUILIBRIUM EFFECTS OF FAKE REVIEWS ON AMAZON.COM

Ashvin Gandhi
Brett Hollenbeck
Zhijian Li

Working Paper 34161
http://www.nber.org/papers/w34161

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2025

We thank Jinzhao Du, Liran Einav, Jessica Fong, Sherry He, Ginger Jin, Malika Korganbekova, Tin Cheuk Leung, Eddie Ning, Ye im Orhun, Ariel Pakes, Jesse Shapiro, Andrey Simonov, Ben Vatter, and Joel Waldfogel, for helpful comments, as well as seminar participants at UPenn, UT Austin McCombs, Santa Clara University, UCSD Rady, UC Riverside, the University of Toronto - Rotman, Yale SOM, Columbia GSB, Tilburg University, Harvard Business School, the FTC Microeconomics Conference, the Bass Forms Conference, the IIOC Conference, the Hal White Antitrust Conference, the Summer Institute in Competitive Strategy, the Southern California Strategy Conference, the Quantitative Marketing and Economics Conference, the BIOMS Conference, the Harvard Business School Digital Competition and Technology Regulation Conference, the Columbia-Wharton Management, Analytics, and Data Conference, the Workshop on Platform Analytics, and the Spring 2025 NBER Industrial Organization meeting. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Misinformation and Mistrust: The Equilibrium Effects of Fake Reviews on Amazon.com
Ashvin Gandhi, Brett Hollenbeck, and Zhijian Li
NBER Working Paper No. 34161
August 2025
JEL No. L00, L1, L10, L11, L15, M3, M31, M38

ABSTRACT

Fake product reviews—and the manipulation of reputation systems by sellers more broadly—are a widespread issue for two-sided platforms. We study two primary channels through which such manipulation can affect market outcomes: (i) creating misinformation about the reviewed product, and (ii) breeding mistrust in ratings system overall. To examine these in the Amazon.com marketplace, we measure misinformation by observing products purchasing fake reviews and measure mistrust by eliciting shoppers' beliefs about the prevalence of fake reviews on Amazon through an incentivized survey experiment. We incorporate these into a structural model of demand in which consumers form beliefs about product quality based on observed reviews and perceptions about their trustworthiness. Counterfactual policy simulations indicate that fake reviews reduce consumer welfare, shift sales from honest to dishonest sellers, and ultimately harm the platform. Welfare losses arise primarily from misinformation that leads to worse purchases. While mistrust also leads to purchasing mistakes, the consumer harms of mistrust are largely offset by increased price competition under a weakened ratings system. Finally, we identify key limitations in platforms' incentives to police manipulation and evaluate enforcement alternatives.

Ashvin Gandhi
University of California, Los Angeles
Anderson School of Management
and NBER
ashvin.gandhi@anderson.ucla.edu

Brett Hollenbeck
University of California, Los Angeles
brett.hollenbeck@gmail.com

Zhijian Li
Northwestern University
zhijian.li@kellogg.northwestern.edu

1 Overview

User-generated ratings and reviews are a core feature behind of the success of online marketplaces (Cabral and Hortacsu, 2010; Tadelis, 2016; Einav et al., 2016). These systems ameliorate asymmetric information by allowing sellers to establish reputations. Surveys show that an overwhelming majority of consumers consult reviews before making purchases. As a result, reputation systems have large impacts on sellers' outcomes and marketplace success, not just online but in many settings such as restaurants, hotels, and healthcare. The importance of these mechanisms creates a powerful incentive for sellers to manipulate their ratings, and recent research has documented that rating manipulation using fake reviews purchased by the seller is widespread (He et al., 2022b; FTC, 2023). With the rising salience of these practices, there has been widespread interest by consumer protection regulators around the globe: the FTC, the UK CMA, the European Commission, and others are all investigating the potential consumer harms from rating manipulation and in some cases have proposed or implemented laws and regulations in response (CMA, 2020; FTC, 2024).

In this paper, we study the implications of rating manipulation through fake reviews for sellers, consumers, and platforms using the setting of the Amazon.com marketplace. We propose two primary channels by which ratings manipulation can shift outcomes. The first channel is that fake reviews create misinformation. By inflating ratings, fake reviews misinform consumers and may mislead them into making different and possibly worse purchasing decisions. The presence of misinformation in markets can also shift equilibrium prices. Products purchasing fake reviews appear higher in quality and can increase prices, while honest products may lower prices to compete with manipulators.

The second channel is that the widespread presence of fake reviews may cause consumers to generally mistrust ratings. This mistrust lessens the ability of the ratings system to solve the asymmetric information problem and may therefore result in worse purchasing decisions. At the same time, if mistrust in ratings makes high-quality products less able to differentiate

from low-quality products, consumers may benefit from increased price competition.

The relative magnitudes of these different forces are unknown, and thus the net impact on aggregate welfare is ambiguous.[1] We quantify these impacts using a model of how consumers form beliefs and make purchasing decisions based on ratings, to which we bring novel data on the Amazon marketplace that includes which products are actually using fake reviews.

To measure fake review activity, we follow He et al. (2022b) in using a hand-collected panel on approximately 1,500 products that purchased fake reviews from private Facebook groups where sellers solicit fake Amazon reviews.[2] We supplement this with a scraped panel of Amazon data for these rating manipulators and a set of their close competitors that includes weekly data on ratings, reviews, sales ranks, prices, and advertising.

The principal component of our model is how Bayesian consumers form beliefs about product quality from ratings, taking into account the possibility of ratings manipulation. To inform key assumptions on consumers' beliefs about the prevalence of fake reviews, we conduct a set of incentivized survey experiments on Amazon shoppers. We also assess the extent to which consumers can detect ratings manipulation and find that, while consumers have reasonable beliefs about the general prevalence of fake reviews, they do poorly at identifying specifically which products use them.

We then estimate a structural model of demand following Berry et al. (1995) that incorporates Bayesian consumers' perception of product quality based on ratings. This models consumer demand as a function of their beliefs over product quality derived from ratings and not simply the ratings themselves. Therefore, the same observed ratings can yield differing demand depending on consumers' beliefs about the presence of fake reviews. This lets us simulate how demand would change not only under different observed ratings but also under different consumer perceptions about the prevalence of fake reviews.

To evaluate the impact of fake reviews, we consider a series of counterfactual policy anal-

[1] A third channel by which ratings manipulation may impact consumers is through dynamic effects, namely the extent to which paying for reviews lowers or raises barriers to entry for high-quality entrants.

[2] While we focus on fake Amazon reviews, similar marketplaces exist for other e-commerce platforms like Wayfair, Walmart, Yelp, and so on.

yses that isolate the different mechanisms at play. We use our knowledge of which products use fake reviews, as well as estimates of the proportion of their reviews that are fake, to adjust products' ratings and consumers' beliefs to what would occur if the platform or regulator had removed or prevented all fake reviews. We then recompute equilibrium prices and calculate consumer welfare and firm profits when fake reviews are present versus when they are absent. In addition, we simulate counterfactuals that isolate the effects of misinformation and mistrust. We isolate misinformation by simulating the market equilibrium in which fake reviews exist but consumers fully trust reviews as if they did not. We isolate mistrust by simulating the market equilibrium without fake reviews but in which consumers still perceive them as prevalent. In all cases, we show results both fixing prices and allowing them to adjust in order to understand the role of competitive responses.

We find that, overall, fake reviews benefit manipulators at the expense of honest products and consumers. Manipulators are enabled to sell 27.2% more units while raising prices by an average of \$0.31. On the other hand, honest products competing against manipulators sell 4.4% fewer units and lower their prices by an average of \$0.07. While consumers do benefit from this discounting of honest products, it does not fully negate the harms of fake reviews. On net, manipulation reduces consumer welfare by 0.77% in the markets we study.

The overall effects mask important differences in the impacts of misinformation and mistrust. We examine these by considering each in isolation. Misinformation misleads consumers into overestimating manipulators' quality. This substantially increases sales and profitability of manipulators but causes an average of \$3.70 in harm to the consumers misled into purchasing lower-quality products with inflated ratings and unduly high prices. It also harms honest competitors, who lose market share, must lower prices, and ultimately lose profit. Mistrust, on the other hand, causes consumers to slightly underestimate quality for virtually all products. This reduces demand and profits for both manipulators and honest products. It also leads consumers to sometimes mistakenly purchase the wrong product or no product at all. These mistakes tend to be more minor, however, averaging just \$1.73. More

importantly, they are almost entirely offset by the benefits of strengthened price competition: because mistrust weakens products' ability to differentiate on quality, both manipulators and honest products reduce their prices on average.

We also study the platform's incentives to combat fake reviews. While our estimates indicate that Amazon is harmed by fake reviews, they also suggest a few key factors that might hinder enforcement. The first is that, while platform revenue does increase with consumer trust, it also increases with misinformation. As a result, simply removing misinformation would actually harm Amazon unless done in a way that was credible and increased consumers' trust. Indeed, if it were feasible, the platform would most prefer to improve trust without substantially reducing misinformation. Likewise, we find that Amazon and consumers are poorly aligned on which fake reviews to prioritize removing. While a relatively small fraction of reviews are responsible for the majority of consumer harm, these reviews also tend to generate substantial additional revenue for Amazon. One alternative that Amazon might consider would be to increase the prevalence of organic reviews. We find that an 48% increase in organic reviews achieves the same welfare gains as eliminating fake reviews and increases platform revenues by $1.84 per additional review.

We contribute to several strands of literature related to information disclosure, platform design, and reputation manipulation. First, and most directly, we contribute to the growing literature on fake reviews which begins with Mayzlin et al. (2014) and Luca and Zervas (2016). Theoretical work on fake reviews has shown that under reasonable circumstances, fake reviews can be efficient and welfare-enhancing. In an extension of the signal-jamming literature on how firms can manipulate strategic variables to distort beliefs, Dellarocas (2006) shows that fake reviews are mainly purchased by high-quality sellers and, therefore, increase market information under the condition that demand increases convexly with respect to user rating. Given how ratings influence search results, it is plausible that this condition holds. Other research modeling fake reviews have also concluded that they may benefit consumers and markets (see Glazer et al. (2020), Saraiva (2020), and Yasui (2020).) Similarly, Johnen

and Ng (2024) considers the welfare gains from sellers lowering their prices to induce positive ratings. These are full equilibrium models of the seller decision to use fake reviews in which consumer beliefs rationally forecast equilibrium seller behavior. Our theoretical framework instead allows consumers to have a range of beliefs, including being naive with respect to the presence and prevalence of fake reviews, but as a consequence should be thought of as a partial equilibrium model.

There have been few attempts to empirically test or quantify the predictions of these models or to empirically assess the impact of fake reviews on welfare or competition. An exception is Akesson et al. (2022), who conduct an incentive-compatible online experiment in which products are shown with random variation in whether and how fake reviews appear. They find via choice tasks that the presence of fake reviews makes consumers more likely to purchase lower-quality products and estimate a welfare loss of $.12 for each dollar spent from this mechanism. This experiment therefore captures the direct effect of misinformation, but does not try to quantify the indirect effects of the change in equilibrium prices that result and does not address the effects of mistrust. Another closely related work is Li et al. (2020), an examination of incentivized reviews on Taobao. They find that high-quality sellers select into the incentivized review system and this improves market efficiency. There are several distinguishing features of incentivised reviews, compared to fake reviews, that we describe in more detail below. While not considering fake reviews, Reimers and Waldfogel (2021) study the welfare impact of consumer reviews as a whole, showing that Amazon user reviews have a large impact on consumer surplus.

We also contribute to an emerging literature on information disclosure. Dranove and Jin (2010) summarize a large body of research on quality disclosure, with a focus on voluntary firm disclosure. When a platform acts as an intermediary and designs a system of quality disclosure, new and complex incentives around competition and welfare arise.[3] Armstrong and Zhou (2022) provide a general treatment of the issues around information signals and

[3]Notable related work on platform reputation systems includes Dai et al. (2018), Hui et al. (2016), Hui et al. (2022), and Chakraborty et al. (2022).

competition, and show that a policy that dampens differentiation can intensify competition and benefit consumers.[4] Hopenhayn and Saeedi (2023) characterize an optimal rating system in the presence of competition and adverse selection by sellers. They show that more precise quality ratings does not always benefit consumers. In ongoing work, Saeedi and Shourideh (2020) studies optimal ratings when firms can potentially manipulate ratings. Vatter (2021) also shows that full information disclosure is not optimal, and characterizes optimal quality scores in the context of Medicare Advantage. Our contributions to this literature are, first, to show how endogenous mistrust of disclosed information could produce similar results as coarse disclosure, and second, empirically characterizing whether consumers are better off by placing less trust in quality ratings.

2 A Simple Model of Misinformation and Mistrust

In this section, we illustrate the different ways that rating manipulation can impact consumer choices and firm outcomes. We present a simple model in which consumers make purchases based on observed product features and user ratings that provide a signal of quality. We divide our analysis into two distinct effects. The first, which we refer to as the "misinformation effect" of rating manipulation, is that fake reviews provide false information that can mislead consumers into making different purchasing decisions. This is the direct effect that purchasing fake reviews has on a product's sales and the sales of its competitors. The second, which we refer to as the "mistrust effect," is the change in outcomes that results from consumer beliefs that some reviews are fake. Mistrust is a more systemic effect, determined by the overall prevalence of fake review purchasing and not the specific purchasing

[4]Related work by Vellodi (2018) focuses on dynamics, and shows that suppressing the reviews of highly-rated firms can stimulate entry and improve consumer welfare through that channel.

of any individual product.[5] Indeed, the effect of mistrust can be felt even in markets where no products have purchased fake reviews. Finally, while misinformation and mistrust represent effects on consumers' behavior, it is important to note that both also affect the equilibrium pricing behavior of both fake review purchasers and honest products.

2.1 Misinformation

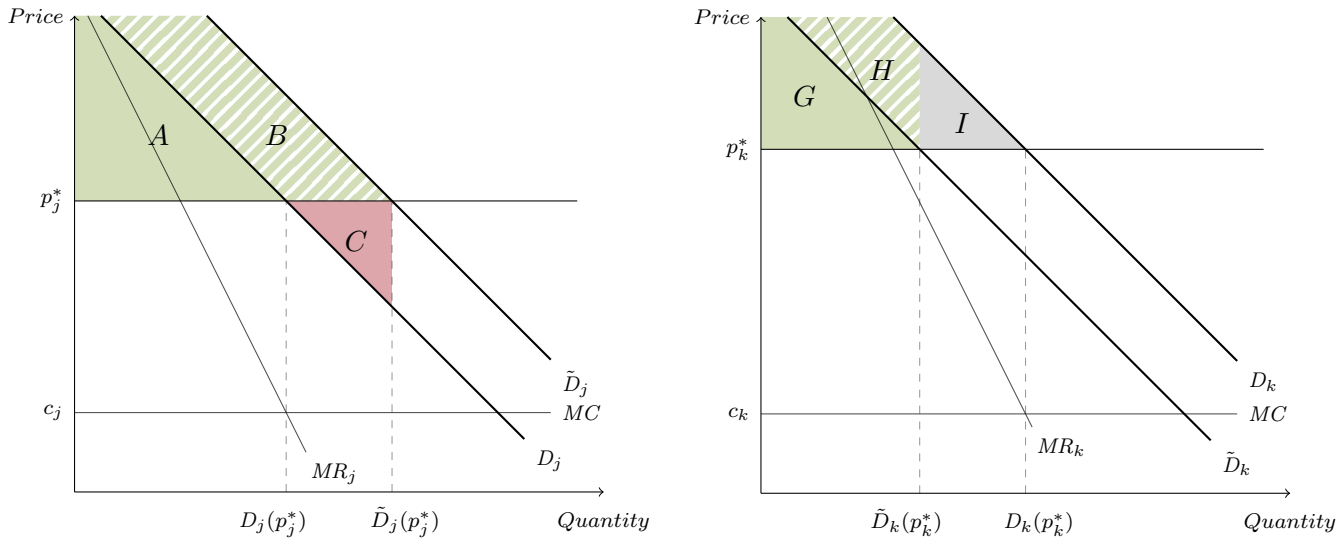
We model consumers' utility from a product j as decreasing in price (p_j) and increasing in quality (q_j). However, when making purchasing decisions, consumers do not directly observe a product's quality and must infer it from the product's reviews (R_j). In our empirical exercise, we think of R_j as a set of reviews that imperfectly reveal a product's quality. However, for simplicity in this toy model, we let R_j be a scalar rating that aggregates all of j's reviews and perfectly reflects j's true quality when j does not purchase fake reviews. Formally, we let $q_j, R_j \in (0, 1)$ and $q_j = R_j$ when j does not purchase fake reviews. On the other hand, if a product purchases fake reviews, then $R_j \geq q_j$, and the ratings no longer perfectly reflects the true quality. We denote j purchasing or not purchasing fake reviews by F_j and $\neg F_j$, respectively.

Our assumptions imply that in a world without fake reviews, rational consumers will interpret a product's rating to be its quality. We describe a consumer as being "trusting" if they interpret reviews in this way. To best illustrate the effect of misinformation, we first consider how fake reviews impact a market with trusting consumers. Such circumstances might reasonably describe settings in which ratings manipulation is too rare, too new, or too difficult to detect, such that consumers have not yet developed meaningful mistrust.

We consider a market with two competing products, j and k. When qualities are observed

[5]The idea of dividing consumer welfare effects into effects from misinformation and mistrust separately has been anticipated in the early literature on fake reviews, most notably Mayzlin et al. (2014), who note in their introduction: "...fake reviews may have at least two deleterious effects on consumer and producer surplus. First, consumers who are fooled by the promotional reviews may make suboptimal choices. Second, the potential presence of biased reviews may lead consumers to mistrust reviews. This in turn forces consumers to disregard or underweight helpful information posted by disinterested reviewers." Here we formalize these two effects and in the following sections we empirically quantify them.

Figure 1: Effect of Misinformation (No Price Changes)



(a) Fake Review Purchaser

(b) Non-Purchaser

by consumers, the demand for product j is $D_j(p_j, q_j, p_k, q_k)$. However, since consumers cannot observe qualities directly, they purchase based on observable ratings. Trusting consumers believe $R_j = q_j$ and $R_k = q_k$, so their demand is characterized by $D_j(p_j, R_j, p_k, R_k)$.

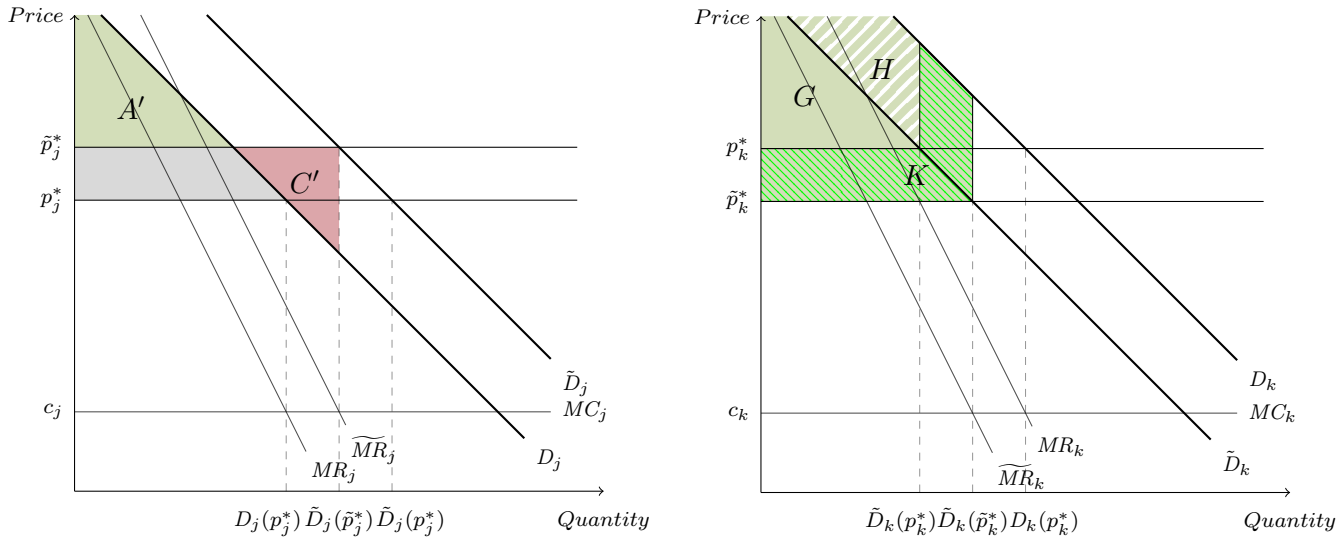
If product j purchases fake reviews, then this increases R_j above q_j and shifts out the demand curve for product j and shifts in the demand curve for competitor product k. Figure 1 shows the effect of these demand shifts when holding prices fixed at the level that would have prevailed without fake reviews. The demand curves D_j and D_k are those that would occur absent fake reviews—i.e., when R_j and R_k accurately reflect q_j and q_k—while $\tilde{D}_j$ and $\tilde{D}_k$ characterize consumer demand given that j purchases fake reviews. Note that while fake reviews cause consumers to purchase according to $\tilde{D}_j$ and $\tilde{D}_k$, the utility actually realized from their purchases are characterized by D_j and D_k. Put simply, the misinformation from fake reviews causes consumers to purchase according to demand curves that do not reflect their informed preferences.

For product j, this entails an increase in quantity demanded from $D_j(p_j^*)$ to $\tilde{D}_j(p_j^*)$, increasing j's profits by $\left(p_j^* - c_j\right)\left(\tilde{D}(p_j^*) - D(p_j^*)\right)$. Consumers purchasing based on $\tilde{D}_j$ anticipate a total consumer surplus of $A + B$. In actuality, however, consumer surplus for

those purchasing j is much lower at $A-C$. Note that while fake reviews cause all consumers to overestimate the utility of purchasing j, not all purchasers of j are actually harmed. For the $D_j(p_j^*)$ consumers who would have purchased j even absent fake reviews, region B only represents a failure of j to meet expectations and not an actual loss in utility. The true harms are borne by the $\tilde{D}_j(p_j^*) - D_j(p_j^*)$ consumers induced to purchase product j by its fake reviews. These consumers would have been better off either purchasing k or nothing at all, and region C represents forgone utility from making a sub-optimal purchasing decision due to misinformation.

Product k, on the other hand, experiences a reduction in demand from $D(p_k^*)$ to $\tilde{D}_k(p_k^*)$, which reduces profits by $(p_k^* - c_k)\left(D_k(p_k^*) - \tilde{D}_k(p_k^*)\right)$. Consumers purchasing based on $\tilde{D}_k$ anticipate receiving consumer surplus G. However, these $\tilde{D}_k(p_k^*)$ consumers underestimate their surplus by H because alternative j is actually worse than its ratings suggest. Of course, these consumers would have purchased k even absent fake reviews, so H does not represent a real benefit. In contrast, the $D_k(p_k^*) - \tilde{D}_k(p_k^*)$ consumers induced by fake reviews to purchase j instead of k experience a real harm shown in region I.[6]

Figure 2: Competitive Responses to Misinformation



(a) Fake Review Purchaser

(b) Non-Purchaser

[6]Note that if fake reviews only steal market share and do not expand total purchasing in the market, then C and I represent the same harms due to misinformation.

Competitive Responses Of course, both firms should adjust their prices in response to j purchasing fake reviews. Figure 2a depicts these competitive responses. The increase in demand from D_j to $\tilde{D}_j$ raises j's optimal price from p_j^* to $\tilde{p}_j^*$.[7] By raising price, j further increases its profit by $\left(\tilde{p}_j^* - c_j\right) \tilde{D}_j(\tilde{p}_j^*)$ and shrinks consumer surplus from $A - C$ to $A' - C'$.[8] Importantly, this price increase harms the $D_j(p_j^*)$ consumers who would have purchased product j even absent fake reviews. It also exacerbates the harms to the $\tilde{D}_j(\tilde{p}_j^*) - D_j(p_j^*)$ consumers still misled into purchasing j even at the higher price. On the other hand, the $\tilde{D}_j(p_j^*) - \tilde{D}_j(\tilde{p}_j^*)$ consumers dissuaded from purchasing j by the price increase actually benefit from the competitive response.

In contrast, the decrease in demand from D_k to $\tilde{D}_k$ lowers k's optimal price from p_k^* to $\tilde{p}_k^*$. By cutting price, k stems its losses to j and earns a profit of $(\tilde{p}_k^* - c_k)\, \tilde{D}_k(\tilde{p}_k^*) > (p_k^* - c_k)\, \tilde{D}_k(p_k^*)$. This also benefits consumers, who see their surplus increase by region K. Indeed, $\tilde{D}_k(\tilde{p}_k^*)$ who still purchase k in spite of j's fake reviews now receive a discount that makes them better off than if j had not purchased fake reviews. This shows that misinformation is not unambiguously bad for consumers, as competitive responses benefit those still purchasing honest products. Which effects dominate ultimately depends on the relative sizes of both the price and quality elasticities of demand.

