Misinformation and Mistrust: The Equilibrium Effects of Fake Reviews on Amazon.com

Ashvin Gandhi
UCLA & NBER

Brett Hollenbeck
UCLA

Zhijian Li
Northwestern

March, 2025 *

This paper investigates the impact of the widespread manipulation of reputation systems by sellers on two-sided online platforms. We focus on a relevant empirical setting: the use of fake product reviews on e-commerce platforms, which can affect consumer welfare via two channels. First, rating manipulation deceives consumers directly, shifting demand towards lower quality products through misinformation. Second, the presence of manipulation lowers trust that ratings reflect true quality, causing consumers to miss out on high quality products. Both misinformation and mistrust can also benefit consumers by increasing competitive pressure on products unable to differentiate through ratings. We study these effects by first modeling how consumers use ratings to form beliefs about quality and estimating consumers' relevant priors about the prevalence of fake reviews using an incentivized survey experiment. We incorporate this beliefs model into a structural model of the Amazon marketplace that we estimate using a large and novel dataset centering on products observed buying fake reviews. We then use counterfactual policy simulations to evaluate how fake reviews impact the marketplace through misinformation and mistrust and explore the implications for platforms and regulators.

*We thank Joel Waldfogel, Sherry He, Jessica Fong, Ginger Jin, Eddie Ning, Jinzhao Du, Ben Vatter, Yeşim Orhun, Andrey Simonov, Tin Cheuk Leung, and Jesse Shapiro 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, and the Spring 2025 NBER Industrial Organization meeting.

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 solve the asymmetric information problem 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 marketplace success and seller outcomes, 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 laws or other measures in response (FTC, 2019; CMA, 2020).

In this paper, we study the implications of rating manipulation for sellers, consumers, and platforms using the setting of the Amazon 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 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 change in beliefs may impede efficient search by lessening the ability of the ratings system to solve the asymmetric information problem. As a result, consumers may make worse purchasing decisions than if they could fully trust product ratings. At the same time, if mistrust in ratings makes high-quality products less able

to differentiate themselves from low-quality products, this may benefit consumers through increased price competition.

The relative magnitude 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 using fake reviews.

To measure fake review activity, we follow He et al. (2022b) in using a novel 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, including 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 designed to elicit these beliefs in the population of Amazon shoppers. We also leverage our knowledge of which products purchase reviews to determine the extent to which consumers can detect ratings manipulation. We 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. Importantly, because consumers' utility incorporates their expectations of quality rather than ratings directly, the same ratings can yield differing demand depending on consumers' beliefs about the presence of fake reviews. In particular, this lets us simulate how demand would change not only under different observed ratings but also under different consumer perceptions about

[1] A third channel by which seller manipulation of ratings may impact consumers is through dynamic effects, namely the extent to which paying for reviews lowers 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.

the prevalence of fake reviews