2.2 Mistrust

Thus far, we have modeled consumers as fully trusting reviews in order to isolate the effect of misinformation. Over time, however, consumers may learn from media, word of mouth, or personal experience that some products' ratings have been manipulated by fake reviews. In this section, we explore the implications of the mistrust in the rating system that occurs when consumers are generally aware of fake reviews but do not know precisely

[7] It is important to note that the competitive responses must solve in equilibrium. As j increases its price, this attenuates the inward shift in k's residual demand curve. Likewise, as k decreases its price, this attenuates the outward shift in j's demand curve. Therefore, when incorporating competitive responses, the equilibrium shifts in demand for j and k are smaller than in Figure 1.

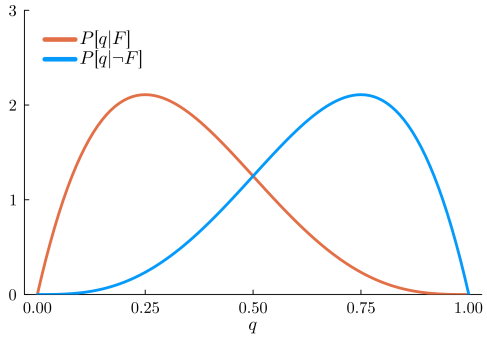
[8] In this example with linear demand and fake reviews shifting only the level of demand, $C' = C$, so the welfare loss is simply $A - A'$.

which ratings are manipulated. To do this, we allow consumers to incorporate the possibility that ratings were manipulated when forming expectations of product quality. Note that we largely suppress product subscripts in order to emphasize that the effect of mistrust works through consumers' beliefs and not a given product's behavior. Indeed, mistrust may affect a market even if none of the products in that specific market purchase fake reviews so long as consumers believe that some products could be doing so.

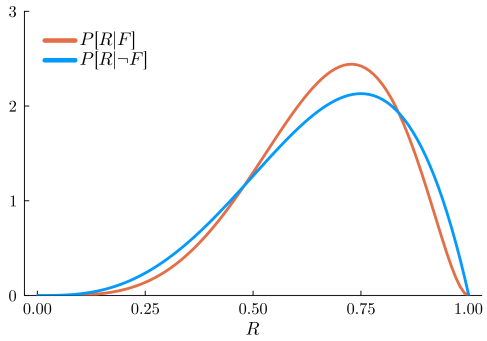
We start by modeling a consumer who cannot identify which products are purchasing fake reviews but has rational expectations about the prevalence of fake reviews. In considering a product with rating R, the mistrustful consumer anticipates some probability $P(F|R) > 0$ that the product purchased fake reviews. If it did, then its rating is inflated, so the expected quality $E[q|R, F]$ is less than R. If it didn't, then R accurately reflects quality. Therefore, the mistrustful consumer forms an expectation about quality that places weight $P(F|R)$ on $E[q|R, F]$ and weight $1 - P(F|R)$ on R:

$$E[q|R] = P(F|R)\, E[q|R, F] + (1 - P(F|R))\, R. \tag{1}$$

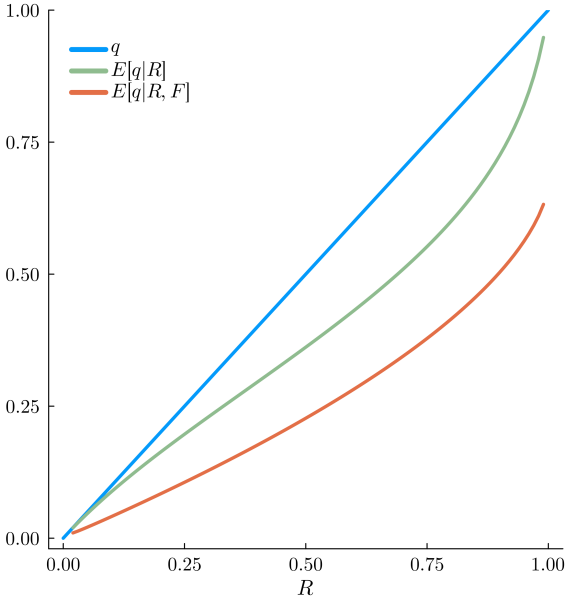
Figure 3: An Illustrative Example



(a) Quality



(b) Ratings



(c) Expected Quality

Notes. True qualities for fake review purchasers and honest products are Beta$(2, 4)$ and Beta$(4, 2)$, respectively. Purchasing fake reviews boosts the rating of a product with quality q by $(1-q)\nu$, where $\nu \sim$ Beta$(3, 3)$. See Appendix A.1 for additional details.

Figures 3 provides an illustrative example in which 50% of products purchase fake reviews. In this example, the products that purchase fake reviews tend to have lower qualities (Figure 3a), and in doing so, it improves their ratings to be fairly similar to the ratings for honest products (Figure 3b). See Appendix A.1 for details.

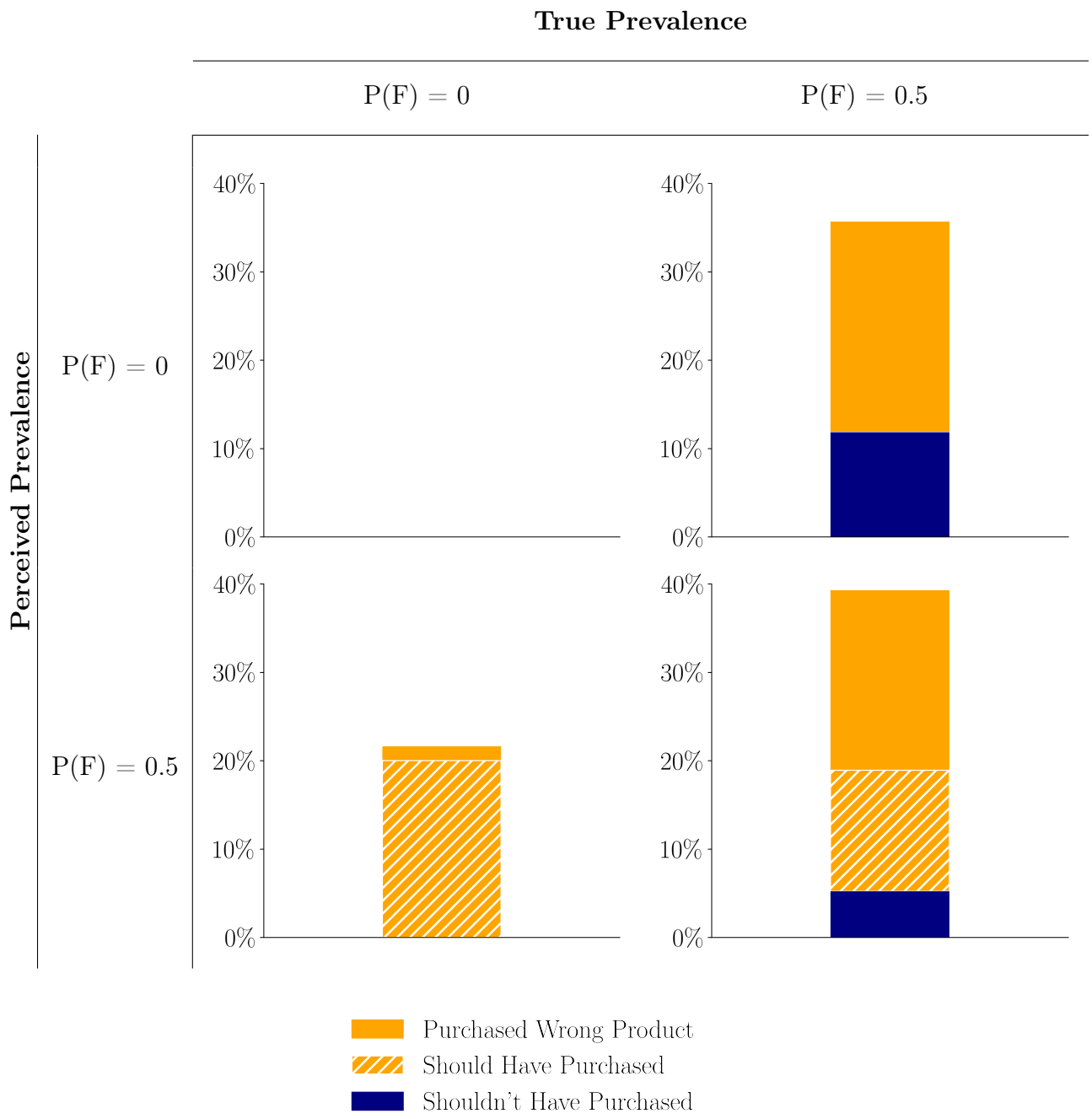
Figure 3c illustrates how a Bayesian consumer with rational expectations infers quality from R. The top line shows R, the quality that the consumer would infer if she were trusting or knew with certainty that the product did not purchase fake reviews. The bottom curve gives $E[q|R, F]$, the expected quality that the Bayesian consumer with rational expectations would infer if she knew for certain that the product purchased fake reviews. Finally,

the middle curve gives $E[q|R]$, the quality that the consumer infers from R given rational expectations about the prevalence of fake reviews and the joint distribution of q and R.

There are a number of instructive features of Figure 3c. The first is that $E[q|R] \leq R$, so mistrust causes consumers to anticipate lower true quality for any given rating. This makes any product less attractive, and all else equal, should reduce purchasing. In fact, if $E[q|R]$ were simply a parallel shift downward from R, the only effect of mistrust would be to shift demand to the outside good. However, the shift downward is not parallel because the mistrusting Bayesian discounts their expectation differently depending on the product's observed rating. Specifically, the Bayesian consumer discounts their expectations most heavily when a product's rating indicates that it likely purchased fake reviews—i.e., $P(F|R)$ is large—or that the products achieving such a rating through manipulation are particularly bad—i.e., $E[q|R, F]$ is much lower than R.

It is important to re-emphasize that the scope of the effect of mistrust may be particularly large because it affects both products that did and did not purchase fake reviews similarly. In fact, it can affect markets in which no products actually purchased fake reviews as long as consumers perceive some probability that they could have. They are also difficult to measure or directly observe since they stem from consumers' perceptions. Finally, they may be difficult to attribute to individual actors, since the change in consumers' beliefs about the relationship between ratings and quality stems from the general prevalence of fake reviews and is not meaningfully shifted by the individual decisions of any single product.

Relaxing Rational Expectations There are a number of reasons that consumers' beliefs about fake reviews may not satisfy rational expectations. For example, consumers may under- or overestimate the prevalence of fake reviews yielding inaccurate beliefs about $P(F|R)$. Likewise, consumers may misunderstand how much fake reviews move R and therefore infer $E[q|R, F]$ incorrectly. Relaxing rational expectations simply requires specifying how the beliefs in equation 1 are determined. In our empirical exercise, we characterize these beliefs using a survey experiment.



Note: All plots are simulated with 10000 random draws from the Beta distributions and 10000 customers, assuming the outside option quality is 0.5. The randomness from the customers is modeled by $Gumbel(0, 0.1)$.

Figure 4: Percentage of Wrong Choices Under Misinfo and Mistrust

Comparing the Effects of Misinformation and Mistrust Of course, both misinformation and mistrust are likely to be present in many markets. Therefore, we return to the illustrative example from Section 2.2 and compare how misinformation and mistrust shift consumer choices. Figure 4 depicts four scenarios in four quadrants, which vary based on the true prevalence of fake reviews (i.e., misinformation) and the perceived prevalence (i.e., mistrust). In the upper-left quadrant, neither misinformation nor mistrust are present, while in the bottom right quadrant, both are present.

When there is only misinformation (upper-right), consumers buy too many product that purchased fake reviews. If fully informed, these consumers would have preferred to purchase other honest products (orange) or not to have purchased at all (blue). When there is only mistrust (bottom-left), the primary distortion in choices is that consumers buy too few products from the marketplace and shift those purchases to the outside option.[9] Finally, when there is both misinformation and mistrust, consumers make all three types sub-optimal choices: they purchase the wrong product, purchase when they should not have, and do not purchase when they should have.

In sum, this toy example suggests that the ultimate implications of misinformation and mistrust for substitution patterns are highly dependent on many empirical factors, including the shape of consumer demand, the prevalence and magnitude of fake reviews, and the distribution of quality for both fake review purchasers and honest products. This underscores the importance of the empirical exercise that we explore in the remainder of our paper.

3 Data

The principal aim of our empirical exercise is to understand the equilibrium impacts of fake reviews on the Amazon marketplace. This requires estimates of consumer demand—especially how demand changes with ratings—as well as information on which products are

[9]Note that neither of the off-diagonal outcomes is a full equilibrium outcome because beliefs and the underlying state of the world are misaligned. These should be interpreted as comparative statics meant to isolate the different mechanisms.

purchasing fake reviews and the extent of their manipulation. In this section, we describe our data on Amazon products used for this analysis.

The primary marketplace for purchasing fake Amazon reviews are a set of private Facebook groups (He et al., 2022b). Amazon sellers wishing to purchase fake reviews post their product to one of these groups and offer to pay for five-star reviews.[10] Interested members privately message the seller to coordinate the transaction. The typical terms essentially entail that the reviewer receives the product for free in return for a positive fake review.[11] In some cases, the reviewer also receives a small commission of around $5 to $10.

Once the terms are set, the reviewer purchases the product on Amazon.com and leaves an authentic-seeming "verified purchase" review that is 5-stars and includes positive text. When the five-star review posts to the product page, the seller reimburses the reviewer via PayPal for the purchase (including taxes and fees) and pays any agreed-upon commission.

We obtain data on fake review activity by collecting information directly from the private Facebook groups where fake reviews are purchased. As scraping Facebook is technically infeasible, this required using a team of research assistants to monitor the top fake review Facebook groups and hand-collect data on a random sample of posting products, including the period over which they were actively recruiting fake reviews. He et al. (2022b) detail these groups and the data collection process. Our data include information on a set of roughly 1,500 unique products observed buying fake reviews between October 2019 and June 2020.

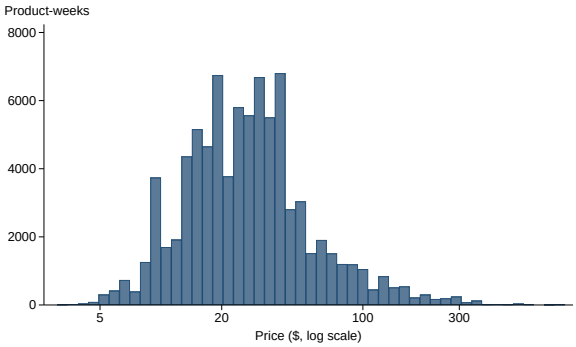
We also conduct a large-scale repeated scraping of Amazon.com during and after this time period. This scraping centers on searches of the product keyword identified by the seller of a product that was soliciting fake reviews. For each keyword, we collect daily data on the full set of products returned in the search. This includes each product's position

[10] In addition to relying on private groups, sellers often take additional steps to avoid detection by Amazon and authorities. First, sellers often use brokers as intermediaries. Second, sellers typically post a unique photo of the product rather than linking to the product's page to make algorithmic enforcement difficult.

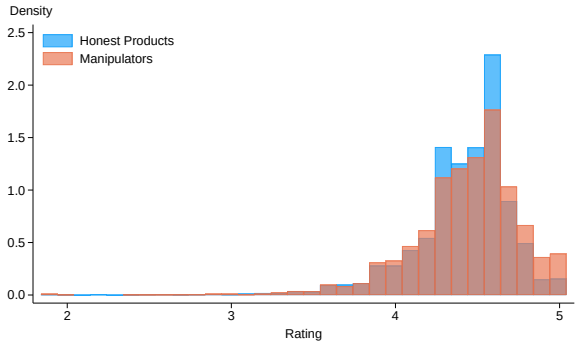
[11] Note that the sanctioned use of "incentivized reviews" differs from purchasing fake reviews in a few key ways. First, Amazon policy dictates that sanctioned reviews must clearly disclose this arrangement and must receive the same payment regardless of whether the review is positive or negative. Second, Amazon's incentivized review programs (known as Amazon Vine) does not allow sellers to choose their own incentivized reviewers in order to prevent hidden payments tied to review content.

in the keyword search results, as well as its price, number of reviews, average rating, and whether it is a sponsored result. We also use the keyword results to define close competitors for each focal product purchasing fake reviews. Specifically, we define these close competitors as the products that show up most frequently near the focal product in the search results around the time the focal product begins soliciting fake reviews. For both the focal products and this set of close competitors, we repeatedly collect the complete history of their reviews, including the text and photos used in each review. For every product review, we also collect the reviewer ID. For a subsample of these users, we are able to identify the set of other products they also reviewed from their Amazon profile.

Figure 5: Distributions of Prices and Ratings



(a) Prices of Fake Review Purchasers



(b) Average rating

Product Information Figure 5 Panel (a) shows the distribution of product prices for the set of products observed buying fake reviews, which we refer to interchangeably as the "Fake Review Purchasers" (FRPs) and ratings manipulators. Most are under $50, and the median price is approximately $24. Panel (b) shows the distribution of the products' average ratings, separately based on whether the product purchases fake reviews or an "honest product" (HP). Most products have average ratings between 4 and 5 stars, with manipulators' ratings being inflated by fake reviews. Table 1 compares fake review purchasers and their honest competitors on a broader set of characteristics. Appendix B.1 additional details on the data, including product categories.

Table 1: Characteristics of Fake Review Purchasers and Comparison Products

	Count	Mean	SD	25%	50%	75%
Displayed Rating						
Fake Review Purchasers	678	4.35	0.37	4.14	4.40	4.61
Close Competitors	3154	4.31	0.37	4.15	4.38	4.56
All Products	221923	4.25	0.61	4.01	4.37	4.62
Number of Reviews						
Fake Review Purchasers	678	239	456	43	101	239
Close Competitors	3154	1214	6088	79	260	852
All Products	222395	317	1844	9	42	179
Price						
Fake Review Purchasers	678	31.65	29.42	16.14	24.39	35.41
Close Competitors	3154	38.02	47.34	15.94	24.99	39.94
All Products	245415	43.48	190.57	12.99	20.99	38.75
Sponsored						
Fake Review Purchasers	678	0.21	0.20	0.03	0.14	0.33
Close Competitors	3154	0.32	0.23	0.13	0.29	0.47
All Products	245452	0.09	0.19	0.00	0.00	0.07
Keyword Position						
Fake Review Purchasers	678	92	50	53	87	127
Close Competitors	3154	97	53	56	92	129
All Products	244160	187	76	133	190	243
Age (months)						
Fake Review Purchasers	678	9.10	7.75	4.79	6.90	10.44
Close Competitors	3154	21.05	23.41	7.25	12.72	26.19
All Products	245936	22.47	26.15	6.00	12.91	29.61
Sales Rank						
Fake Review Purchasers	678	140726	191631	28050	81921	173623
Close Competitors	3154	115962	215764	12740	49420	134613
All Products	246051	365923	691609	51652	166437	411728

Sales Data While Amazon does not report sales precise sales quantities for each product, it does report key statistics from which sales can be inferred. The first is a measure called Best Seller Ranking or "sales rank," which ranks products within broad categories based on their recent units sold. Second, Amazon product pages detail precise product inventories unless the product has more than 1,000 units in stock.

We follow He and Hollenbeck (2020) in calculating sales quantities using these data. For most products, we are able infer quantities sold based on daily inventory changes. For

products without inventory data, we impute sales quantities using an estimated model that maps sales ranks to quantities and fits the data well in-sample. See He and Hollenbeck (2020) for additional details.

3.1 Estimating the Frequency of Fake Reviews

While we directly observe which products use fake reviews, we cannot identify with certainty which reviews are fake. Even during the period a product is observed actively buying fake reviews, some of the reviews it receives are likely organic. It is useful for our empirical exercise, however, to estimate the share of each product's reviews that are fake. To do so, we rely on the insight from He et al. (2022a) that products buying fake reviews must rely on a relatively small set of common reviewers participating in the Facebook groups. Therefore, products that share reviewers to an unusual degree are more likely to be rating manipulators.

We use this prediction algorithm from He et al. (2022a) to classify all products in the product-reviewer network as buying fake reviews or not. For a subsample of reviewers, we observe all their Amazon reviews from their Amazon profile. We label the subset of reviewers observed to leave five-star reviews for multiple fake review purchasers as "fake reviewers." Using this labeling of reviewers, we can estimate the proportion of the five-star reviews for each fake review purchaser that came from fake reviewers. For the products we observe buying fake reviews, the average estimated share of fake reviews is 47% with a median share of 50%. See Appendix B.2 for additional details on our procedure.[12]

[12]Feldman et al. (2025) also classify fake reviewers by constructing a Graphical Neural Network model of reviewers to exploit the relational patterns between fake reviewers alongside with behavioral and review content features. This method is more data intensive, and hence only provides results for roughly one third as many reviewers as we include. Nevertheless, the correlation between estimates derived from their predictions and ours is .95.

4 Empirical Model of Consumers' Beliefs

Section 2 models misinformation and mistrust in quite general terms. To make things more concrete for our empirical analysis, we precisely specify a model of how consumers interpret the ratings they observe. Section 4.1 presents a simple model in which mistrusting Bayesian consumers infer quality based on a product's number of positive and negative reviews.

This model suggests a few key components that we must either estimate or assume. The first is consumers' priors about the distribution of product quality for honest products and ratings manipulators. For these, we assume rational expectations and estimate the distributions in Section 4.2. The second is consumers' perceptions about the prevalence of ratings manipulation, which we estimate using an incentivized experiment in Section 4.3.

4.1 Consumer's Beliefs About Quality Given Ratings

In this section, we describe our model of how a Bayesian consumer forms beliefs about product quality based on observed ratings. Because the consumer is Bayesian, this entails detailing the assumptions the consumer makes about how reviews are generated.

We define a product's quality q as the probability that an organic (i.e., not fake) reviewer has a positive, five-star experience with the product. Therefore, the number of positive reviews that a given product receives out of N organic reviews is binomially distributed $B(N, q)$. Note that this model treats reviews as binary, while Amazon reviews are on a five-star scale. Most reviews, however, are either one or five stars. We therefore map Amazon ratings onto our binary framework by modeling consumers as viewing average ratings as being generated entirely by one and five star reviews.[13]

When a product manipulates its rating by purchasing fake positive reviews—which we denote using indicator F—then some of its reviews are not organic. We model this as each

[13]Formally, for a product with N reviews and an average rating of $\bar{r} \in [1, 5]$, consumers interpret the product as having N^+ positive reviews (and $N^- \equiv N - N^+$ negative reviews) such that $\frac{5N^+ + 1N^-}{N} \approx \bar{r}$.

review for manipulators (i.e., products with $F = 1$) having θ^F probability of being fake. Taking this into account, the probability of a review being positive for a given product with quality q and manipulation behavior F is:

$$p_{Fq} := \begin{cases} q & \text{if } F = 0 \\ \theta^F + (1 - \theta^F)q & \text{if } F = 1. \end{cases} \tag{2}$$

Therefore, accounting for fake reviews, the number of positive reviews a product receives out of N total reviews is binomial $B(N, p_{Fq})$:

$$P(N^+|q, N, F) = \binom{N}{N^+} p_{Fq}^{N^+} (1 - p_{Fq})^{N^-}. \tag{3}$$

Given this model, a Bayesian consumer's posterior belief about the quality of a product with N^+ positive reviews out of N reviews is a straightforward application of Bayes' rule:

$$\begin{aligned} P\left(q \mid N^+, N\right) &= \sum_F P\left(F \mid N^+, N\right) P\left(q \mid N^+, N, F\right) \\ &= \sum_F P\left(F \mid N^+, N\right) \frac{P\left(N^+ \mid q, N, F\right) P\left(q \mid N, F\right)}{\int P\left(N^+ \mid q, N, F\right) dP\left(q \mid N, F\right)}. \end{aligned} \tag{4}$$

Crucially, equation (4) suggests that a few key terms required for our empirical model. The first is $P(N^+|q, N, F)$. This is the binomial from equation (3) and notably incorporates θ^F, consumers' perceptions about the fraction of manipulators' reviews that are fake. The second is $P\left(q \mid N, F\right)$, the latent distribution of quality for fake review purchasers and honest products, which we estimate in Section 4.2. The third is $P\left(F \mid N^+, N\right)$, consumers' beliefs about the probability of ratings manipulation given N^+ positive reviews out of N reviews.

Importantly, both $P\left(F \mid N^+, N\right)$ and θ^F regard consumers' *perceptions* on the prevalence of fake reviews, which need not align with the true prevalence. Therefore, we estimate these perceptions through a survey experiment described in Section 4.3.

When characterizing demand in Section 5, we model consumers are forming expectations

about product quality based on the posterior from equation (4):

$$\mathbb{E}\left[q|N^+, N\right] := \int q dP\left(q|N^+, N\right). \tag{5}$$

4.2 Estimating the Distribution of Quality

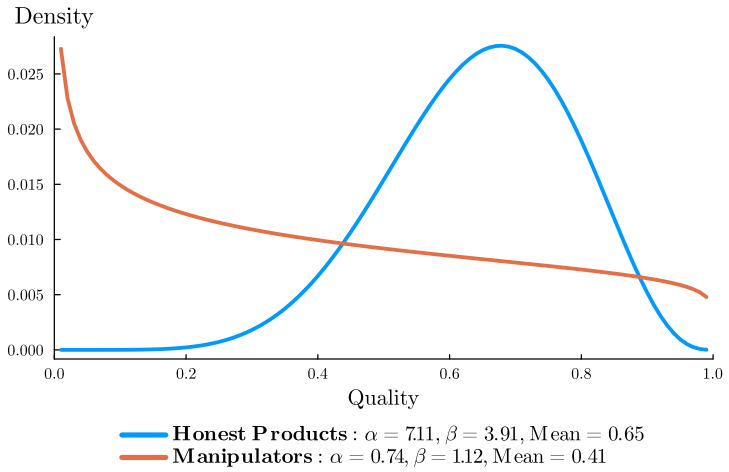
Our model of consumers' Bayesian inference about product quality above requires consumers' priors about the distribution of quality for products that do and do not purchase fake reviews. We assume that consumers have correct priors about these distributions but do not condition their prior on the number of product reviews. The former assumption is one of rational expectations and allows us to represent consumers' priors with an econometric estimate of the distributions of quality. The latter is that consumers implicitly assume $P\left(q \mid N, F\right) = P\left(q \mid F\right)$, which substantially reduces the dimensionality of the priors we must estimate. Note that this does not imply that consumers entirely ignore the number of reviews, which still plays a key role how consumers update their beliefs based on ratings in equation (4).

We estimate the distributions of quality as those that maximize the average log-likelihood of the observed organic ratings. To do this, we first leverage our inferences in Section 3.1 to identify the products that purchase fake reviews and our estimate of the number of fake reviews purchased by each manipulator. Knowing this, we can infer the number of positive organic reviews—i.e., the number of positive reviews after excluding fake reviews—which we denote by N^{o+}. Likewise, we denote the number of organic reviews as $N^o := N^{o+} + N^-$.

We denote by $P(q|F;\gamma)$ the parameterization of $P(q|F)$ by γ. We let q be Beta distributed conditional on F.[14] In other words, $\gamma = \{(\alpha_F, \beta_F)\}_F$ and $q|F \sim \text{Beta}(\alpha_F, \beta_F)$. Using this,

[14]Note that the Beta distribution is conveniently the conjugate prior of the Binomial distribution, which aids in computing consumers' posterior expected quality (equations 4 and 5).

Figure 6: Estimated Priors



the likelihood of N^{o+} organic positive ratings out of N^o organic ratings is:

$$LL\left(N^{o+}, N^o; \gamma\right) := \log\left(\int \binom{N^o}{N^{o+}} q^{N^{o+}} (1-q)^{N^o - N^{o+}} dP(q|F;\gamma)\right) \tag{6}$$

We estimate γ as the maximizer of the average log-likelihood of organic reviews in the data. Figure 6 presents the estimates of γ and implied distributions $P(q|F;\hat{\gamma})$. These indicate that products purchasing fake reviews tend to be of substantially lower quality: the average quality of manipulators is just 0.41, while the average quality of honest products is 0.65. Importantly, this suggests that manipulators on Amazon are not high-quality products trying to overcome a cold-start problem. While some theoretical models of fake review behavior predict an increasing relationship between product quality and likelihood of using fake reviews (Dellarocas, 2006; Yasui, 2020), we instead find that fake reviews on Amazon are probably best characterized as enabling low-quality products to masquerade as high-quality ones.

4.3 Survey Experiment to Measure Beliefs

The are are two key components in our model of beliefs in Section 4.1 that represent consumers' *perceptions* about ratings manipulation are therefore not directly observable in

market data. The first is $P(F|N^+, N)$, the perceived probability that a product with a given rating purchases fake reviews. The second is θ^F, the perceived fraction of manipulators' reviews that are fake.

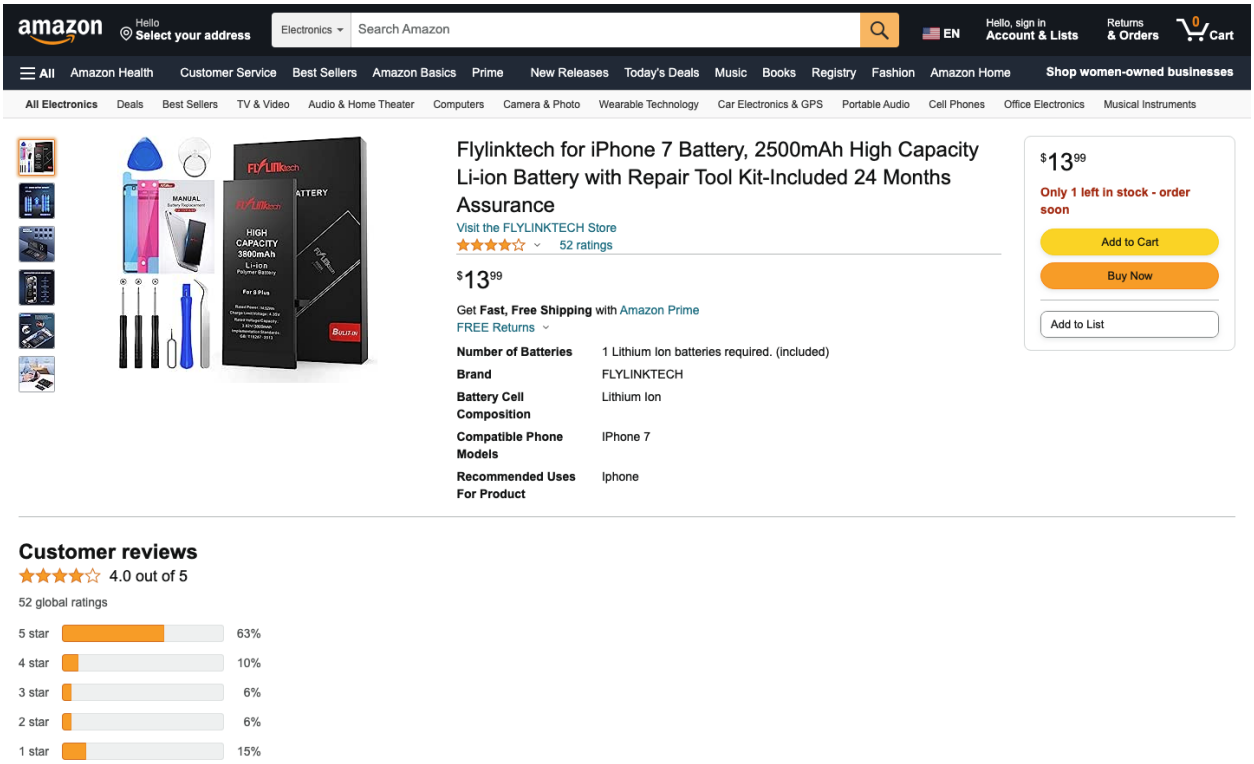
In this section, we describe an incentivized survey experiment that we use to characterize consumers' perceptions. The principal survey task that we describe in detail below aims to elicit consumers' perceptions of the prevalence of fake reviews and determine how beliefs vary with product characteristics. We use the fact that we observe the ground truth about ratings manipulation to incentivize respondents by paying them more for selecting responses that better align with the truth. The survey implementation clearly communicates these payoffs to participants. We also leverage randomization to assess how participants' responses vary with different product observables.

The main survey task takes place after the participant has completed a reading comprehension check, answered a series of demographic questions, indicated whether they shop on Amazon, and identified which 5 of Amazon's 19 primary product categories they most frequently shop for online. We implement a host of best-practices to screen for bots and poor engagement, including an initial reading comprehension check, a mid-survey attention check, and an additional comprehension check during the main component of the survey. Finally, in addition to the experiment, we also ask participants directly about the prevalence of fake reviews: "Out of 100 randomly chosen products on Amazon.com, how many would you expect to have purchased fake reviews?"

Main Prediction Task The principal survey task asks participants to view an Amazon product page and provide their best guess of the probability that that product uses fake reviews. The pages participants view are constructed using the HTML code from the pages of products in our sample. As illustrated in Panel (a) of Figure 7, each product page displays the product name, image, price, average star rating, number of reviews, and other details.

We include a slider under the product page that asks "Using the slider below, please select the percentage probability on a scale of 0 to 100 that the product purchases or has

Figure 7: Example Survey Page and Slider



(a) Example product page shown as in survey

(b) Slider showing respondents' beliefs and payoffs.

purchased fake reviews." The payments are straightforward: a respondent that selects $x\%$ probability receives x cents if the product purchased fake reviews and $100 - x$ cents if the product did not purchase fake reviews. When participants move the slider, it automatically updates to provide a full description of their conditional payouts, as illustrated in Panel (b). This ensures that respondents are fully informed about their payoffs and, in particular, the fact that placing greater probability on the truth earns a larger payment. To ensure that respondents understand the slider's mechanics, they must demonstrate its use prior to the prediction exercises as well as in the middle of the exercises as an attention check.

Respondents repeat the prediction task for 10 different products: two products drawn randomly from each of the five Amazon categories the participant indicated they are most likely to purchase online. Participants can therefore earn a maximum of $10 in addition to a base payment of $1 for participation. Within each Amazon category, the candidate products that a participant might see consists of two randomly selected manipulators and the closest competitor for each. For each question, approximately one third (32%) of respondents saw a ratings manipulator, and approximately two thirds (68%) saw an honest competitor. For some respondents, we randomly replace one prediction task with the product page for an Amazon gift card to verify that respondents are generally attentive enough to indicate a low probability that Amazon purchased fake reviews.

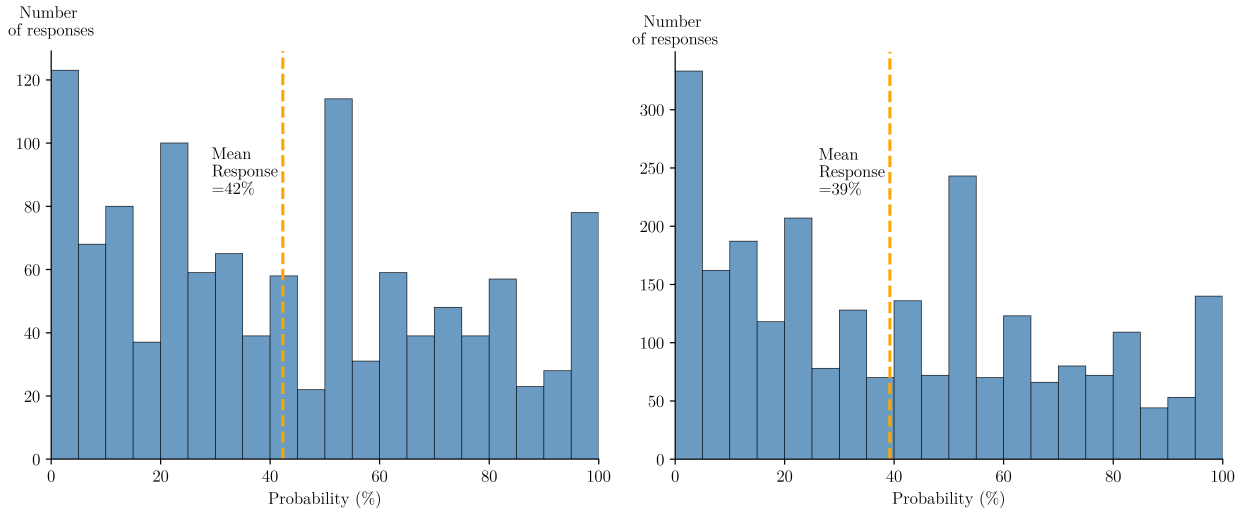
In addition to randomizing products, we also modify the HTML code of product pages to independently randomize the number of reviews and average rating that each participant sees for each product. This randomization allows us to identify how ratings affect consumers' perceptions about the likelihood a product purchasing fake reviews—i.e., $P(F|N^+, N)$—using experimental variation that is decoupled from the product itself. In implementation, we modify not only the number of reviews and the average rating, but also the histogram of ratings. We show most participants a histogram matching the distribution of reviews for all products with that rating and number of reviews. A 40% subsample of respondents are instead shown either highly unimodal or bimodal distributions (5th or 95th percentiles of variance) to determine the importance of histogram shape. See Appendix B.3.6 for details.

Finally, for a randomly selected product, we ask respondents a follow-up question intended to elicit beliefs about θ^F: "For this question, please assume that this product has purchased fake reviews. Guess the fraction of fake reviews among all its reviews." To incentivize respondents, we pay them based on how close they get to our measure of θ^F. As with the other prediction exercises, respondents respond using a slider that automatically updates and displays their payoffs under each state.

4.3.1 Survey Results

We ran the online survey experiment on Prolific in July 2023. Our sample includes 401 qualified respondents who report shopping on Amazon and passed all comprehension and attention checks. Figure B.2 summarizes the demographics of the qualified participants.

Figure 8: Perceptions of Fake Review Purchasing by Actual Behavior



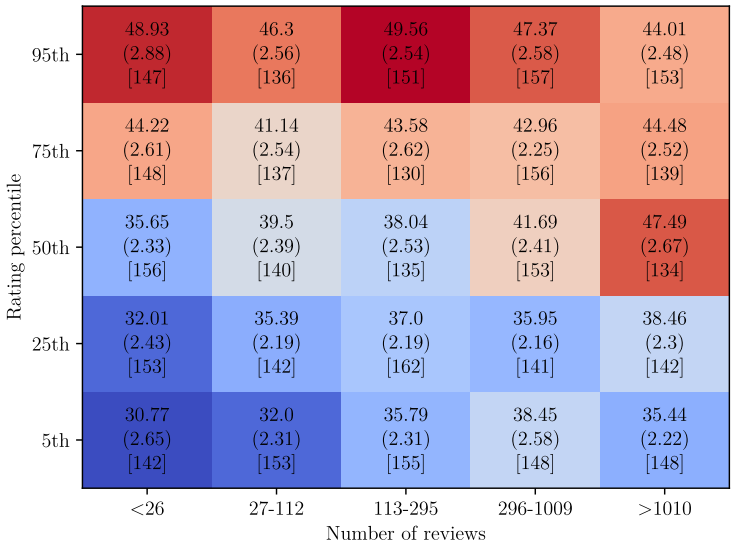
(a) Fake Review Purchasers

(b) Honest Products

When asking directly about the percent of products purchasing fake reviews, the mean response is 31% and the median is 26%. This is slightly higher than the 19% of products we observe in our data. For the prediction task, beliefs about fake review prevalence are somewhat higher. Figure 8 shows the distribution of responses for products that did and did not purchase fake reviews. In instances where the respondent is shown a fake review purchaser (Panel a), the mean response is 42% and the median is 40%. In cases where the product shown does not use fake reviews (Panel b), the mean is 39% and the median is 36%. That these probabilities are so similar suggests that the characteristics of a product's page do little to help consumers discern the truth about whether it purchases fake reviews.

Figure 9 depicts how these probabilities vary with average rating and number of reviews. Consumers appear generally suspicious of products with high ratings and perceive products

Figure 9: Beliefs by Rating and Number of Reviews



Notes: Figure depicts respondents' average perceived probability that a product with a given rating and number of reviews purchased fake reviews. We also present the standard error of the mean in parentheses and the number of responses in brackets. Figure B.3 shows the relationship separately for ratings and number of reviews. Figure B.4 presents the relationships in our real-world sample.

with low ratings as being comparatively unlikely to have purchased fake reviews. Surprisingly, the number of reviews has little effect on responses. This represents an important consumer misperception, as our real-world data indicate that products with many reviews are empirically much less likely to purchase fake reviews (Figure B.4).

The values in Figure 9 represent $P(F|N^+, N)$, a key prior required for our model of consumer beliefs in Section 4.1. Note that we do not condition on features other than N^+ and N, as Figure 8 indicates that consumers are not able to productively use other product characteristics to identify fake review purchasers.

The final unknown governing consumers' beliefs in Section 4.1 is θ^F. Figure B.5 shows participants' responses when asked to estimate θ^F. The mean response is 38% (median: 31%), so we model consumers' perceptions using $\theta^F = 38\%$. Note that this is lower than the 47% measured empirically in Section 3.1.

For additional analyses of responses by product category, differing histogram shapes, and for Amazon gift cards, see Appendices B.3.5, B.3.6, and B.3.7.

4.3.2 Supplemental Survey With Review Text

Our primary survey did not include the text of reviews both for simplicity of implementation and in order to emphasize ratings and the number of reviews. To assess whether the content of reviews might aid in consumers' ability to identify fake reviews, we ran a supplemental survey in April 2024 on a different set of 100 Prolific participants that included the option for participants expand and view a sample of reviews for each product (Figure B.12).[15]

Survey participants clicked to expand and view the reviews 86% of the time, indicating that they believed review text would be informative. The click-through rates were similar when viewing fake review purchasers and honest products. Figure B.13 shows the distribution of responses after viewing the text of reviews. Viewing review text does not appear to improve participants' ability to distinguish fake review purchasers: the mean prediction was 35.2% if participants saw the reviews for a fake review purchaser and 34.8% if they saw the reviews for an honest product.[16]

5 Consumer Demand

In this section, we specify a model of consumer demand as a function of ratings, prices, and other product attributes. Section 5.1 characterizes consumers' indirect utility, Section 5.2 details our estimation procedure, and Section 5.3 presents our estimates.

5.1 Consumer Indirect Utility

We model demand using the standard discrete choice random utility framework following Berry et al. (1995). Consumer i makes a purchase decision about product j at time t based

[15] For fake review purchasers, we show the first 10 reviews received after date the product began purchasing fake reviews, which were between December 2019 and June 2020. For the honest products, we select the earliest 10 reviews among our data, which were scraped between August and December 2020.

[16] Interestingly, if the participant did not click to see reviews, they did slightly better at distinguishing: 48.5% for fake review purchasers and 39.8% for honest products. This may reflect randomness or could indicate that a small number of sophisticate respondents are able to spot features outside of the review text that serve as a weak signals of fake review purchasing activity.

on their indirect utility function:

$$u_{ijt} = \beta_i \mathbb{E}\left[q_{jt}|N_{jt}^{+}, N_{jt}\right] - \alpha_i p_{jt} + \beta^X X_{jt} + \lambda_t + \zeta_j + \xi_{jt} + \epsilon_{ijt} \tag{7}$$

where $\mathbb{E}\left[q_{jt}|N_{jt}^{+}, N_{jt}\right]$ is the consumer's expectation about quality given its star rating and number of reviews. Section 4 describes our model of how consumers form beliefs about quality based on ratings, as well the procedure we use to estimate their priors. Price p_{jt}, product age (cumulative time listed on Amazon), and position in search results also enter into indirect utility. We also include time fixed-effects, λ_t, to capture general seasonality in demand, and product fixed effects, ζ_j, that capture unobserved product characteristics. Since the typical product we examine is only about $25, we assume that consumers are not forward-looking or strategic in the timing of their purchases. To allow for heterogeneity in individuals' preferences, we model consumer utility over price and expected quality as $\binom{\alpha_i}{\beta_i} \sim \log \mathcal{N}(\mu, \Sigma)$. The use of a lognormal distribution restricts preferences such that all consumers place positive weight on expected quality and negative weight on price. To allow for flexible patterns of substitution to the outside good, we also include a mean-zero, normally distributed random coefficient in the utility of the outside good. The error term ϵ_{ijt} is assumed to be Type-I extreme value distributed.

We define markets at the keyword-week level and denote the set of products in the market as $\mathcal{J}$. To construct this set of competitors, we use our data from several months of scraping keyword search results and calculate the frequency with which products co-occur on the same page of search results. Then, for each focal product that purchases fake reviews, we choose the set of up to ten products that co-occur most frequently. We define market size by taking the moving average of total weekly sales for the products in $\mathcal{J}$ at the monthly level and multiplying by a constant.

5.2 Estimation and Identification

We estimate demand using weekly data on market shares, ratings, number of reviews, and prices for all products in consumers' consideration sets. To estimate demand parameters $\theta = (\beta^X, \mu, \Sigma)$, we use a GMM estimator that interacts the structural demand side error with a set of instruments Z, where the demand parameters. We also follow MacKay and Miller (2024) in implementing a covariance restriction between the demand-side error and the supply-side error.

For all specifications, we employ the second-stage heteroskedasticity robust optimal weighting matrix and the Chamberlain (1987) approximation to the optimal instruments as described in Conlon and Gortmaker (2020). In order to obtain a first-stage estimate to construct the weighting matrix and approximation to the optimal instruments, we need to choose initial instruments. For the supply specification we use the intercept, product age, and sponsorship status. For demand, we follow a standard approach and use Gandhi & Houde-style instruments constructed from the product characteristics of competing products. We rely on product fixed effects to absorb mean product quality. Thus, we treat variation in ratings over time as largely exogenous.

5.3 Results of Demand Estimation

Table 2 shows the results from demand estimation. We find significant coefficients in the expected direction for price and quality, as well as significant heterogeneity in price sensitivity. However, heterogeneity in preferences for quality are estimated as being close to zero and not significant. The results imply the elasticity of demand with respect to expected product quality is fairly high at roughly 1.6. This is not directly comparable to previous estimates since this elasticity is to the posterior expectation of quality rather than the rating itself. We find a mean price elasticity of -3 with a median of -2.3. We also find a significant negative coefficient on the listing rank, which is consistent with greater demand for better-positioned products. The random coefficient on the intercept term is significant, indicating significant variance in the preference for the outside good.

Table 2: Demand Estimates

Age	0.0037
	(0.023)
Listing Rank	-0.047
	(0.0013)
σ_1	0.38
	(0.041)
μ_{-p}	-2.3
	(0.099)
σ_{-p}	0.12
	(0.010)
μ_q	0.78
	(0.041)
σ_q	0.038
	(0.012)

Product FEs	Yes
Week FEs	Yes
Gandhi-Houde IVs	Yes
Median Own-Price Elast.	-2.3
Mean Own-Price Elast.	-3
Median Own-Quality Elast.	1.6
Mean Own-Quality Elast.	1.6
Observations	83,530

Notes: The random coefficients on price and expected quality are parameterized as $\binom{\alpha_i}{\beta_i} \sim \log\mathcal{N}(\binom{\mu_{-p}}{\mu_q}, \begin{pmatrix}\sigma_{-p} & 0\\ 0 & \sigma_q\end{pmatrix})$. The standard deviation of the random coefficient on the intercept term is denoted as σ_1.

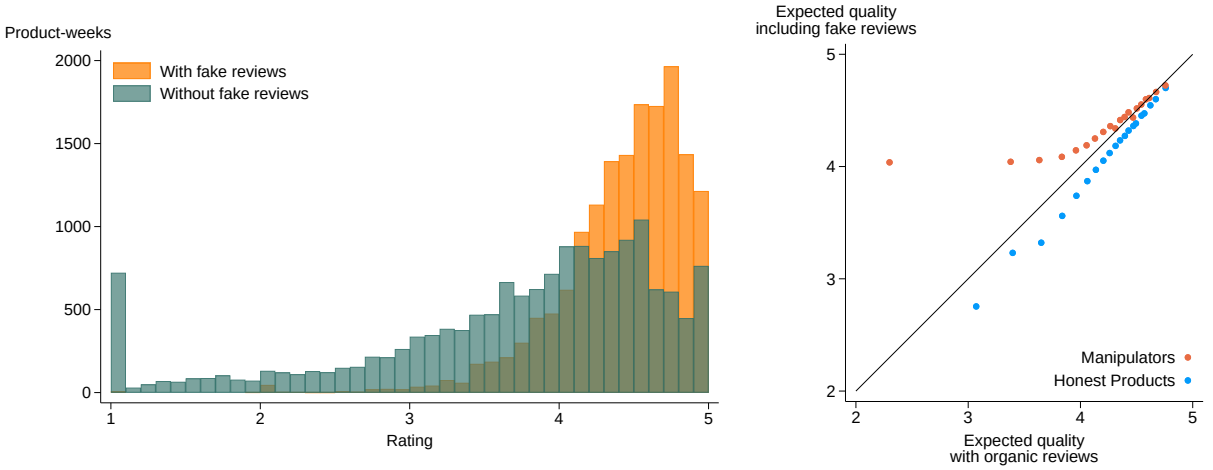
6 Counterfactuals

To understand the effects of rating manipulation on the Amazon marketplace, we simulate a series of counterfactuals in which the platform eliminates fake reviews. In order to present our results as the effect of fake reviews—as opposed to the effect of deleting fake reviews—our comparisons treat the counterfactual world without fake reviews as the baseline. Our first counterfactual examines the full equilibrium effects when removing fake reviews eliminates both misinformation and mistrust. We then isolate the misinformation and mistrust channels by simulating separate counterfactuals in which (i) fake reviews are eliminated but consumers remain mistrustful and (ii) fake reviews persist but consumers fully trust reviews. In each case, we highlight the role of competition by comparing outcomes when prices are held fixed to those when firms react by changing prices. Finally, we examine alternative policies to fully deleting fake reviews and explore Amazon sellers' incentives to purchase fake reviews.

6.1 The Equilibrium Effect of Fake Reviews

To understand the equilibrium effect of fake reviews, we contrast the factual Amazon marketplace—where fake reviews are prevalent and consumers mistrust ratings—with a simulated counterfactual without fake reviews or the attendant misinformation and mistrust. Figure 10 uses our Section 3.1 estimates of fake reviews for each product in our data to recompute product ratings and consumer beliefs in the absence of fake reviews. Panel (a) shows that fake reviews dramatically inflate manipulators' ratings by an average of 0.7 stars. Panel (b) depicts how the presence of fake reviews changes consumers' perceptions of products' qualities (Equations 4 and 5): consumers overestimate the quality of manipulators—especially those with poor organic reviews—and underestimate the quality of honest products. Figure C.1 separates the effects of misinformation (Panel a)—which inflates manipulators' ratings and perceived quality—and mistrust (Panel b)—which deflates perceived quality for all products.[17] Importantly, the net effect shown in Figure 10 Panel (b) advantages manipulators, improving their perceived quality relative to honest competitors.

Figure 10: Effect of Removing Fake Reviews on Product Ratings and Perceived Qualities



(a) Ratings of Fake Review Purchasers

(b) Perceived Qualities

[17]Note that under our assumed binomial model of fake reviews, the mistrustful Bayesian infers that a small number of products are so poorly rated they are unlikely to have purchased fake reviews. For these products, mistrust actually increases perceived quality since honest products have much higher average quality. In the interest of clarity, we disregard these small number of cases in our exposition.

These changes in perceived quality affect consumers' demand according to the model estimated in Section 5. In turn, firms adjust their prices to reach a Bertrand Nash equilibrium. In a fully static market, simulating counterfactual demand would simply entail modifying $\mathbb{E}\left[q_{jt}|N_{jt}^{+}, N_{jt}\right]$ in equation (7) with updated beliefs. However, eliminating fake reviews may have important dynamic effects because equilibrium outcomes in each period affect both organic reviews and search positions in future periods. We summarize how we incorporate these dynamics below and provide further details in Sections C.1, C.2, and C.3.

Organic Reviews. Organic reviews are unpaid reviews left voluntarily by a small fraction of purchasers. Since each purchaser has the potential to leave an organic review, the number of new organic reviews should scale with a product's recent sales. To capture this, we model the number of new organic reviews that each product receives as being Poisson with a mean that scales with sales in each of the previous two weeks. We estimate this relationship via Poisson regression and find that new organic reviews scale particularly strongly with the prior week's sales (elasticity of 0.73).

Search Position. We model search position as being determined stochastically by a rank-ordered logit (Beggs et al., 1981). Our estimates indicate that Amazon's search algorithm preferences products with recent sales, recent positive reviews, that have paid for a sponsored position, and that have been on Amazon longer.

Simulating Dynamics. We simulate entire outcome *paths* in order to incorporate path-dependence in organic reviews, search position, and sales. To construct a path, we simulate the dynamic evolution of the market period-by-period. Each period, we draw a realization of both the new organic reviews and search ranking according to the models above based on the simulated equilibria in the previous two periods along the simulated path. These determine demand curves in the current period. This approach allows early changes in counterfactual demand to compound or attenuate according to the dynamics induced by the

path-dependence of organic reviews and search position.[18]

Figure 11 depicts our principal findings on the effect of fake reviews. Note that we present our results as the change in marketplace outcomes moving from the counterfactual without fake reviews to the factual with fake reviews that result in both misinformation and mistrust.

Panel (a) shows that, in equilibrium, fake reviews tend to shift market share toward ratings manipulators and away from honest products. On average, manipulators sell 27.2% more units while honest products sell 4.4% fewer. Since honest products outnumber manipulators and tend to have higher baseline sales, fake reviews result in a reduction of 0.9% in units sold on the marketplace. Panel (b) indicates that the increase in demand allows manipulators to profitably increase their prices (average increase of \$0.31), while honest products must lower their prices to compete with manipulators (average decrease of \$0.07). These shifts in prices and quantities translate to substantially greater revenues and profits for manipulators at the expense of honest products (Panels c and d). Fake reviews increase the total profitability of manipulators by a dramatic 29.9% while reducing the profits of honest products by 5%.

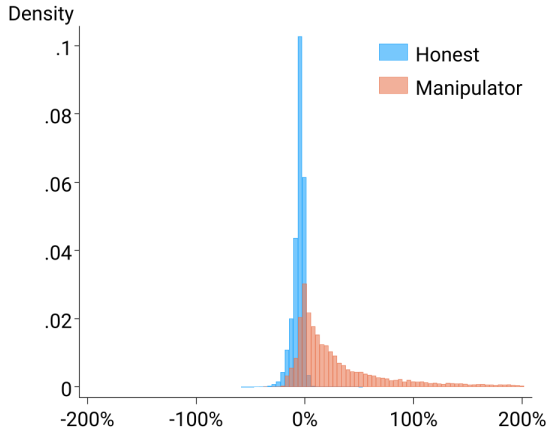
The total revenue across all products in our sample decreases by 1.3%. This likely harms Amazon, which typically takes a fixed percentage of revenue from each sale as a "referral fee." Still, if identifying fake reviews or taking enforcement action is costly, Amazon may prefer to simply allow the level of misinformation and mistrust in our sample. Moreover, we show in Section 6.2 that Amazon benefits from improving trust but actually prefers misinformation. Enforcement that is credible and improves trust may be particularly difficult or costly, which would further weaken Amazon's incentives to combat fake reviews.

Fake reviews decrease total consumer welfare by 0.77%, with substantial variation in the impact across individuals (Panel e).[19] Welfare changes have two principal drivers that we explore further in Section 6.2. The first are purchasing mistakes: 3.18% of consumers are induced to purchase the manipulator's product due to misinformation or dissuaded from

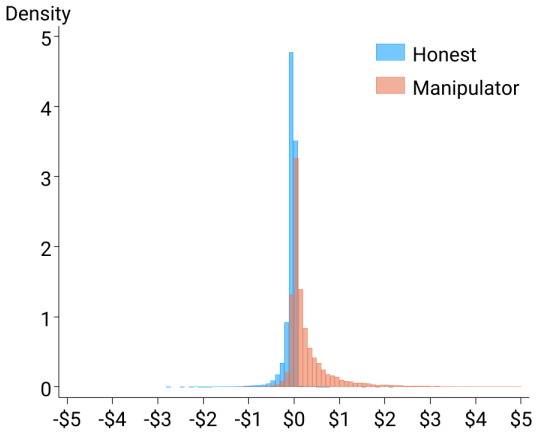
[18]Note that we model firms as ignoring dynamic incentives: even as demand evolves dynamically, we model firms as setting prices to reach a static Nash Bertrand each period.

[19]See Appendix C.4 for details on how we compute consumer welfare under misinformation and mistrust.

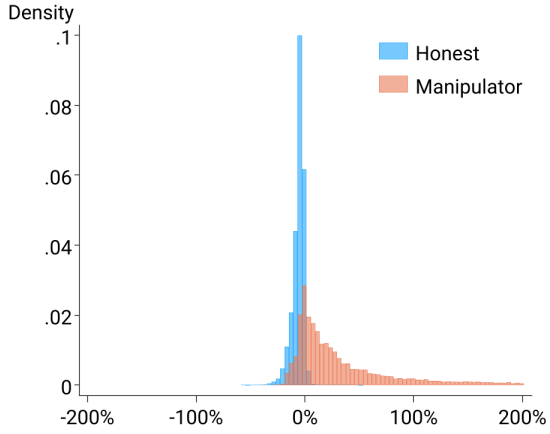
Figure 11: Equilibrium Effect of Fake Reviews



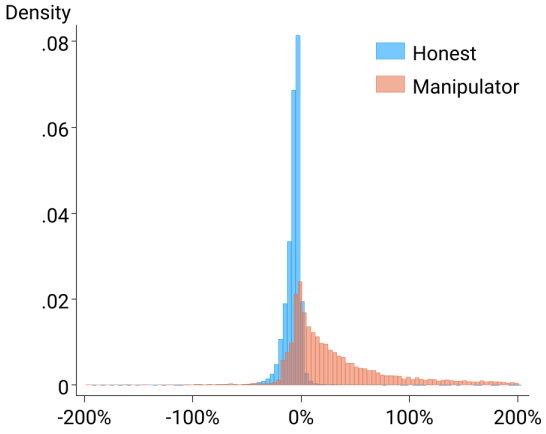
(a) Change in Quantities



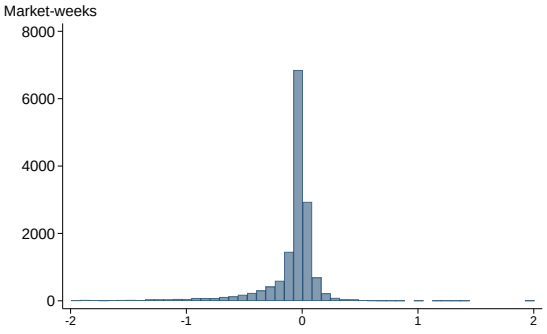
(b) Change in Prices



(c) Change in Revenues



(d) Change in Profits



(e) Change in Welfare

Notes: Figure presents changes in equilibrium outcomes attributable to the level of misinformation and mistrust due to fake reviews in the factual world. Fewer than 4% of observations were beyond the axis limits and were trimmed. Table C.3 summarizes the changes.

purchasing a preferable product due to mistrust (Figure C.2).[20] The second are price changes: consumers who still purchase honest products tend to be better off due to discounted prices, while those who purchase manipulators tend to be worse off because of price increases.

6.2 Isolating the Effects of Misinformation and Mistrust

Table 3: Counterfactuals Varying Misinformation and Mistrust

	No FR	Misinfo		Mistrust		Misinfo+Mistrust	
		Fixed prices	Floating prices	Fixed prices	Floating prices	Fixed prices	Floating prices
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Manipulators							
Units Sold	1,241,188	1,667,653	1,619,013	1,174,963	1,181,126	1,622,939	1,579,209
Change (%)	-	+34.36	+30.44	-5.34	-4.84	+30.76	+27.23
Average Price ($)	31.32	31.32	31.67	31.32	31.27	31.32	31.63
Change (%)	-	-	+1.12	-	-0.16	-	+0.99
Revenues ($)	38,579,454	50,851,284	50,225,380	36,466,417	36,564,405	49,560,286	48,972,656
Change (%)	-	+31.81	+30.19	-5.48	-5.22	+28.46	+26.94
Profits ($)	15,211,127	20,113,398	20,318,445	14,401,311	14,402,343	19,574,512	19,761,716
Change (%)	-	+32.23	+33.58	-5.32	-5.32	+28.69	+29.92
Honest Products							
Units Sold	9,760,784	9,558,423	9,596,630	9,478,091	9,495,613	9,271,686	9,329,038
Change (%)	-	-2.07	-1.68	-2.90	-2.72	-5.01	-4.42
Average Price ($)	37.95	37.95	37.91	37.95	37.92	37.95	37.88
Change (%)	-	-	-0.11	-	-0.08	-	-0.19
Revenues ($)	359,226,738	352,003,143	353,003,741	349,937,739	349,837,157	342,334,268	343,535,154
Change (%)	-	-2.01	-1.73	-2.59	-2.61	-4.70	-4.37
Profits ($)	122,692,205	120,126,778	120,283,241	119,189,010	119,084,844	116,586,494	116,590,465
Change (%)	-	-2.09	-1.96	-2.86	-2.94	-4.98	-4.97
Platform							
Revenue ($)	39,780,619	40,285,443	40,322,912	38,640,416	38,640,156	39,189,455	39,250,781
Change (%)	-	+1.27	+1.36	-2.87	-2.87	-1.49	-1.33
Consumers							
Welfare ($)	192,998,456	191,651,163	191,584,960	192,228,282	192,815,989	191,028,358	191,514,061
Change (%)	-	-0.70	-0.73	-0.40	-0.09	-1.02	-0.77
Sophisticates							
Welfare Change (%)	-	+0.07	+0.04	-0.02	+0.28	+0.05	+0.31

To better understand how fake reviews impact the marketplace, we isolate the effects of misinformation and mistrust. To isolate the effect of misinformation, we simulate a counterfactual in which fake reviews exist but consumers interpret ratings as if all reviews were

[20]Note that the percentage of mistakes appears small because the average market share of manipulators is just 9.9% since only 19.4% of our sample are manipulators and the outside good share averages 51.3%.

organic. That is, we simulate consumers as trusting reviews in spite of fake reviews being prevalent. To isolate the effect of mistrust, we simulate a counterfactual without fake reviews but in which consumers still mistrust ratings to the extent estimated from our survey experiment. That is, we simulate consumers as still mistrusting reviews in spite of their absence. In each case, we first compute the results holding prices fixed and then allowing firms to adjust prices in order to isolate the competitive responses to each mechanism.

Table 3 presents the principal results of our counterfactual simulations. As before, we contrast each scenario against a baseline without either misinformation or mistrust (column 1). Correspondingly, contrasting the last column with the first column summarizes the results from Figure 11. Additionally, to better understand how misinformation and mistrust affect consumers, we also detail the frequency and harms of purchasing mistakes under each counterfactual in Figures C.2 and C.3.

Only Misinformation. Misinformation misleads consumers into substantially overestimating manipulators' quality, especially for low-quality products masquerading as highly-rated products (Figure C.1a). Columns (2) and (3) of Table 3 present the counterfactuals in which consumers face misinformation but still fully trust ratings. Absent price adjustments, manipulators' inflated ratings mislead consumers into purchasing 34.4% more of their products, which increases manipulators' profits by 32.2%. When prices adjust, manipulators raise theirs by an average of \$0.35, while honest products reduce theirs by an average of \$0.04. These price changes further harm purchasers of manipulators and slightly benefit purchasers of honest products. The net impact, however, is negligible. The majority of harms occur to the 1.81% of consumers that are misled into making the wrong purchase (Figure C.2), mistakes that average \$3.70 in harm (Figure C.3). Welfare across all consumers that consider purchasing on the platform decreases by a small but meaningful 0.73%. Finally, honest products are also harmed, as they sell 1.7% fewer units and earn 2% less profit.

Only Mistrust. Absent misinformation, mistrust causes consumers to slightly underestimate quality for all products and especially those with mediocre ratings (Figure C.1b).

This wariness leads 1.9% of consumers to mistakenly purchase the wrong product or not purchase at all (Figure C.2). These mistakes have an average harm of $1.73 (Figure C.3), which is more modest than for mistakes driven by misinformation. This is because slight changes in perception due to mistrust tend to affect purchasing decisions primarily when a consumer is almost indifferent, whereas a heavily inflated rating due to misinformation can cause consumers to mistakenly purchase a substantially inferior product.

When prices are fixed, mistrust results in slightly fewer sales for both manipulators and honest products, as well as a 0.4% reduction in consumer welfare. When prices adjust, both manipulators and honest products lower their prices on average. Intuitively, products must compete more aggressively on price since mistrust makes it more difficult to differentiate through ratings. Unlike misinformation, mistrust induces sufficient additional price competition to largely mitigate its consumer harms, which shrink to just 0.1%.

Sophisticated Consumer. The benefits of competitive responses are clearest when considering a single consumer that is "sophisticated" in that they purchase products only based on organic reviews.[21] This approximates what one might expect from a consumer who uses software to hide fake reviews or otherwise relies on a different trustworthy signal of quality.

Although fake reviews do not directly affect the sophisticate's perception of quality, they affect her welfare indirectly by changing equilibrium prices. In isolation, misinformation increases her welfare by 0.04% due primarily to discounting of honest products, while mistrust alone increases her welfare by 0.29% due to discounting of both honest products and manipulators. Under misinformation and mistrust together, her welfare increases by 0.31%.

Dynamic effects (Section C.1) offer a second channel through which fake reviews affect the sophisticate. Most notably, manipulation affects the arrival rate of organic reviews. For example, misinformation increases manipulators' sales and therefore their organic reviews, providing valuable information to the sophisticate about manipulators' true quality. Therefore, misinformation increases the sophisticate's welfare by 0.07% even when prices are fixed.

[21] We assume that she fully trusts organic reviews and has a rational expectations prior on product quality.

Platform Incentives These counterfactuals also shed light on the incentives of the platform, which receives a fixed share of revenue.[22] Table 3 shows that the platform benefits from misinformation—which causes consumers to overestimate quality and spend more on the platform—but is harmed by mistrust— which causes consumers to underestimate quality and spend less on the platform.

Indeed, if it were possible, the platform would most prefer for fake reviews to exist but for consumers to be entirely trusting.[23] In contrast, the platform is particularly harmed if consumers mistrust ratings when no fake reviews exist. This creates a key challenge in relying on the platform to address ratings manipulation: the benefits to the platform derive from reducing consumers' perceptions of manipulation and not from the actual removal of misinformation. Therefore, inconspicuously identifying and removing fake reviews is the most harmful policy to Amazon in the short-run. Instead, Amazon should principally aim to inspire confidence that manipulation is rare—such as by conspicuously advertising strict anti-manipulation policies and sophisticated enforcement technology—while only actually removing or preventing fake reviews to the extent required for credibility.

6.3 Alternative Policies

Our principal counterfactual considers fully eliminating fake reviews. However, Amazon and regulators might also consider less extreme alternatives.[24] We consider two in this section. The first is to selectively delete only a fraction of fake reviews. The second is to dilute the influence of fake reviews by increasing the number of organic reviews.

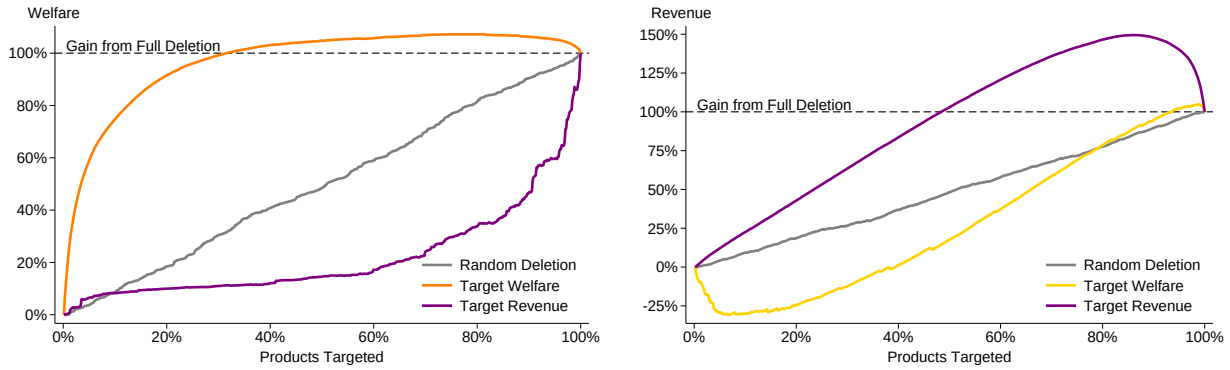
Partial Deletion Deleting fake reviews may be costly to Amazon for a number of reasons. For example, even an extremely accurate process for identifying fake reviews would have a

[22]Platform incentives are also shaped by the related issue of sponsored advertising. In theoretical work, Long and Liu (2024) show that under certain conditions, a platform may tolerate fake reviews in order to intensify advertising competition and increase advertising revenue.

[23]This is likely infeasible in the long run as consumers learn about the prevalence of fake reviews or experience systematic discord between product ratings and quality.

[24]They may also consider more extreme alternatives. For example, we examine de-listing of manipulators in Section C.5.8 and find that consumers are substantially harmed by the loss in product variety.

Figure 12: Gains from Partial Deletion Relative to Full Deletion



(a) Consumer Welfare

(b) Platform Profits

non-zero false-positive rate. Widespread deletions would therefore likely entail adjudicating a large number of instances in which a seller disputes that a deleted review was fake.

Figure 12 examines the extent to which Amazon could improve consumer welfare or platform revenue while deleting fake reviews for only a fraction of manipulators. We present three alternative assumptions about which fake reviews Amazon deletes. In the first, Amazon identifies manipulators at random. In the second and third, Amazon applies a greedy algorithm to prioritize deleting the fake reviews on products where the manipulation most harms consumer welfare and platform revenues, respectively. In each simulation, deletions proportionately reduce both misinformation and mistrust. See Appendix C.5.4 for details.

We find that while random deletion increases both welfare and profits roughly linearly, selective deletion achieves gains considerably more quickly: half of the benefits of to consumers and the platform can be achieved by targeting just 4% and 24% of manipulators, respectively.[25] Importantly, however, Figure 12 also highlights a crucial misalignment between consumers and the platform: the fake reviews most harming consumers and those most harming the platform are not the same. Fake reviews for low-quality products in large markets with high ratings-elasticities cause the most harm to consumers. However, delet-

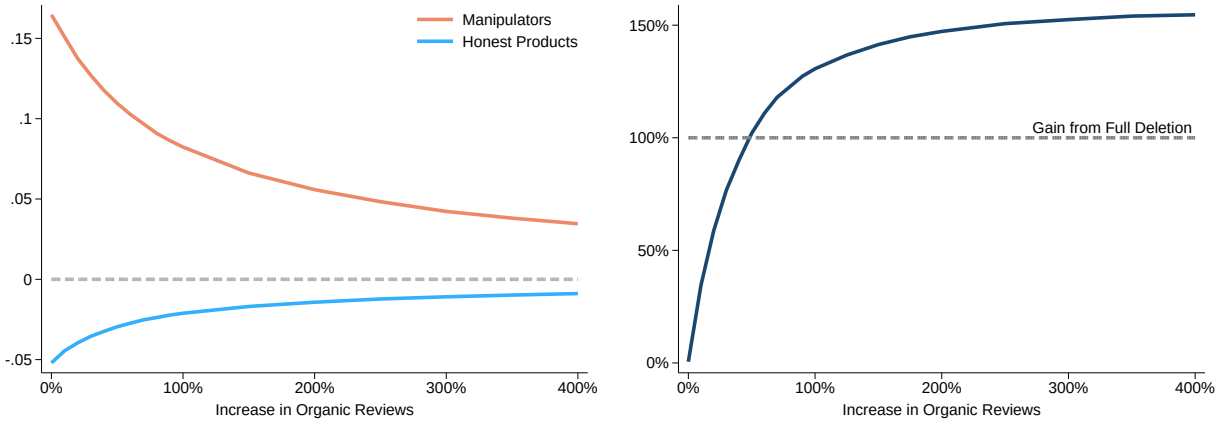
[25]Note that the 4% of manipulators represent 14% of fake reviews, since increasing consumer welfare typically entails targeting products with many fake reviews that create harmful misinformation. In contrast, the 24% of manipulators represent just 6% of fake reviews, since increasing platform revenue entails targeting products with few fake reviews to improve trust without substantially reducing misinformation.

ing these reviews actually harms the platform since they frequently generate substantial additional revenue.

In contrast, prioritizing platform revenue generally entails deleting fake reviews to improve trust without jeopardizing the revenue generated by misinformation. This approach targets fake reviews that lead to few purchasing mistakes and have little harm to consumers. Indeed, the greedy algorithm prioritizing platform revenues removes fake reviews from 64% of manipulators before it can achieve just 20% of the potential gain in consumer welfare. Importantly, this should caution against relying on platform-directed enforcement to benefit consumers, even when enforcement appears substantial.

Increasing the Number of Organic Reviews Even without identifying and deleting fake reviews, Amazon may be able to mitigate their impact by increasing the number of organic reviews on the platform. They might achieve this by offering compensation for honest reviews either directly from the platform (Fradkin and Holtz, 2023) or from the seller (Li et al., 2020). To study the effect of such policies, we simulate counterfactuals in which the number and arrival rate of organic reviews are scaled up. See Appendix C.5.5 for details.

Figure 13: Increasing the Number of Organic Reviews



(a) Mean Error in Perceived Quality

(b) Welfare Gains Relative to Elimination

Figure 13 presents our findings. Panel (a) shows that increasing organic reviews increases the accuracy of consumers' perceptions about both manipulators and honest products. Dou-

bling organic reviews reduces the mean absolute error in perceived quality by 50% and 59% for manipulators and honest products, respectively. Panel (b) presents the welfare gains from additional information. Increasing organic reviews by 48% achieves the same welfare gains as eliminating fake reviews.

Importantly, both the information and welfare gains are quite concave, implying that the cost-efficiency of increasing fake reviews is likely to decline rapidly. We find this to be the case when examining the per-review gains to profits and consumer welfare in Figure C.4. Panel (a) shows that increasing organic reviews by 1% increases consumer welfare by more than \$3.10 per review, while doubling the number of organic reviews increases welfare by just \$1.10 per review. Likewise, Panel (b) shows that initial increases in organic reviews raise platform revenues by more than \$0.47 per review, while doubling organic reviews generates only approximately \$0.20 in additional revenue per organic review.

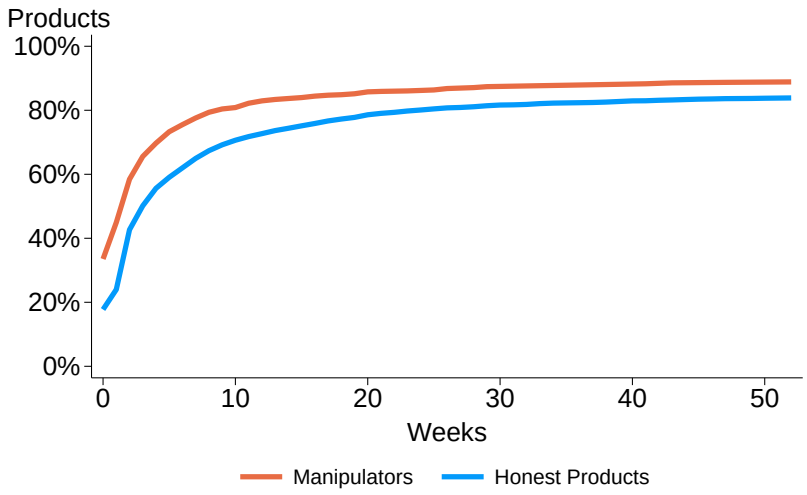
6.4 Incentives to Purchase Fake Reviews

In this section, we explore the financial incentives to purchasing fake reviews. We follow He et al. (2022b) in modeling the typical cost for purchasing a fake review as approximately $mc + 0.25p$. This includes the marginal cost of the product the reviewer receives for free, as well as the total of Amazon fees, PayPal fees, and sales taxes required for the reviewer to purchase through Amazon and be reimbursed via PayPal. See Appendix C.5.6 for details.

The benefits to purchasing fake reviews depend on a product's margin and the extent to which additional positive reviews induce additional demand. Unlike the costs, which are paid only once, the benefits of purchasing a fake review accrue over time. So long as the review remains, prospective consumers observe a rating that is slightly higher than they would have inferred from organic reviews alone.

In order to assess the benefits of purchasing fake reviews relative to their costs, we calculate the length of time required for each product to break even on its first purchased fake review. Figure 14 shows that the typical ratings manipulator breaks even on their first

Figure 14: Weeks to Break Even on Purchasing the First Fake Review



fake review very quickly: 80% earn sufficient additional profit to cover the cost within just 9 weeks. While purchasing fake reviews would also likely be lucrative for many honest products, the economics are not quite as favorable: it would take 24 weeks for 80% of honest products to break even. We explore this further in Appendix C.5.6 and find that manipulators tend to have both a lower cost to purchasing fake reviews and anticipate greater benefits relative to their honest competitors. The former is due to manipulators typically having lower marginal costs and prices. The latter is due to manipulators typically having both fewer and worse organic reviews, meaning that additional positive reviews shift consumers' perceptions more.

Note that while the costs and benefits we examine do explain some variation in which products purchase fake reviews, they are far from capturing the full story. Most notably, our model suggests that the majority of honest products would eventually generate additional profits from a fake review that exceeds the cost. Therefore, other factors—such as concerns about penalties and non-financial motives—must explain these products' decisions not to purchase fake reviews. Future research might aim to better understand these other factors.

Given the substantial financial benefits of purchasing positive fake reviews, a natural question is whether firms could also benefit from purchasing negative fake reviews for competitors' products. We explore this in Appendix C.5.7 and find that, in general, purchasing

negative fake reviews for competitors is far less profitable than purchasing positive fake reviews for one's own product. Two factors drive this. The first is that purchasing a fake review for a competitor is more costly because it entails paying full price for the competitor's product. The second is that the fake review purchaser does not capture all of the demand diverted from its competitor. A third unmodeled factor is that negative fake reviews may be particularly risky, since the affected competitor is strongly incentivized to identify and report the behavior to regulators or the platform.

7 Conclusion

A core mission of consumer protection regulators is protect consumers from deceptive practices. An increasingly prevalent form of deception is the manipulation of reputation systems by sellers on two-sided online platforms. In this paper we bring new empirical evidence on the magnitude and nature of consumer harms from this practice in a highly relevant empirical setting: the use of fake product reviews by third-party sellers on Amazon.com.

Our approach is to formally model the different ways in which fake reviews can affect outcomes for consumers, sellers, and the platform, and then use novel data sources to assess these empirically. We simulate equilibrium effects of fake reviews on the Amazon marketplace and quantify the different channels by which consumers are impacted. We show that the presence of fake reviews harms consumers, and do so principally through the misinformation channel. In contrast, mistrust has more modest net effects, as the consumer harms from the mistakes it generates are largely offset by increased price competition. While Amazon benefits from consumers' trust in ratings, they also benefit from misinformation increasing sales. As such, Amazon doesn't benefit from preventing manipulation, and is especially harmed by enforcement that reduces misinformation without increasing consumers' trust.

Our findings highlight that both misinformation and mistrust have important implications for the marketplace. They shift consumer demand, induce competitive responses, and guide the platforms' incentives to limit rating manipulation. Regulators should look to these

channels when evaluating the implications of rating manipulation for two-sided marketplaces.

References

Akesson, J., Hahn, R. W., Metcalfe, R. D., and Monti-Nussbaum, M. (2022). The impact of fake reviews on demand and welfare. *Working Paper*.

Armstrong, M. and Zhou, J. (2022). Consumer information and the limits to competition. *American Economic Review*, 112(2):534–77.

Beggs, S., Cardell, S., and Hausman, J. (1981). Assessing the potential demand for electric cars. *Journal of econometrics*, 17(1):1–19.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63:841–890.

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

Chakraborty, I., Kim, M., and Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, 59(3):600–622.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restriction. *Journal of Econometrics*, 34:305–334.

CMA (2020). Cma investigates misleading online reviews. `https://www.gov.uk/government/news/cma-investigates-misleading-online-reviews`. Accessed: 2024-03-18.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with pyblp. *RAND Journal of Economics*.

Dai, W. D., Jin, G., Lee, J., and Luca, M. (2018). Aggregation of consumer ratings: an application to Yelp.com. *Quantitative Marketing and Economics (QME)*, 16(3):289–339.

Dellarocas, C. (2006). Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management science*, 52(10):1577–1593.

Dranove, D. and Jin, G. Z. (2010). Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature*, 48(4):935–963.

Einav, L., Farronato, C., and Levin, J. (2016). Peer-to-peer markets. *Annual Review of Economics*, 8(1):615–635.

Feldman, E., Tosyali, A., and Overgoor, G. (2025). Addressing large-scale reviewer recruitment on amazon: A reviewer-centric approach to the fake review problem. *SSRN*.

Fradkin, A. and Holtz, D. (2023). Do incentives to review help the market? evidence from a field experiment on airbnb. *Marketing Science*, 42(5):853–865.

FTC (2023). Trade regulation rule on the use of consumer reviews and testimonials. 16 CFR 465: 88 FR 49364, RIN: 3084-AB76.

FTC (2024). Trade regulation rule on the use of consumer reviews and testimonials. 16 CFR 465: RIN: 3084-AB76.

Glazer, J., Herrera, H., and Perry, M. (2020). Fake reviews. *The Economic Journal*.

He, S. and Hollenbeck, B. (2020). Sales and rank on amazon.com. Technical Note.

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).

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

Hopenhayn, H. and Saeedi, M. (2023). Optimal Information Disclosure and Market Outcomes. *Theoretical Economics*.

Hui, X., Jin, G. Z., and Liu, M. (2022). Designing Quality Certificates: Insights from eBay. NBER Working Papers 29674, National Bureau of Economic Research, Inc.

Hui, X., Saeedi, M., Shen, Z., and Sundaresan, N. (2016). Reputation and regulations: Evidence from ebay. *Management Science*, 62.

Johnen, J. and Ng, R. (2024). Harvesting ratings. Technical report, University of Bonn and University of Mannheim, Germany.

Lam, H. T. (2021). Platform search design and market power. *Job Market Paper, Northwestern University*.

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

Long, F. and Liu, Y. (2024). Platform Manipulation in Online Retail Marketplace with Sponsored Advertising. *Marketing Science*, 43(2):317–345.

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

MacKay, A. and Miller, N. (2024). Estimating models of supply and demand: Instruments and covariance restrictions. *American Economic Journal: Microeconomics*.

Mayzlin, D., Y., D., and Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *The American Economic Review*, 104:2421–2455.

Punj, G. N. and Staelin, R. (1978). The choice process for graduate business schools. *Journal of Marketing research*, 15(4):588–598.

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):1944–1971.

Saeedi, M. and Shourideh, A. (2020). Optimal Rating Design under Moral Hazard. Papers 2008.09529, arXiv.org.

Saraiva, G. (2020). Incentives to fake reviews in online platforms. Working Paper.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8:321–340.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552.

Vatter, B. (2021). Quality disclosure and regulation: Scoring design in medicare advantage. Working Paper.

Vellodi, N. (2018). Ratings design and barriers to entry. Working Papers 18-13, NET Institute.

Yasui, Y. (2020). Controlling fake reviews. Working Paper.

Yu, C. (2024). The welfare effects of sponsored product advertising. *Available at SSRN 4817542*.

A Model Appendix

A.1 Relationship between quality and rating for fake review purchasers

A product j with quality q_j receives organic reviews such that its rating $R_j = q_j$ deterministically. Fake reviews shift ratings such that R_j lies above q_j. The impact of fake reviews on ratings, $R_j - q_j$, is governed by a beta distribution with mean 0.5 that is scaled to lie on the interval $[q_j, 1]$. Formally, $R = q + (1-q)\nu$, where $\nu \sim Beta(3,3)$ and $E[\nu] = 0.5$. Figure A.1 describes the shape of the distribution of R_j for a given q_j. Figure A.2 depicts the joint distribution of (q_j, R_j).

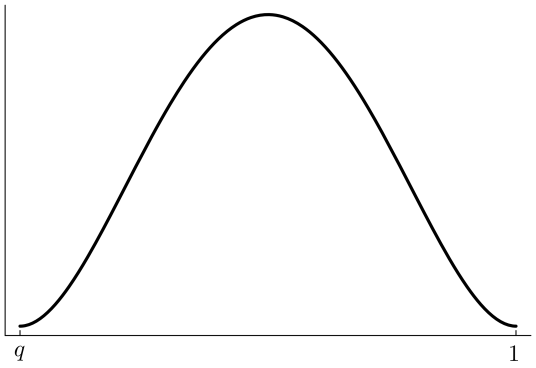


Figure A.1: Distribution of R_j with fake reviews.

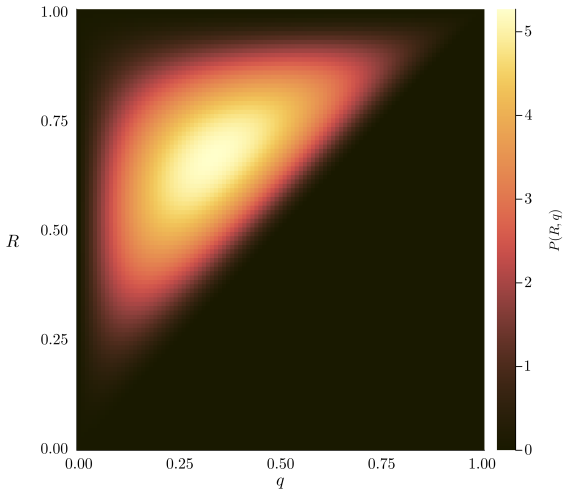


Figure A.2: Joint distribution of quality and R

B Data Appendix

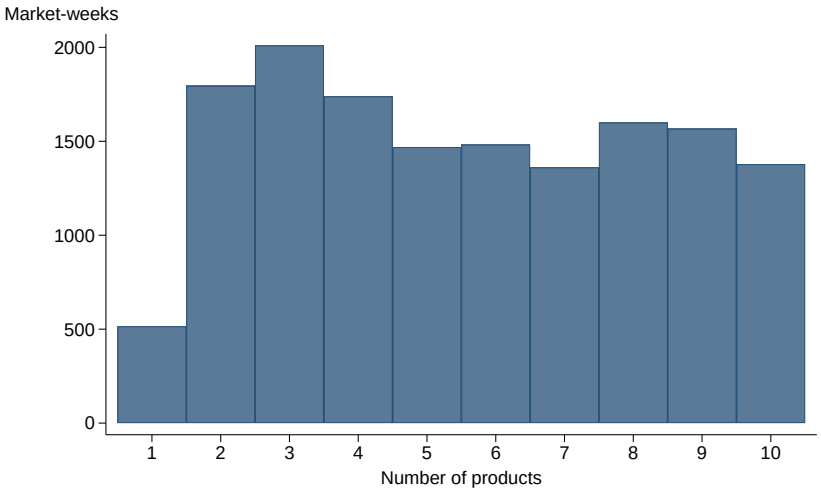
B.1 Data Description

Table B.1 reports the top product categories and subcategories in the dataset. Notably, products using fake reviews are found across a wide range of categories and subcategories. Our definition of markets is similar in specificity to the subcategories as defined by Amazon. Each product belongs to a subcategory, which in turn belongs to a category. For the demand estimation, we consider each week to be a separate market. We then remove products observed in 2 or fewer weeks, and remove markets that have 2 or fewer products across all weeks. The final dataset has 617 "markets" (as defined by search co-appearance), 47 weeks, and 3832 unique products. There are 14915 market-weeks, with the modal market-week having 3 products. The distribution of the number of products in a market-week is depicted in Figure B.1 .

Table B.1: **Top Categories of Fake Review Purchasers**

Category	Product-weeks
Beauty & Personal Care	106
Health & Household	95
Home & Kitchen	75
Kitchen & Dining	59
Tools & Home Improvement	59
Cell Phones & Accessories	43
Pet Supplies	38
Sports & Outdoors	35
Patio, Lawn & Garden	32
Electronics	27

Figure B.1: Products per market



B.2 Estimating the Share of Fake Reviews

We discuss in general terms in section 3.1 how we estimate the share of a product's reviews that are fake. In this section we provide more details on this procedure. We rely on the classification model from He et al. (2022a), which details a way to discern which products are fake review purchasers, given the network structure of reviews. They train a classifier model on features derived from the product-reviewer network as well as review features, text and photo features, and product metadata. This method performs well out of sample, detecting fake review buyers with an accuracy rate of .86 and AUC score of .93.

We use this prediction algorithm from He et al. (2022a) to classify all products in the product-reviewer network as buying fake reviews or not. This network is composed of all the FRPs and their competitors, as well as any other products that reviewers of these products also left reviews for. This consists of 25,840 products and 1.7 million reviews. For each of the fake review products and their close competitors, for a random sample of roughly 25% of their reviews, we also scraped the pages of the authors of those reviews in order to know the full set of products reviewed by those authors.

We use this data to identify any reviewers observed leaving multiple five-star reviews for products classified as purchasing fake reviews. We label these reviewers as "fake reviewers" and find 27,045 fake reviewers out of the 368,386 unique reviewers in this data, or roughly 7%. Then, for each product j that we know purchases fake reviews, we can compute the fraction of j's five-star reviews that came from these fake reviewers. This is measured as a fraction of the subsample of reviewers for which we observe their full rating history. That is, we do not compute the fraction of all reviewers that are designated as fake reviewers, but the fraction of all reviewers with observed ratings histories that are designated as fake reviewers. This provides an estimate of the proportion of fake reviews for that product, but with some noise due to the fact that we only observe ratings histories for a sample of each product's reviewers. For the set products we observe buying fake reviews, the average estimated share

of fake reviews is 47% with a median share of 50%.[26]

Feldman et al. (2025) also classify fake reviewers by constructing a Graphical Neural Network model of reviewers to exploit relational patterns between fake reviewers alongside with behavioral and review content features. This method is more data intensive, and hence only provides results for roughly one third as many reviewers as we include. Between the two studies, there are 30,220 common reviewers that have been classified, and there was a 77.21% agreement between the two approaches. Of the 22.79% reviewers for whom the two methods disagreed, 13.82% reviewers were only classified as fake by the GNN model, and 8.97% were only classified as fake by our method. Nevertheless, with the high agreement and roughly symmetric disagreement, the correlation between the estimates of the share of fake reviews at the weekly product level derived from their predictions and ours is .95.

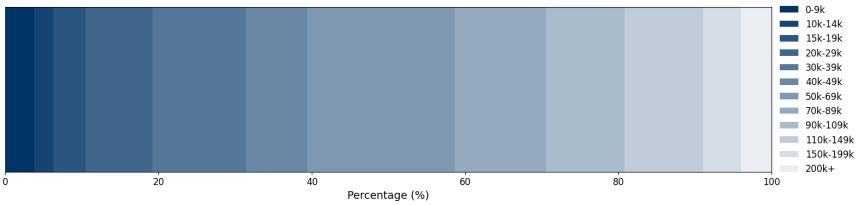
B.3 Survey Experiment

B.3.1 Demographics

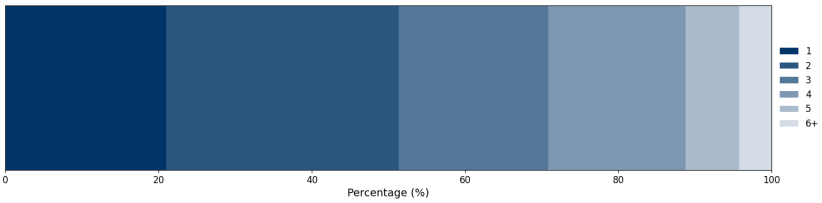
[26] By contrast, among honest products, we observe only .6% of their reviews are left by these fake reviewers.

Figure B.2: **Demographics of Survey Respondents**

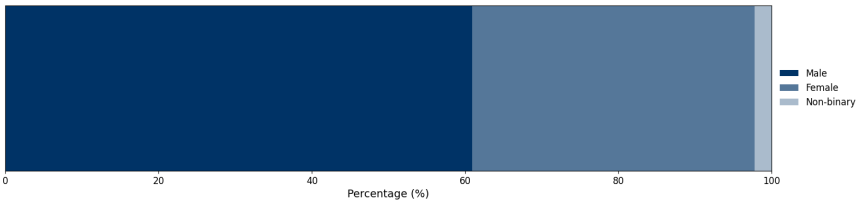
(a) Income



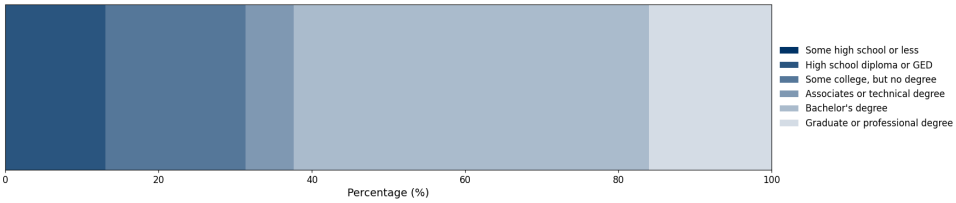
(b) Household Size



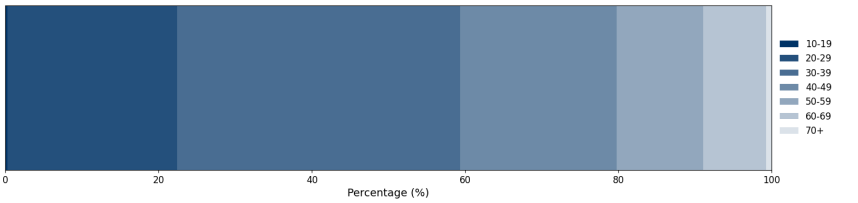
(c) Gender



(d) Education



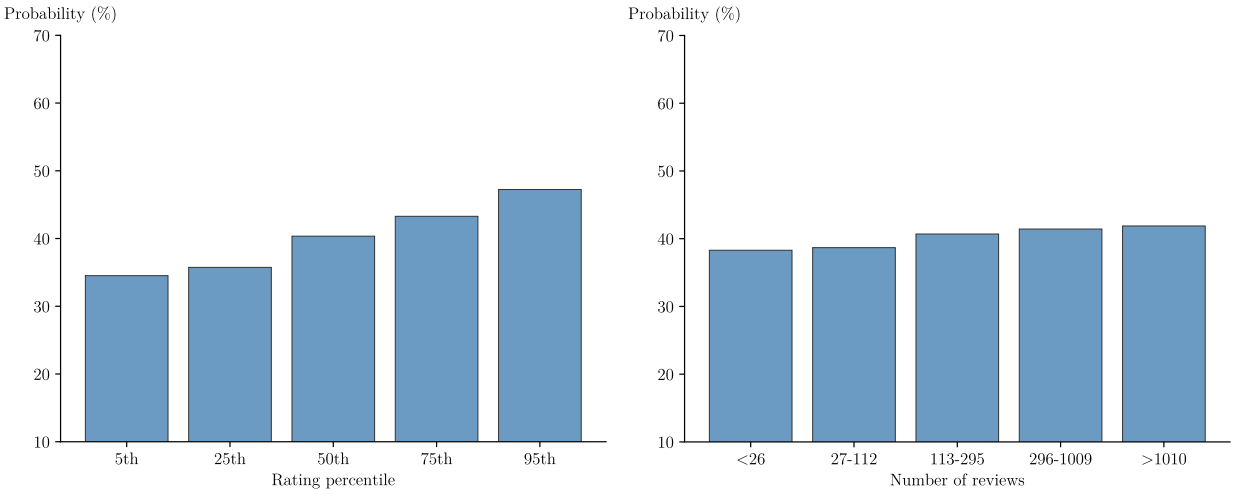
(e) Age



Notes: For subfigure (B.2d), 0.5% percent of participants put "Prefer Not to Say".

B.3.2 Responses by Rating and Number of Reviews

Figure B.3: Beliefs About Fake Reviews by Product Characteristics

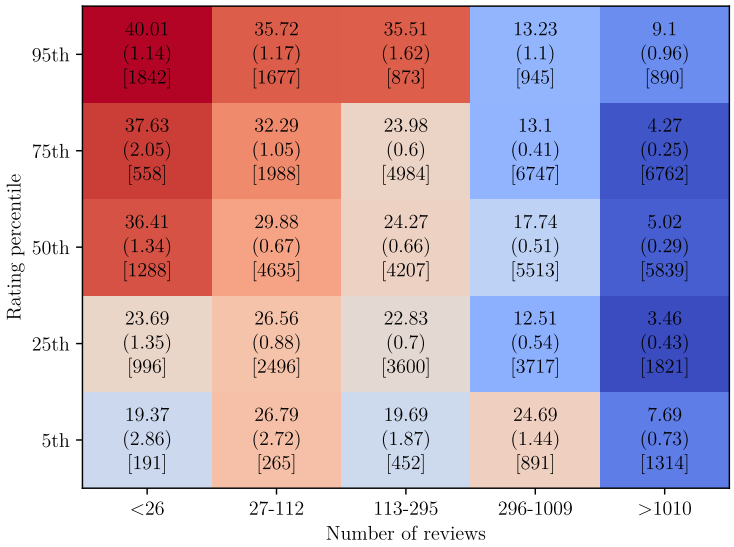


(a) By Rating

(b) By # Reviews

B.3.3 Empirical Prevalence by Rating and Number of Reviews

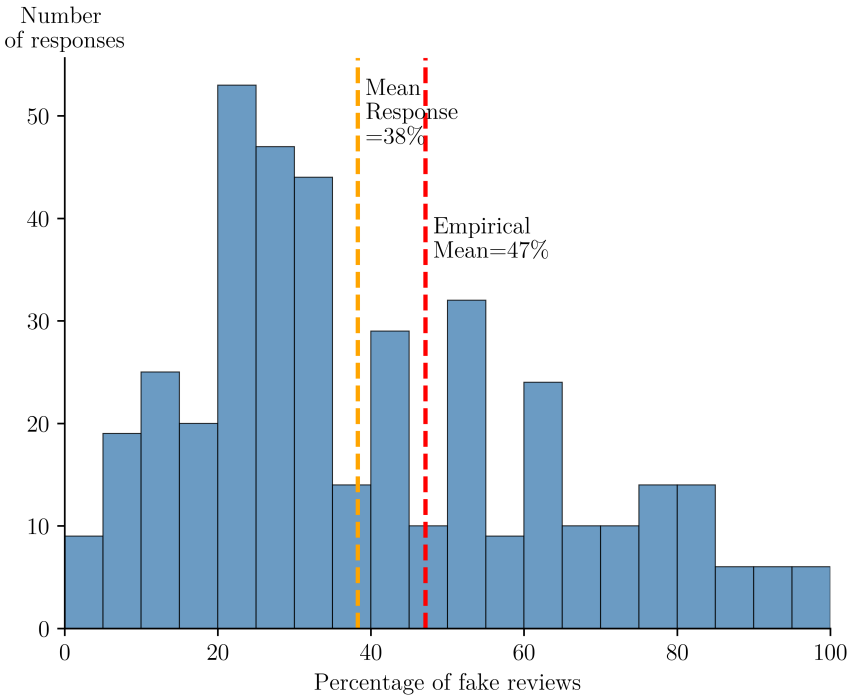
Figure B.4: Empirical Prevalence by Ratings and Number of Reviews



Notes: In each cell, the first number is the mean response for the probability of manipulation, the standard error of the mean is parenthesized, and the number of responses is in square brackets.

B.3.4 Beliefs about the Frequency of Fake Reviews

Figure B.5: Surveyed Perceptions of θ^F



B.3.5 Responses by Product Category

Table B.2: Average Response by Product Category

Category	Total	FRP	HP
Arts, Crafts, & Sewing	33.3	34.7	32.4
Automotive	38.5	38.8	38.2
Baby	37.2	31.8	39.8
Beauty & Personal Care	39.9	50.4	34.3
Camera & Photo	43.7	46.3	42.6
Cell Phones & Accessories	39.0	41.8	37.6
Clothing, Shoes & Jewelry	40.0	42.2	39.0
Computers & Accessories	46.1	52.3	43.6
Electronics	44.1	46.5	42.9
Health & Household	35.3	37.9	34.1
Home & Kitchen	44.4	40.6	46.2
Industrial & Scientific	18.0	18.5	17.8
Kitchen & Dining	37.0	36.2	37.2
Office Products	36.9	36.0	37.4
Patio, Lawn & Garden	39.3	35.7	41.6
Pet Supplies	41.0	43.3	39.9
Sports & Outdoors	27.7	25.2	28.7
Tools & Home Improvement	38.2	39.4	37.6
Toys & Games	44.8	49.3	43.2

B.3.6 Review Histograms

We show 60% of our sample the "overall" histogram depicting the distribution of all review for all products with the given number of reviews and rating (Figure B.6). The remaining 40% are randomized between highly bimodal (Figure B.7) and unimodal (Figure B.8) histograms (95th and 5th percentiles of variance, respectively). Figures B.9 and B.10 contrast responses under different histogram shapes. First, we observe that respondents are more trusting of products with few reviews or low ratings when they have a highly unimodal distribution. Second, respondents appear to report slightly higher probabilities when shown ratings histograms with higher variances. In general, however, the contrast in responses is not dramatic. We therefore do not condition on histogram shape in our main analysis.

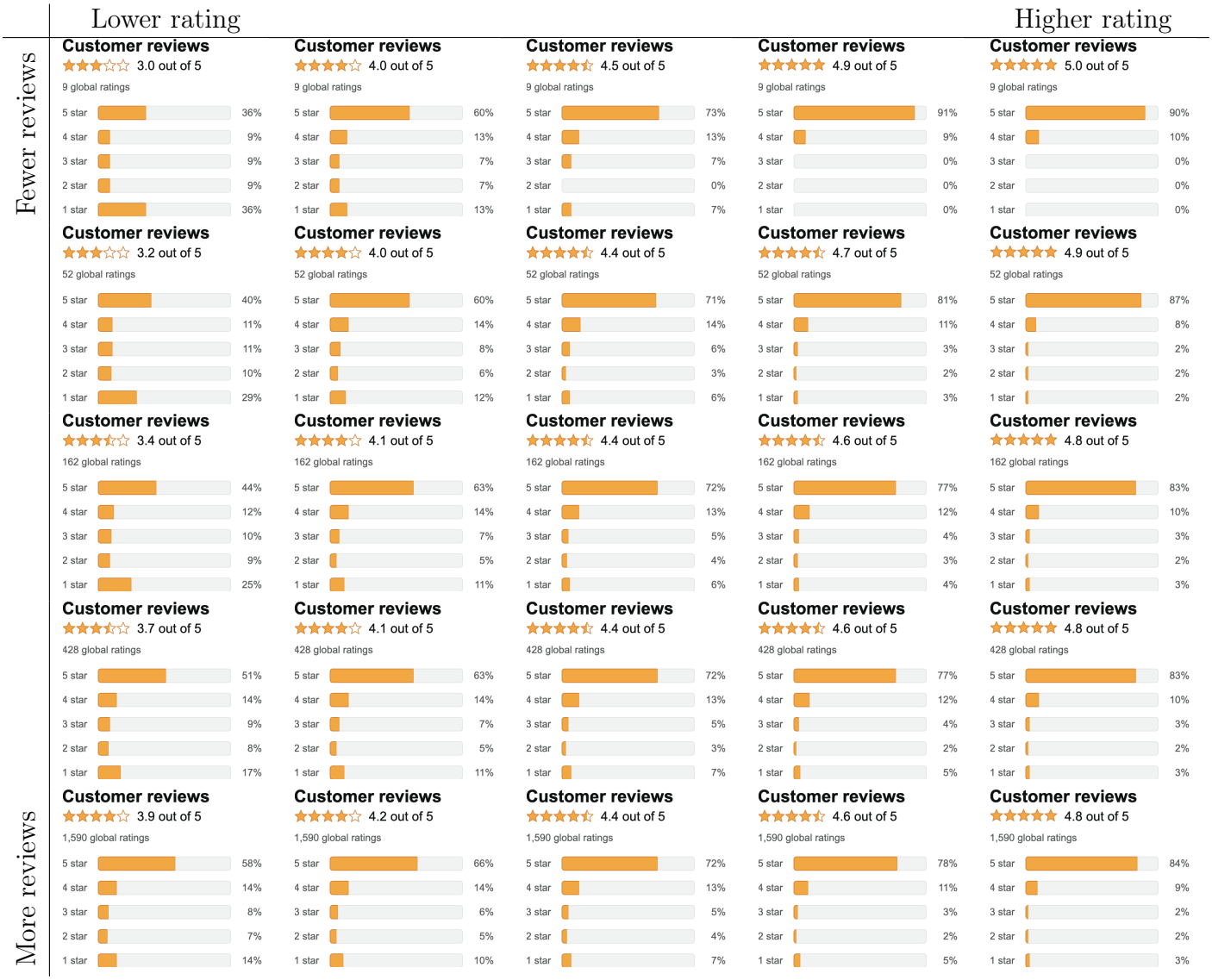


Figure B.6: Overall histograms

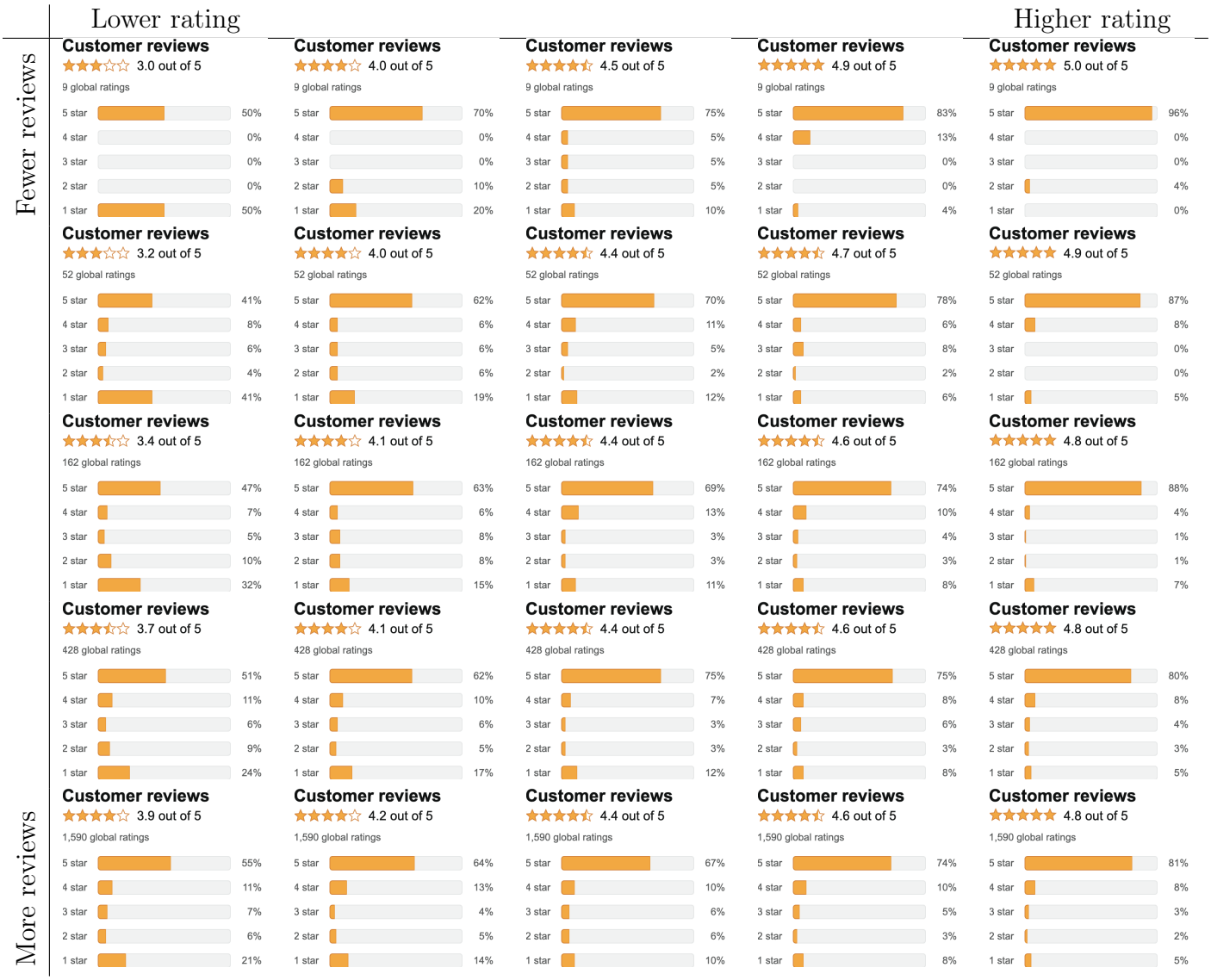


Figure B.7: Bimodal histograms

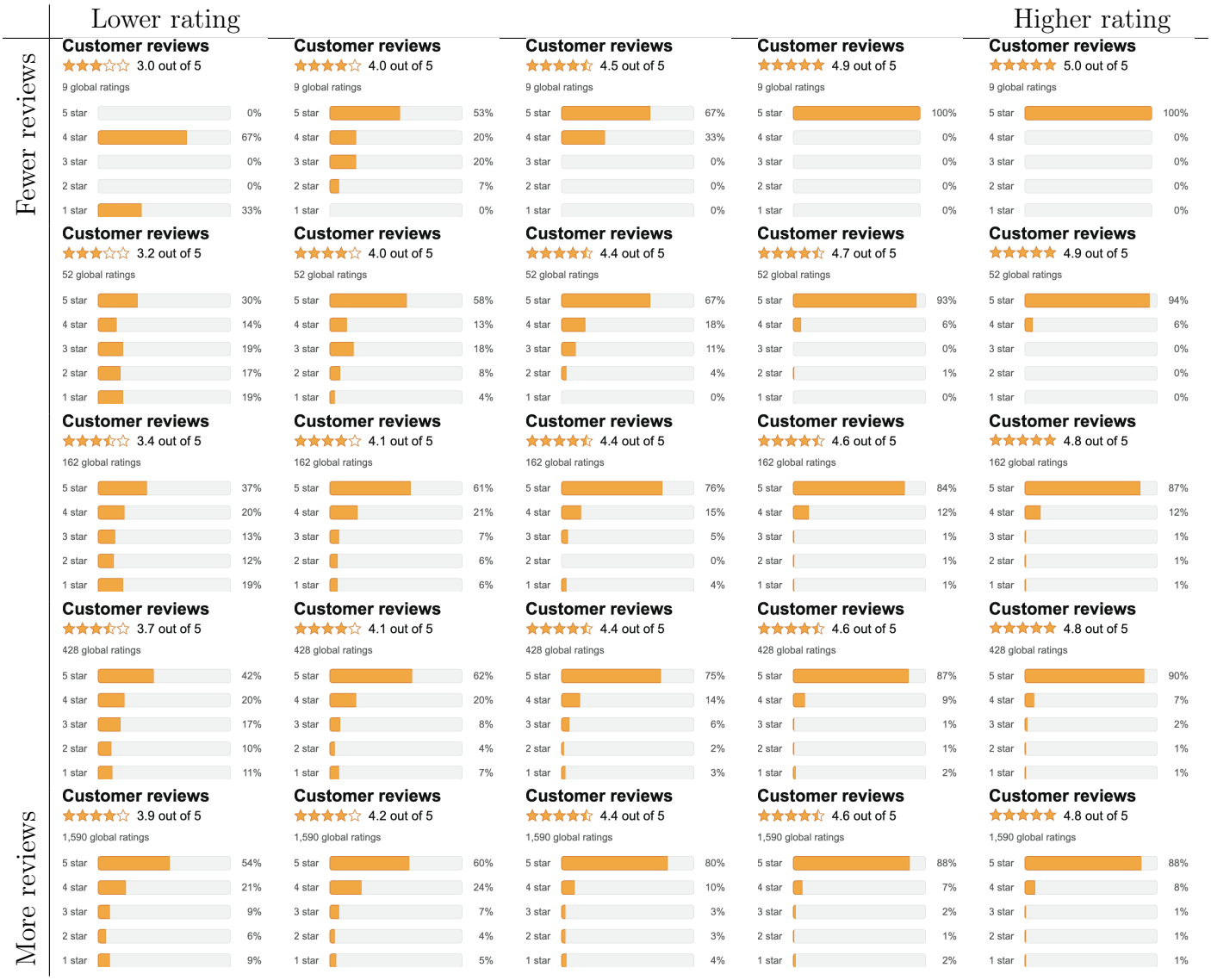
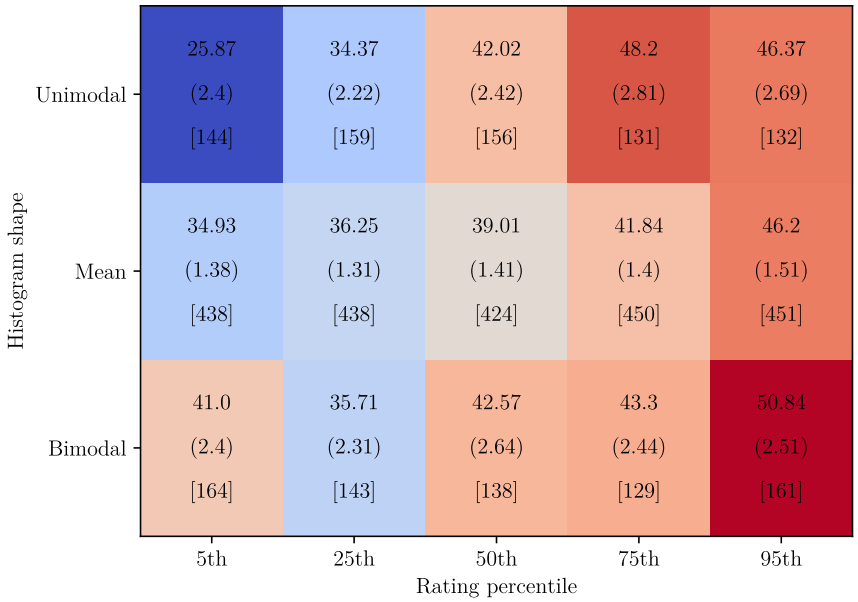


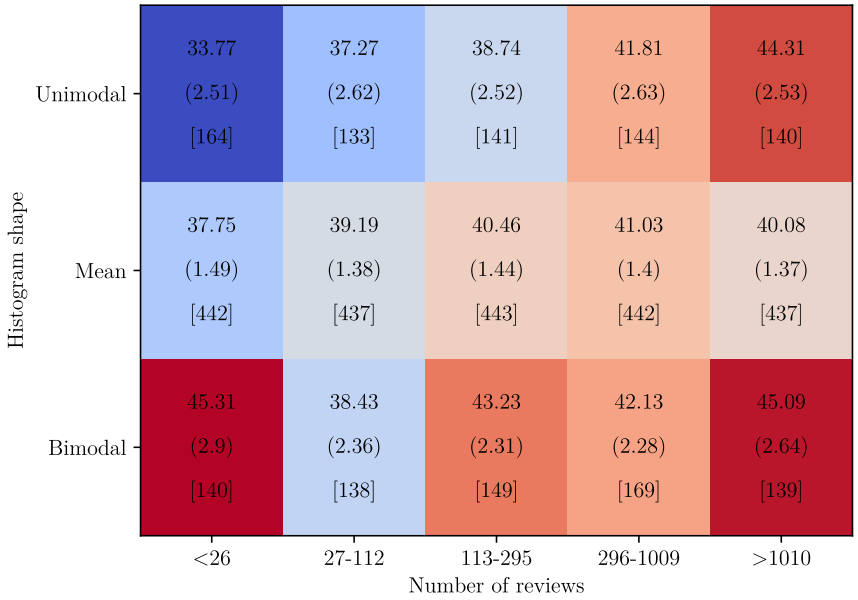
Figure B.8: Unimodal histograms

Figure B.9: Comparing Across Histograms by Rating



Notes: In each cell, the first number is the mean response for the probability of manipulation, the standard error of the mean is parenthesized, and the number of responses is in square brackets.

Figure B.10: Comparing Across Histograms by Number of Reviews



Notes: In each cell, the first number is the mean response for the probability of manipulation, the standard error of the mean is parenthesized, and the number of responses is in square brackets.

B.3.7 Amazon Gift Card Sanity Check

For the question that displays the Amazon gift card, 0% of the respondents correctly responded 0%, and the mean response is 11%. Figure B.11 shows the histogram of responses. We test for a relationship between giving a response greater than 10% to the gift card question and other survey responses and find no relationship, and overall results are similar when this group are excluded.

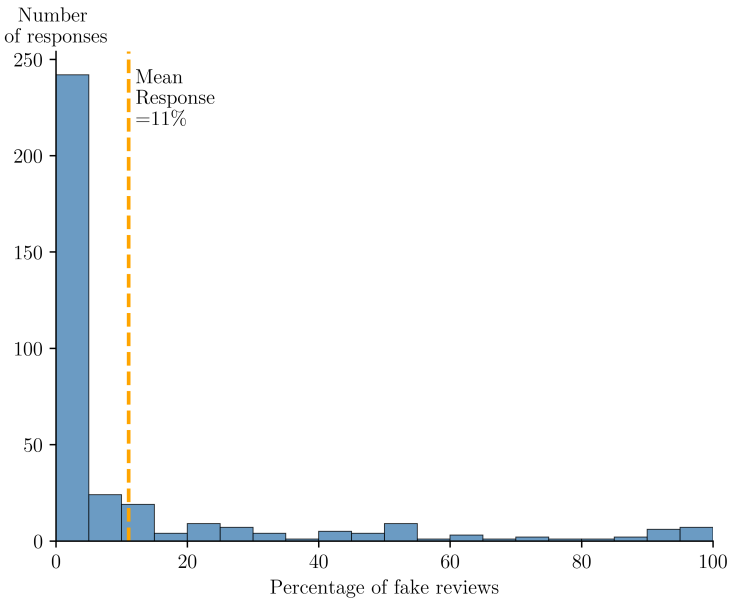


Figure B.11: Responses for Amazon Gift Card

B.3.8 Expanding Review Text

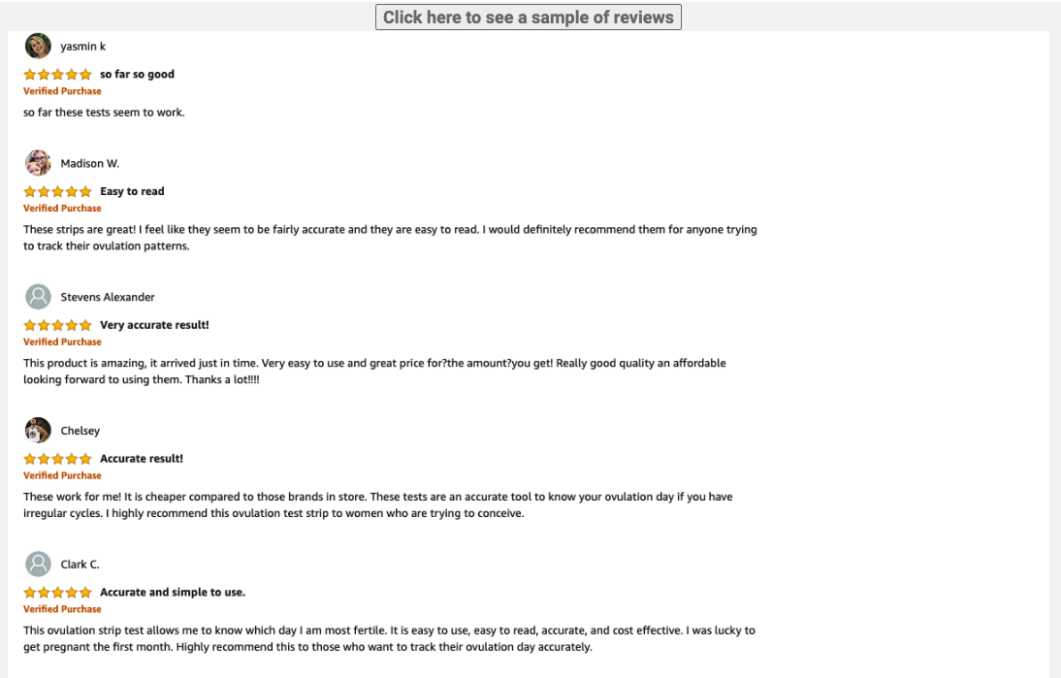
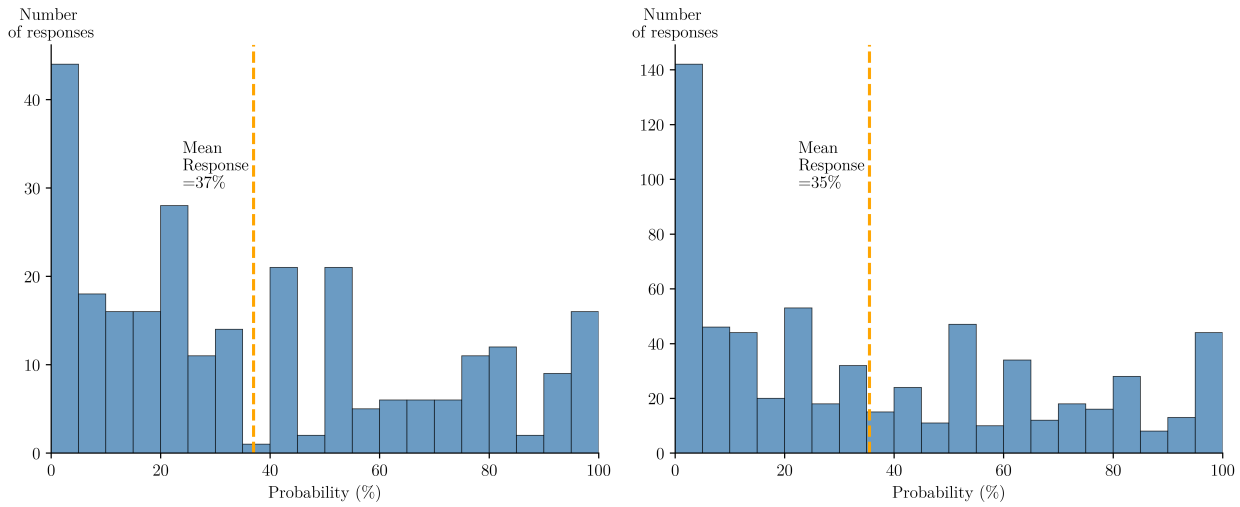


Figure B.12: Example expanded reviews.

Figure B.13: Perceptions of Fake Review Purchasing with Review Text



(a) Fake Review Purchasers

(b) Honest Products

C Details on Counterfactual Simulations

In this section, we detail our procedure for simulating counterfactuals. Sections C.1 and C.2 respectively present the models for how organic reviews and product search position

evolve stochastically. Section C.3 describes how we simulate and integrate over the distribution of counterfactual realizations. Section C.4 discusses how we compute welfare in counterfactual simulations given the potential disparity between consumers' perceptions of a product's quality and the product's true quality.

C.1 Organic Reviews

Organic reviews are *unpaid* reviews for a product left by consumers who purchased the product on Amazon.com. Our counterfactuals change the quantities of products sold and therefore the number of individuals able to leave an organic review. To incorporate this, we model the new organic reviews for a product in a given week as Poisson with a mean determined by the number of sales in the last two weeks. We focus on the previous two weeks, as the option to leave a review is likely to be most salient for recent purchasers. We estimate the extent to which recent sales affect the arrival of new organic reviews (n_{jt}) using a Poisson regression:

$$\log(E[n_{jt}|m(j), Q_{j,t-1}, Q_{j,t-1}]) = \alpha_{m(j)} + \gamma_1 \log(Q_{j,t-1}) + \gamma_2 \log(Q_{j,t-2}), \tag{8}$$

where $Q_{j,t-1}$ and $Q_{j,t-1}$ are the sales quantities in the previous two weeks, and α_m is a market-level intercept allowing the general propensity to leave reviews to differ across types of products.

Table C.1: Poisson Model of the Number of New Organic Reviews

Log(Quantity)		
Lag 1	0.771***	0.732***
	(0.0398)	(0.0413)
Lag 2		0.0483***
		(0.0128)
Market FEs	Y	Y
N. Obs.	56499	56499
Mean Dep. Var.	57.41	57.41
SD	300.6	300.6

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.1 presents the estimated relationship between recent sales volume and new organic reviews. As expected, we find that the arrival of new organic reviews scales substantially with recent sales. Our preferred specification used in our counterfactual simulations includes lags for each of the previous two weeks. The estimates imply that a log-point increase in the prior week's sales increases the expected number of new organic reviews by 0.73 log-points. Consistent with reviews deriving from very recent purchases, a log-point increase in sales two-weeks prior results in only a 0.05 log-point increase in expected new organic reviews.

Implementation While the estimates in Table C.1 effectively captures the timeseries dependence of organic reviews on recent sales, it does not necessarily capture product-specific variation in the number of organic reviews. To better allow our counterfactuals to hold this

constant, we leverage the model in equation (8) to simulate *perturbations* from the factual.

For ease of notation, let $\hat{n}_{jt}(Q_{j,t-1}, Q_{j,t-2}) := E[n_{jt}|m, Q_{j,t-1}, Q_{j,t-2}]$ be the expected number of organic reviews for product j in period t given an arbitrary $(Q_{j,t-1}, Q_{j,t-2})$ and $\hat{n}^o_{jt} := E[n_{jt}|m, Q^o_{j,t-1}, Q^o_{j,t-2}]$ be the expected number of organic reviews given the observed $(Q_{j,t-1}, Q_{j,t-2})$. When simulating n_{jt} in the counterfactual, we simulate in a way that ensures $E[n_{jt} - n^o_{jt}] = \hat{n}_{jt}(Q_{j,t-1}, Q_{j,t-2}) - \hat{n}^o_{jt}$.

When $\hat{n}_{jt}(Q_{j,t-1}, Q_{j,t-2}) - \hat{n}^o_{jt} > 0$ we draw additional counterfactual organic reviews from a Poisson with mean $\hat{n}_{jt}(Q_{j,t-1}, Q_{j,t-2}) - \hat{n}^o_{jt}$. When $\hat{n}_{jt}(Q_{j,t-1}, Q_{j,t-2}) - \hat{n}^o_{jt} < 0$ we delete each observed new organic review with probability $\frac{\hat{n}^o_{jt} - \hat{n}_{jt}(Q_{j,t-1}, Q_{j,t-2})}{\hat{n}^o_{jt}}$. In both cases, the distribution of additional (or fewer) reviews is the same as implied by Equation 8.

This approach has two key advantages. The first is that it allows us to better respect observed differences in organic reviews across products due to factors outside our Poisson regression. The second is that it is "smoother" in ensuring that counterfactuals with $\frac{\hat{n}^o_{jt} - \hat{n}_{jt}(Q_{j,t-1}, Q_{j,t-2})}{\hat{n}^o_{jt}}$ close to zero are simulated to be very close to the observed data. In turn, this reduces the computational burden of the counterfactual analysis.

C.2 Search Position

When a consumer searches a keyword on Amazon.com, the website returns an algorithmically sorted list of products. Both our demand estimates and previous research (Ursu, 2018; Lam, 2021) indicate that a product's positioning on the search page will affect its sales. Since our counterfactuals affect both product ratings and sales, it is likely that it would also affect the search position for products on Amazon. In turn, these changes to search position may lead to further changes in ratings and sales. Therefore, incorporating changes to search positioning may be important in evaluating the counterfactuals. For example, if eliminating fake reviews lowers manipulator products' search position, then this intensifies the reduction in demand they would experience due to deletion. Moreover, these effects may evolve dynamically if reduced sales in a given week affect search positioning in future weeks. In this section, we discuss how we incorporate such dynamic adjustments to search position in our counterfactual simulations.

While Amazon's precise search and ranking algorithm is proprietary, we can infer a number of Amazon's incentives in search. First, given that Amazon earns a percentage of marketplace sales, Amazon benefits from prioritizing products that consumers are likely to purchase. Second, since Amazon's business model relies heavily on repeated customers and subscriptions, Amazon is incentivized to prioritize products that the consumer will not regret purchasing. Third, Amazon also earns revenue from sponsored listings, a form of advertising where products pay for favorable search positioning.[27]

We try to capture these incentives in a simple model in which Amazon assigns a score to each product and sorts products in order of score. We model the score v_{jt} for product j at time t as linear in product characteristics X_{jt} plus a Gumbel-distributed error ϵ_{jt}:

$$v_{jt} = \beta \mathbf{X}_{jt} + \epsilon^v_{jt} \tag{9}$$

[27] See Yu (2024) for a careful analysis of sponsored product advertising on Amazon and its implications for the platform.

This approach is often referred to as a rank-ordered logit (Beggs et al., 1981) or exploded logit (Punj and Staelin, 1978).

We choose the covariates $\mathbf{X}_{jt}$ to capture Amazon's three incentives in search detailed above. First, we include recent sales, as these are a strong predictor of which products consumers are likely to purchase this week. Second, we include product ratings, as these predict product quality and therefore whether consumers are likely to regret their purchase. Third, we include an indicator for whether product purchased a sponsored position, as this will mechanically improve their search position. We also include product fixed effects, which capture time-invariant product-level variation in search position due to factors such as unobserved quality.

We estimate β via maximum likelihood. Without loss of generality, let a search ordering of products in a market be $1, \ldots, J$. This occurs when $v_1 \geq ... \geq v_J$. Therefore, the likelihood of observing ordering $1, \ldots, J$ is:

$$\begin{aligned}
P(v_1 \geq ... \geq v_J) &= P(v_1 \geq v_2 | v_2 \geq ... \geq v_J) P(v_2 \geq v_3 | v_3 \geq ... \geq v_J) ... P(v_{J-1} \geq v_J) \\
&= \prod_{j=1}^{J-1} P(v_j \geq v_{j+1} | v_{j+1} \geq ... \geq v_J) \\
&= \prod_{j=1}^{J-1} P(v_j \geq v_{j+1} | v_{j+1} \geq \max_{k=j+2}^{J} v_k) \\
&= \prod_{j=1}^{J-1} P(v_j \geq \max_{k=j+1}^{J} v_k | v_{j+1} \geq \max_{k=j+2}^{J} v_k) \\
&= \prod_{j=1}^{J-1} P(v_j \geq \max_{k=j+1}^{J} v_k) \\
&= \prod_{j=1}^{J-1} \frac{\exp(\beta \mathbf{X}_j)}{\sum_{k=j}^{J} \exp(\beta \mathbf{X}_k)}.
\end{aligned}$$

Table C.2 shows the maximum likelihood estimates given varying sets of covariates. Across all specifications, lagged demand, lagged ratings, and sponsorship all strongly predict more favorable search positioning. Column 1 is our preferred specification that we use in simulating counterfactuals. Under this model, a log-point increase in lagged demand in both of the prior two weeks improves the average product's relative position by 0.41. This is a relatively large effect given that our median market has 5 products. A log-point increase in the number of positive reviews in each of the last two weeks results in an improvement of relative position by 0.07. Finally, sponsorship improves relative position by an average of 0.49.

C.3 Simulating the Dynamic Evolution of Markets

In evaluating counterfactuals, we must account for the fact that organic reviews and search positions are stochastic and evolve dynamically. To do this, we simulate the distribution of paths that each market could have followed. To draw one path for market m, we

Table C.2: Hedonic model of product rank

Log Shares				
Lag 1	0.238***	0.246***	0.244***	0.243***
	(0.0121)	(0.0124)	(0.0123)	(0.0123)
Lag 2	0.174***	0.183***	0.183***	0.182***
	(0.0109)	(0.0111)	(0.0111)	(0.0111)
Log N. Positive Reviews				
Lag 1	0.0381***			
	(0.00651)			
Lag 2	0.0281***			
	(0.00594)			
Cumulative rating				
Lag 1		0.0625***		0.0596**
		(0.00791)		(0.0226)
Lag 2		0.0633***		0.0648**
		(0.00782)		(0.0219)
Weekly rating				
Lag 1			0.0528***	0.00351
			(0.00740)	(0.0212)
Lag 2			0.0567***	0.0000425
			(0.00728)	(0.0206)
Log Cumulative N. Reviews				
Lag 1		0.0387***		0.0318***
		(0.00645)		(0.00700)
Lag 2		0.0287***		0.0233***
		(0.00586)		(0.00634)
Log Weekly N. Reviews				
Lag 1			0.0311***	0.0179**
			(0.00604)	(0.00661)
Lag 2			0.0213***	0.0115
			(0.00570)	(0.00617)
Sponsored	0.490***	0.484***	0.489***	0.484***
	(0.0353)	(0.0353)	(0.0352)	(0.0352)
Log Age	0.236**	0.165*	0.234**	0.183*
	(0.0821)	(0.0830)	(0.0827)	(0.0834)
Product FEs	Y	Y	Y	Y
Observations	330484	330484	330484	330484

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

initialize the path with the first two weeks observed in the data and simulate forward iteratively starting with week 3. For period t, we draw the number of new organic reviews n_{jt} for each product j in market m based on the quantities of j sold in $t-1$ and $t-2$ as described in Section C.1. Each of the n_{jt} new reviews is positive with probability q_j. We likewise simulate the week t search positions of products in m according to the model described in Section C.2. These simulated positions in t are determined by the lagged market shares and ratings that were simulated for $t-1$ and $t-2$ along the given path. Search positions also depend on products' sponsorship decisions, which are assumed not to change from the factual. Our results are averaged over 30 simulated paths for each counterfactual.

C.4 Computing welfare

Consumers' purchasing decisions are based on their *decision utility*, which includes the expected quality they perceive for a product. Their *experience utility*, however, is based on the true quality of the product. While we also do not observe true quality, we are able to form a more accurate expectation than the consumer by leveraging our inference about the product-specific prevalence of fake reviews from Section 3.1. This approach allows us to infer the number of organic reviews (N^o) and the number of positive organic reviews (N^{o+}) and yields the following econometrician's posterior on quality:

$$\begin{aligned} P(q|N^{o+},N^o,F;\hat{\gamma}) &= \frac{P(N^{o+}|q,N^o,F)P(q|F;\hat{\gamma})}{P(N^{o+}|N^o,F)} \\ &= \frac{q^{N^{o+}}(1-q)^{N^o-N^{o+}}P(q|F;\hat{\gamma})}{\int q^{N^{o+}}(1-q)^{N^o-N^{o+}}dP(q|F;\hat{\gamma})}, \end{aligned} \tag{10}$$

where the first equality applies the assumption that $P(q|N^o,F) = P(q|F)$. We use this econometrician's posterior to characterize the quality consumers experience from their purchases.

To construct consumer welfare, we first define $\Delta\mathbb{E}q$ to be the difference between the consumer's and econemtrician's expectations about quality:

$$\Delta\mathbb{E}q := \int q dP(q|N^+,N;\hat{\gamma}) - \int q dP(q|N^{o+},N^o,F;\hat{\gamma}). \tag{11}$$

For a given good j in market t, consumer i's expected experience utility is $u_{ijt} - \beta_i\Delta\mathbb{E}q_{ijt}$. The consumer's welfare is then:

$$\begin{aligned} W_{it} &= \mathbb{E}_{\epsilon,\alpha_i,\beta_i}[u_{ij^*t}] - \mathbb{E}_{\epsilon_i,\alpha_i,\beta_i}[\beta_i\Delta\mathbb{E}q_{ij^*t}] \\ &= \mathbb{E}_{\epsilon_i,\alpha_i,\beta_i}[\max_j\{u_{ijt}\}] - \mathbb{E}_{\epsilon,\alpha_i,\beta_i}[\beta_i\Delta\mathbb{E}q_{ij^*t}] \\ &= \bar{W}_{it} - \sum_i\sum_j \beta_i s_{ijt}\Delta\mathbb{E}q_{ijt}, \end{aligned}$$

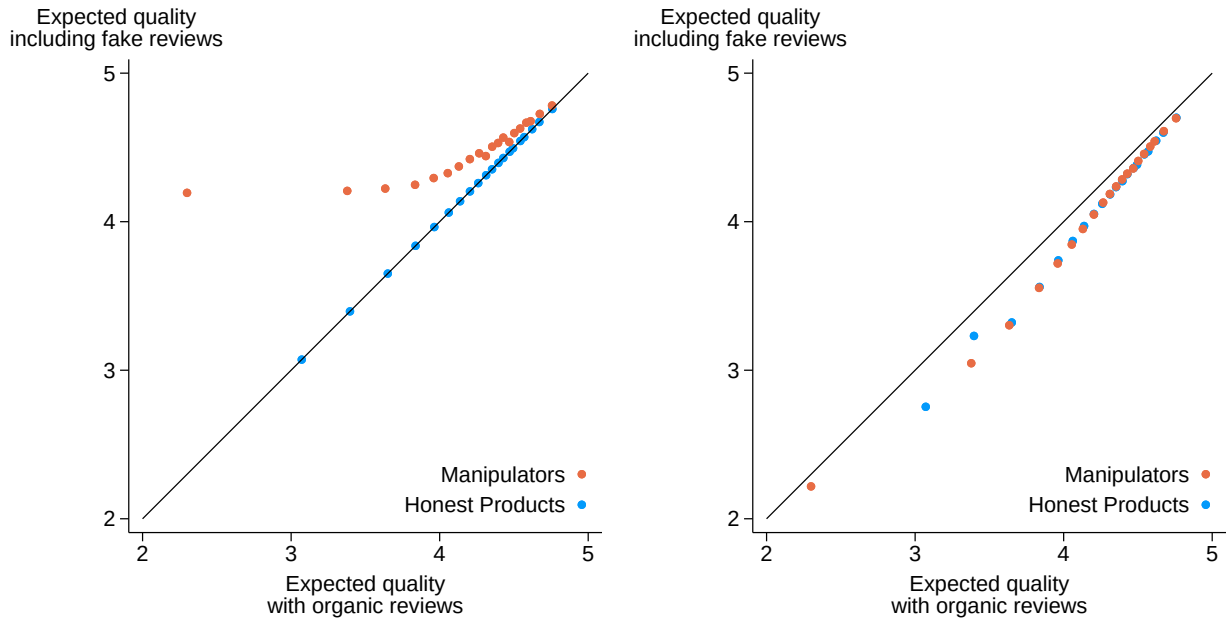
where j^* is chosen based on perceived quality, and $\bar{W}_{it}$ is expected decision utility.

C.5 Additional Counterfactual Analysis

C.5.1 Changes in Perceived Quality Under Misinformation and Mistrust

Figure C.1 presents binscatters of how perceived qualities change under misinformation or mistrust in isolation. The horizontal axis gives a product's expected quality given only its organic reviews, and the vertical axis gives consumer's beliefs. By inflating manipulators' ratings, misinformation increases the perceived quality of manipulators. Mistrust, on the other hand, makes consumers skeptical that observed ratings were earned and therefore reduces the perceived quality of all products.

Figure C.1: Perceived Qualities Under Misinformation and Mistrust in Isolation



(a) Misinformation (b) Mistrust

C.5.2 Summary of the Equilibrium Effect of Fake Reviews

Table C.3 presents a summary of the results from Figure 11.

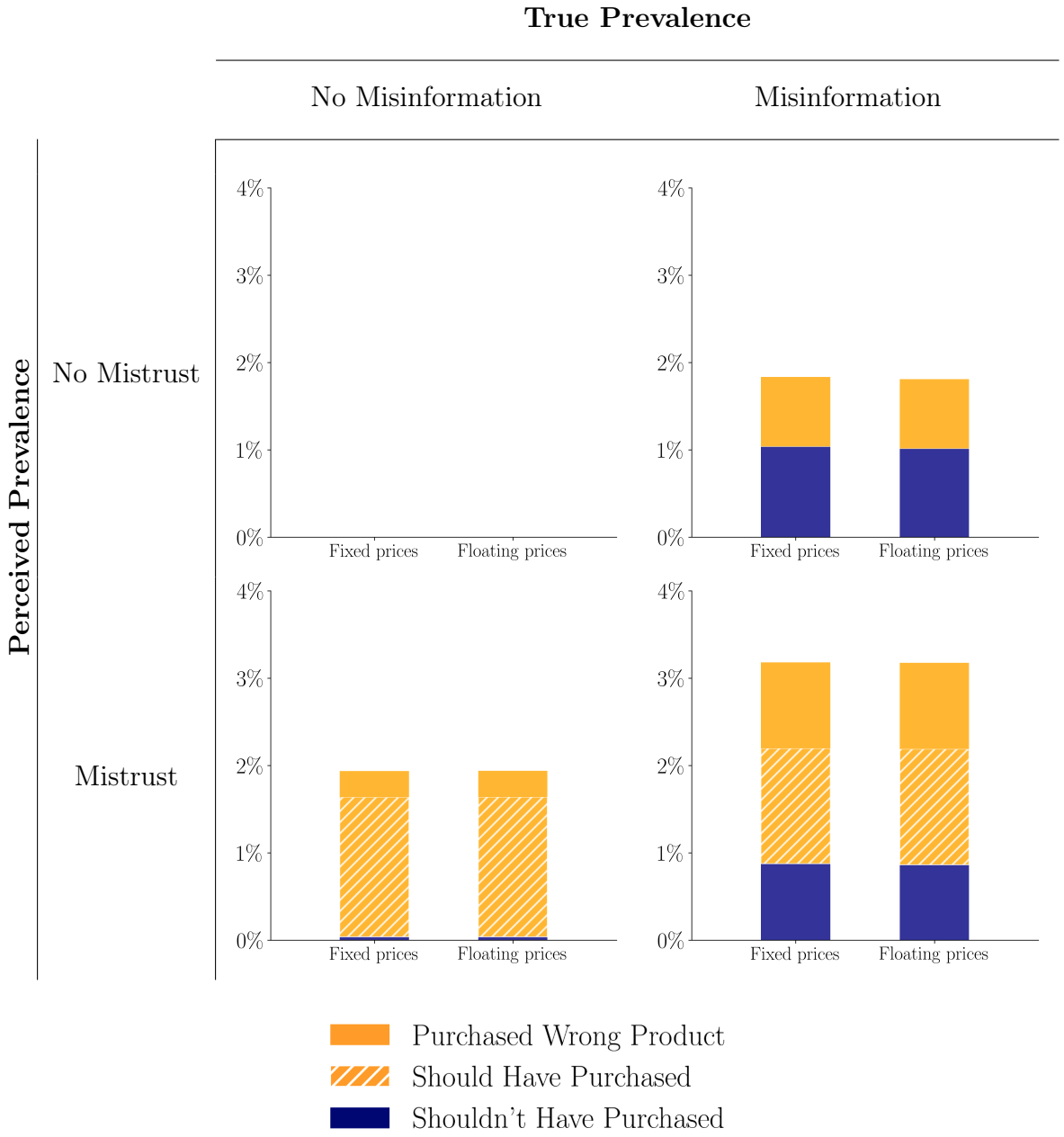
Table C.3: Summary of Counterfactual Changes in Figure 11

	Mean	P25	Median	P75	Total
Welfare (%)	-2.0	-1.0	-0.3	0.1	-0.8
FRP prices (%)	1.6	0.1	0.4	1.4	
HP prices (%)	-0.3	-0.3	-0.1	-0.0	
Overall prices (%)	0.1	-0.3	-0.1	0.0	
FRP quantity (%)	60.1	1.9	18.5	63.7	27.5
HP quantity (%)	-5.7	-8.6	-4.6	-2.1	-4.4
Overall quantity (%)	7.0	-7.7	-3.6	-0.3	-0.8
FRP revenue (%)	63.5	2.1	20.2	67.8	27.2
HP revenue (%)	-6.0	-9.0	-4.9	-2.2	-4.4
Overall revenue (%)	7.5	-8.1	-3.9	-0.3	-1.3
FRP profits (%)	65.7	2.3	21.3	70.6	30.1
HP profits (%)	-6.2	-9.4	-5.1	-2.3	-5.0
Overall profits (%)	7.7	-8.4	-4.0	-0.2	-1.1

C.5.3 Decomposing Mistakes

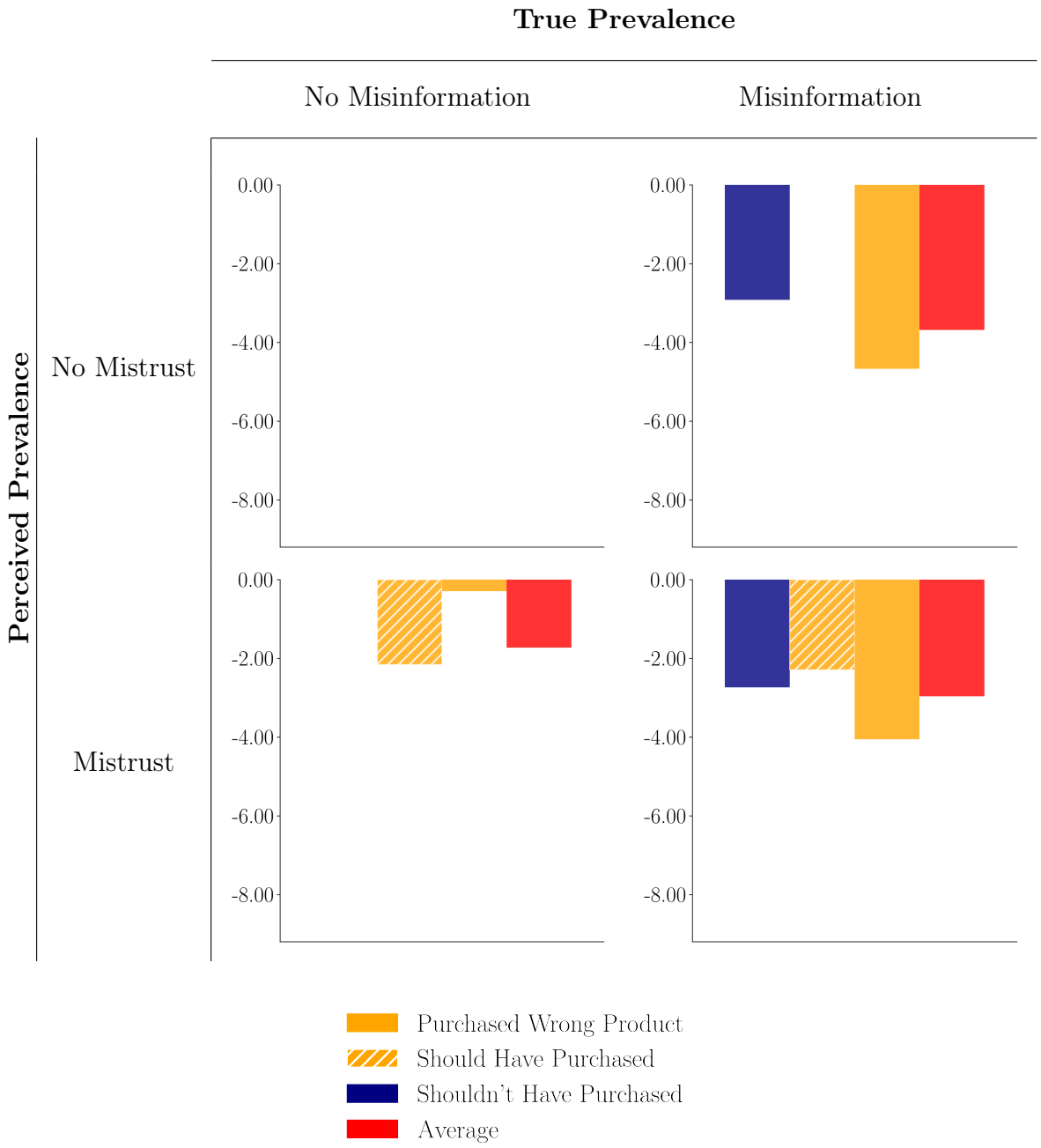
Figure C.2 characterizes the "mistakes" made by consumers when facing misinformation, mistrust, or both. In all cases, we define a mistake to be a purchasing decision that differs from what the consumer would have made as a "sophisticate" who trusts organic reviews and purchases only based on organic reviews. Not all of these mistakes are equally harmful to the consumer. Figure C.3 details the average harms for each type of mistake, as well as for the average mistake under each counterfactual.

Figure C.2: Mistakes Under Misinformation and Mistrust



Note: Figure tabulates the number of consumers who make each type of mistakes under combinations of misinformation and mistrust. Note that mistrust inflates the perceived quality of a small number of products, resulting in a very small number of cases where mistrust induces a consumer to purchase when it would have been optimal not to purchase anything. (See footnote 17 for details.)

Figure C.3: Welfare Harms by Type of Mistake



Note: Figure reports the average welfare loss in dollars per mistake made in each counterfactual. Mistakes are defined as making a different choice than a sophisticate that purchases based only on organic reviews.

C.5.4 Partial Deletion of Fake Reviews

Figure 12 depicts how consumer welfare (panel a) and platform profits (panel b) change with the partial deletion of fake reviews from the platform. We consider two scenarios. In one, a fraction of fake reviews are deleted randomly. In the second, the deleted fake reviews are selected to maximize consumer welfare according to a greedy algorithm. Specifically, under this algorithm, reviews are deleted iteratively, where at each step, we evaluate each product with fake reviews remaining, and compute the gain in the market's consumer welfare from deleting the product's fake reviews, considering the dynamic effects of product rank and organic reviews. Note that we could theoretically achieve further optimality by (i) computing the optimal fraction of reviews to delete per product, and (ii) including the global effect of each deletion on mistrust in the optimization objective. Therefore, our current analysis likely understates the concavity of the gains curve.

To account for the increase in consumer trust that review deletion would gain, we adjust the level of mistrust with each deletion such that consumer beliefs at the end of the deletion process are consistent with the scenario with no Fake Reviews. This entails, with every deletion, scaling down commensurately the probability placed on each product being a manipulator. Additionally, as manipulators join the ranks of honest products, we re-compute the prior distribution of qualities within honest products using a linear combination of their original priors and the priors on manipulator qualities.

C.5.5 Increasing Organic Reviews

In this exercise, we seek to answer whether increasing the prevalence of organic reviews can mitigate welfare harms from fake reviews. We repeat the counterfactual simulations with increased organic reviews at magnitudes ranging from zero to 400%, that is, from the factual scenario to one with five times the number of organic reviews. The proportion of good reviews among the extra reviews is governed by true quality of the product, and fake reviews are kept fixed at their factual levels. We increase both the stock of organic reviews at the start of the observation period and the predicted organic reviews that arrive each week.[28] Increasing the stock of existing organic reviews is necessary to ensure that there is sufficient perturbation in the number of reviews, especially since there are products whose existing reviews dominate the review arrivals in the observation period. Thus, we are simulating a counterfactual world in which organic reviews have been increased to a higher level in the long run. As such, we also assume that consumers adjust their beliefs with increased organic reviews. For example, consumers who believe that half of the reviews of a given product are fake reviews would adjust this proportion to one-third when organic reviews are increased by 100%. Consumers do not adjust the perceived probability that each product is a fake review purchaser conditional on the reviews.

Figure 13a describes how increasing organic reviews shifts consumer's quality perceptions closer to the econometrician's estimate of true quality. This shift is reflected in a reduction of perceived quality among fake review purchasers and an increase among honest products.

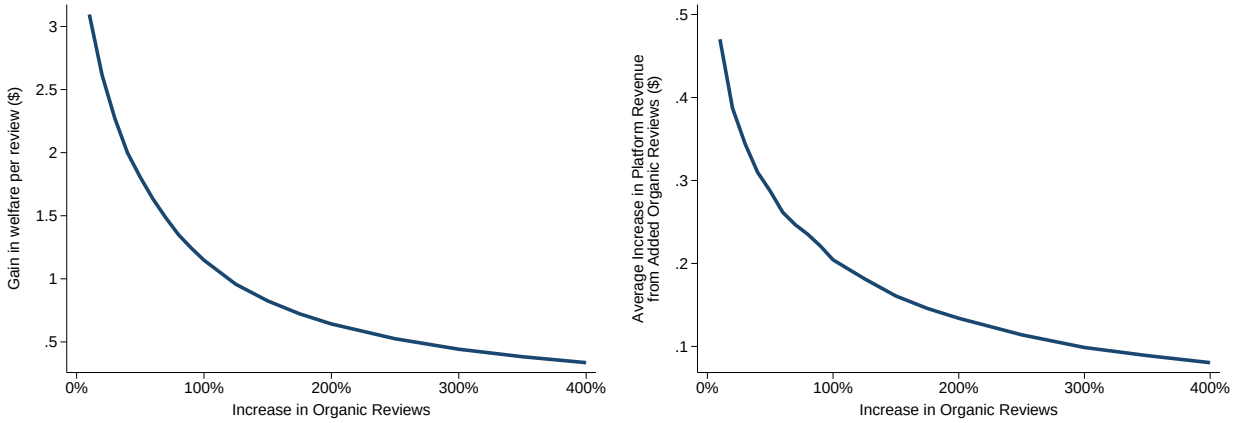
These shifts are composed of multiple mechanisms through which organic reviews affect quality perceptions. Firstly, for fake review purchasers, the increased organic reviews end up

[28] See Section C.1 for details on how organic review arrivals are predicted from lagged sales.

dominating the fake reviews, shifting the mean ratings closer to the true quality. Secondly, we assume that mistrust adjusts to an environment with more organic reviews, eventually approximating a world without fake reviews. Finally, for both fake review purchasers and honest products, increasing the number of reviews allows consumers to form more precise posterior beliefs of the qualities of individual products. This last channel in particular suggests that the benefits of increasing organic reviews potentially exceeds that of simply deleting fake reviews.

In figure 13b, we plot the change in consumer welfare under different levels of organic reviews. Here, the vertical axis is normalized by the total welfare harm that arise from fake reviews. We find that an increase in organic reviews by 48% would fully recover the lost welfare.

Figure C.4: Per-Review Gains from Increasing Organic Reviews



(a) Gain in Consumer Welfare per Review

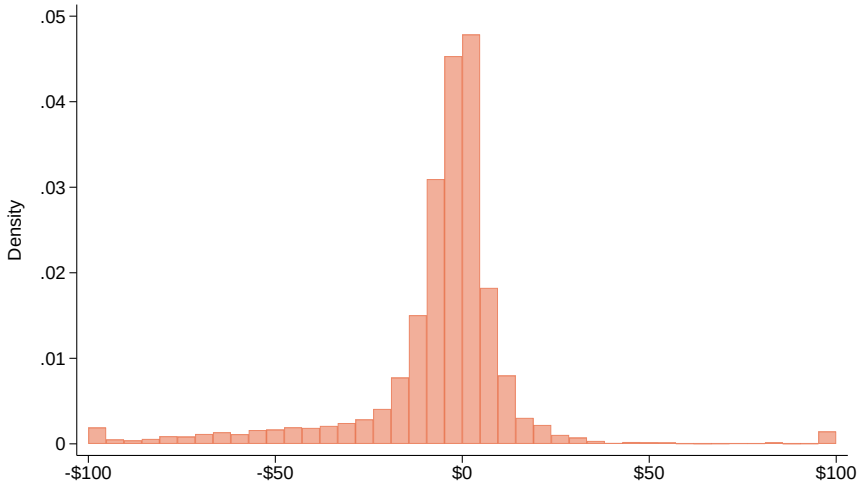
(b) Platform Profits

C.5.6 Costs and Benefits to Purchasing Fake Reviews

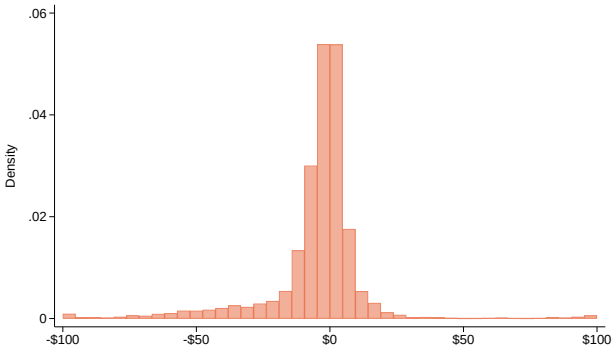
Costs to Purchasing a Fake Review In purchasing a fake review, the seller produces a unit at marginal cost mc that it sells through Amazon to the fake reviewer at price p. Amazon keeps $c^A p$ as its commission, and the seller receives $(1 - c^A)p$ from the sale. The seller then reimburses the reviewer via PayPal for the purchase price (p) and sales tax ($\tau^G p$). Sometimes the seller provides an additional small commission (c^R) of around \$5 to \$10. Finally, PayPal charges a fee of τ^{PP} times the payment amount. Therefore, the net cost of the transaction for the seller is:

$$\begin{aligned} c^{FR} =& mc + \left(1 + \tau^{PP}\right)\left(1 + \tau^G\right)p + c^R - \left(1 - c^A\right)p \\ =& mc + \left(\left(1 + \tau^{PP}\right)\left(1 + \tau^G\right) - \left(1 - c^A\right)\right)p + c^R \\ =& mc + \left(\tau^{PP} + \tau^G + \tau^{PP}\tau^G + c^A\right)p + c^R \end{aligned}$$

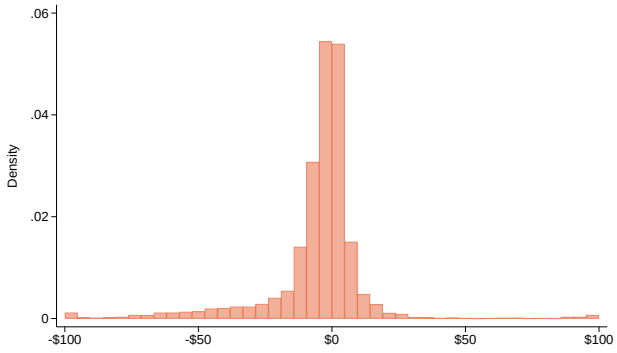
Figure C.5: Comparing Manipulators' Costs to their Honest Competitors



(a) Relative Cost of Purchasing Fake Reviews



(b) Relative Marginal Costs



(c) Relative Prices

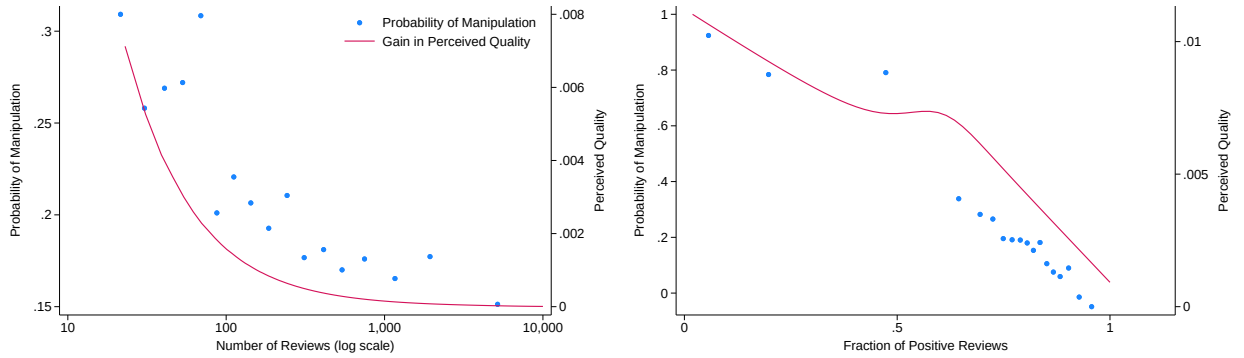
Note: These histograms show the difference between Fake Review Purchasers and the mean levels of the three variables within their markets. The differences are winsorized at −$100 and $100.

We follow He et al. (2022b) in assuming $\tau^G = 6.56\%$ (the average state and local sales tax), $\tau^{PP} = 2.9\%$, $c^A = 15\%$, and $c^R = 0$. Under these assumptions, the cost of purchasing a fake review is approximately $mc + .25p$. This places greater weight on marginal cost than price and suggests that, holding other factors fixed, high-margin products with low marginal costs will find purchasing fake reviews particularly attractive. This may explain the prevalence of fake reviews in categories such as Beauty & Personal Care.

Figure C.5 contrasts the costs of purchasing a fake review for empirical manipulators and their honest competitors. We find that in the majority of cases, purchasing fake reviews tends to be cheaper for manipulators than their honest competitors (panel a). This is attributable to the fact that manipulators tend to both have lower marginal costs (panel b) and lower prices (panel c) than their competitors.

It is important to emphasize that there are two important costs we do not capture in this analysis. The first is the risk of enforcement action by Amazon or regulators. In addition to removing the fake reviews, Amazon may choose to de-list a seller's product or even deactivate the sellers account if they are identified to be purchasing fake reviews. Regulators may go further in imposing sanctions such as fines. The second costs are the psychological or moral costs of defrauding consumers. These may be substantial and are impossible to observe directly.

Figure C.6: Increase in Perceived Quality from Purchasing a Fake Review



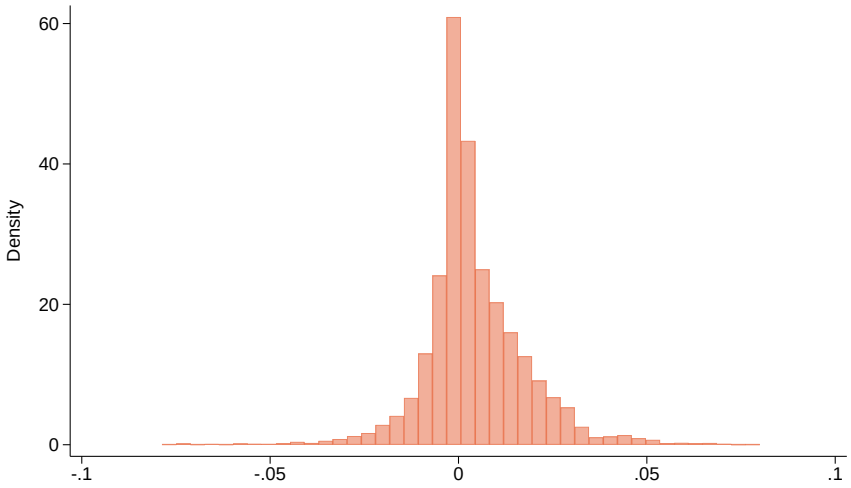
(a) Number of Organic Reviews

(b) Fraction of Positive Organic Reviews

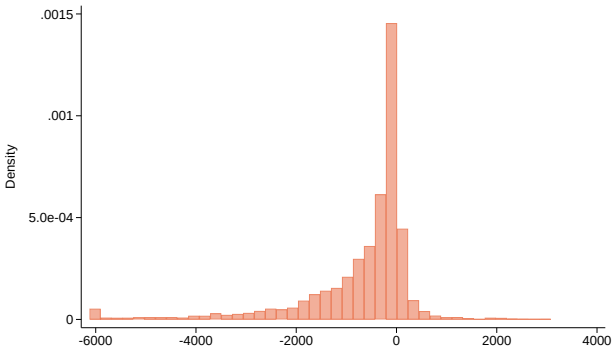
Note: Panel (a) presents the relationship between a product's number of organic reviews and the increase in perceived quality it would obtain by purchasing a fake review. The relationship is presented holding fixed the fraction of good reviews at the median of 87%. We overlay this with a binscatter of the empirical relationship between purchasing fake reviews and number of organic reviews, controlling for the share of reviews that are positive. Panel (b) presents the relationship between the fraction of a product organic reviews that are positive and the increase in perceived quality it would obtain by purchasing a fake review. The relationship is presented holding fixed the number of organic reviews at the median of 52 for fake review purchasers. We overlay this with a binscatter of the empirical relationship between purchasing fake reviews and fraction of positive reviews, controlling for logarithm of the number of reviews.

Benefits from Purchasing a Fake Review The benefits of purchasing a fake review accrue through misinformation inflating consumers' perception of the product. Consider product j at time t with N^o_{jt} organic reviews of which N^{o+}_{jt} are positive. Consumers at time t expect the quality of j to be $\mathbb{E}[q_j | N^+_{jt} = N^{o+}_{jt}, N_{jt} = N^o_{jt}]$ based on its rating. If the product

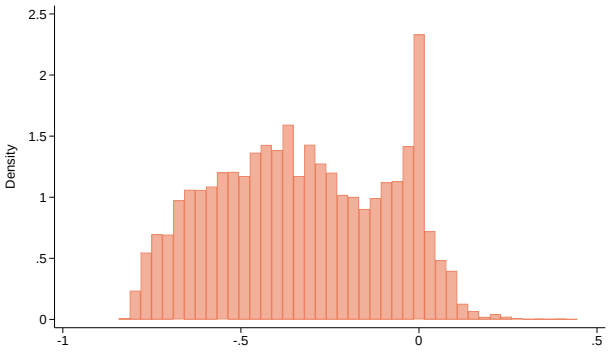
Figure C.7: Comparing Manipulators' Benefits to their Honest Competitors



(a) Relative Increase in Perceived Quality



(b) Relative Number of Organic Reviews



(c) Relative Review Positivity

Note: These histograms show the difference between Fake Review Purchasers and the mean levels of the three variables within their markets. Panel (b) has outliers winsorized at the 0.5^{th} and 99.5^{th} percentiles.

purchases a fake review at time t, consumers observe $N^o_{jt}+1$ ratings, of which $N^{o+}_{jt}+1$ are positive. Given this, consumers expect the quality of j to be $\mathbb{E}[q_j|N^+_{jt}=N^{o+}_{jt}+1, N_{jt}=N^o_{jt}+1]$.

Figure C.6 shows how $\mathbb{E}[q_j|N^+_{jt}=N^{o+}_{jt}+1, N_{jt}=N^o_{jt}+1]-\mathbb{E}[q_j|N^+_{jt}=N^{o+}_{jt}, N_{jt}=N^o_{jt}]$—i.e., increase in perceived quality from purchasing a fake review—varies with both N^o_{jt} and $\frac{N^{o+}_{jt}}{N^o_{jt}}$. Panel (a) indicates that consumers' perception of quality increases most substantially when the review is purchased for a product with few organic reviews. Panel (b) suggests products with worse organic reviews tend to increase their perceived quality more when purchasing fake reviews. Consistent with these, we find both that products with fewer reviews and with worse organic reviews are more likely to purchase fake reviews.

Figure C.5 compares these benefits of greater perceived quality for empirical manipulators to their honest competitors. In most cases, the products that we observe purchasing fake reviews are expected to improve their perceived quality more than their honest competitors would if they were to purchase fake reviews instead (panel a). This largely because the products we observe purchasing fake reviews tend to have both fewer organic reviews (panel b) and lower average ratings from organic reviews (panel c).

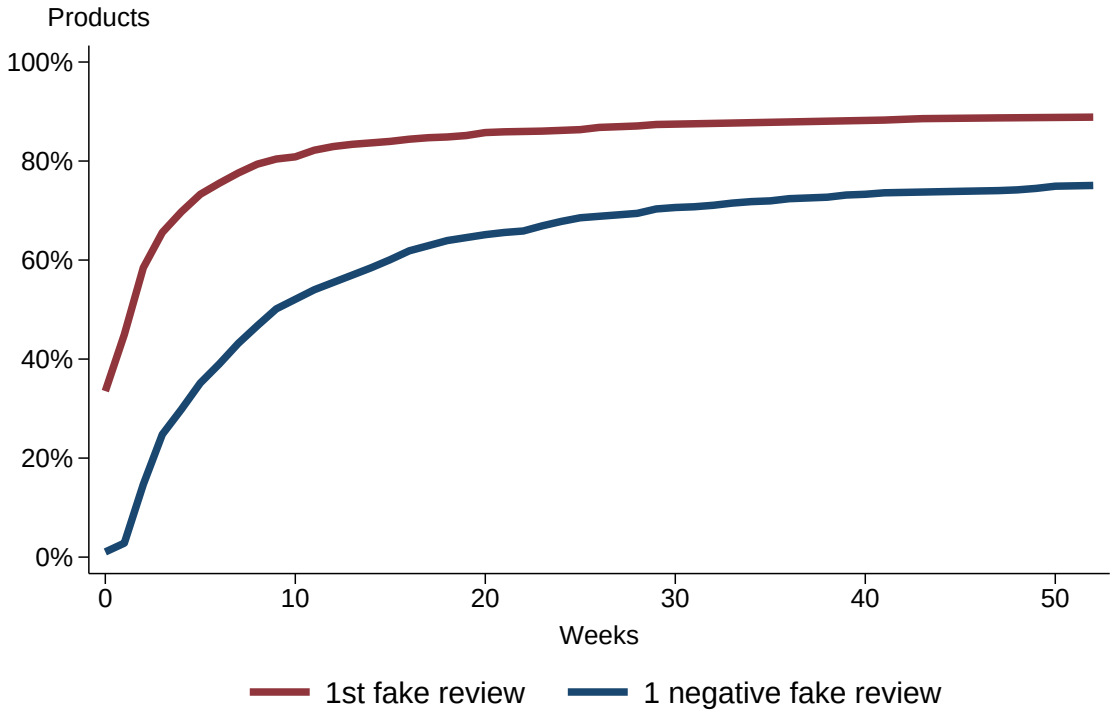
Note that when a product purchases fake reviews, it translates to demand benefits that accrue over multiple periods. The increase in j's period profits for time t is straightforward to compute: the increase in perceived quality that product j experiences at t typically results in an outward shift in demand for j and a corresponding static price increase from t. However, the dynamic changes j's expected profits in future $t'>t$ are less straightforward. Most directly, the fake review that j purchased t will continue to appear in both $N^+_{jt'}$ and $N_{jt'}$ for all $t'>t$. In addition, the increased sales of j at time t stochastically affects both search position and the arrival of new organic reviews for $t'>t$. We incorporate these dynamics into our computation of additional profits that j earns by simulating a distribution of paths that could occur. See C.1, C.2, and C.3 for details.

C.5.7 Negative Fake Reviews

In this section, we compare purchasing a fake negative review for one's competitor to purchasing a positive fake review for oneself. For each Fake Review Purchaser, we first find its closest competitor as determined by the elasticity of demand with respect to the competitor's expected quality. We then simulate the market equilibrium under the counterfactual in which the competitor receives an additional negative review. We compare the additional profit earned by the negative fake review purchaser to the cost of purchasing the negative fake review. Note that the cost of purchasing a negative fake review on a competitor's product is considerably higher than the cost of purchasing a positive review on one's own product. This is because when purchasing a negative review of a competitor, the $(1-c^A)p$ proceeds of the sale after Amazon takes its commissions go to the competitor. When purchasing a positive fake review for one's own product, these proceeds are returned to the fake review purchaser.

The median net cost of a negative fake review for a close competitor is \$40, which is much greater than the median net cost of a positive fake review of \$34. Additionally, the benefits of a negative fake review tend to be much lower since diverted consumers do not fully shift

Figure C.8: Weeks to Break Even from Purchasing One Negative Review for a Close Competitor



to purchasing the product being sold by the fake review purchaser. The median additional profits accrued over 4-weeks after purchasing a negative fake review is \$1.10 compared to \$84 for a positive fake review. Figure C.8 shows that in general, it takes much longer for products to break even when purchasing negative fake review for competitors than when purchasing positive fake reviews for themselves.

C.5.8 Deleting Manipulators

In this section, we consider the counterfactual policy of deleting manipulators from the platform. We find this policy to be detrimental to consumer welfare in the aggregate. This is true even if we can delete a fraction of Fake Review Purchasers in an ordering that optimizes welfare, as shown in Figure C.9. Intuitively, the negative welfare effect arises because any improvement to average quality is outweighed by the price increase from reduced competition in equilibrium. When all Fake Review Purchasers are deleted, the Honest Products are able to set prices that are 0.46% higher and gain profits that are around 10% higher than the factual equilibrium.

We also consider the effect of Fake Review Purchasers exiting the market due to lost profits after a counterfactual policy that removes both misinformation and mistrust. We model Fake Review Purchasers exiting according to how much a full deletion policy impacts their profits, and find that exits have an unambiguously negative effect on consumer welfare. The mechanisms governing the equilibrium effects is similar to the deletion counterfactual

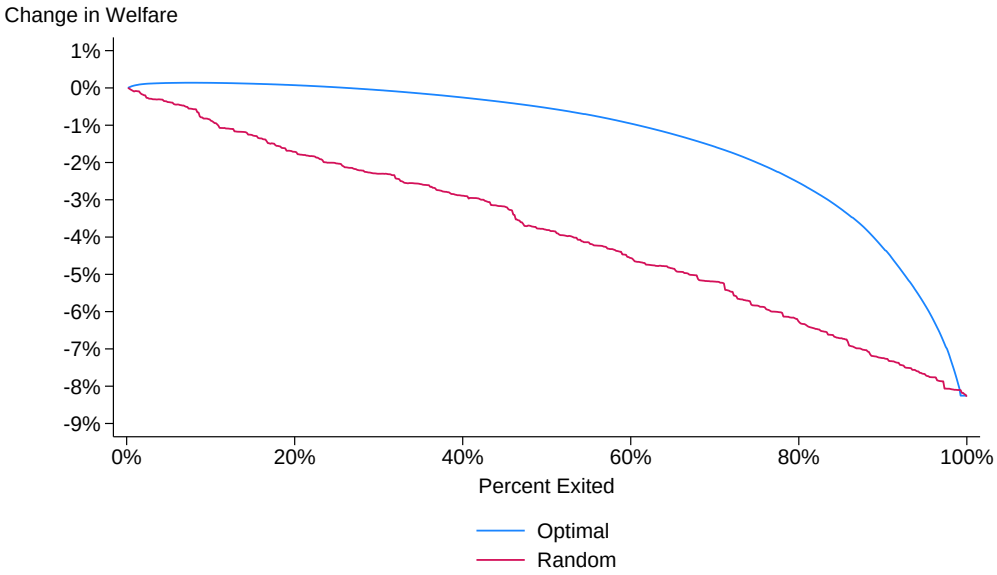


Figure C.9: Change in welfare from deletion of Fake Review Purchasers

above, as is the magnitude of the welfare changes.

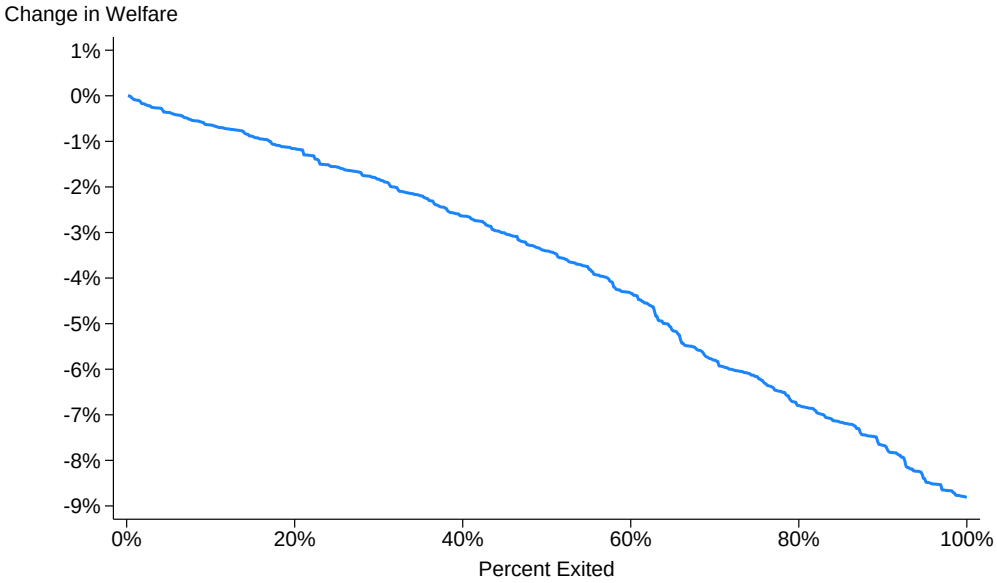


Figure C.10: Change in welfare from exit of Fake Review Purchasers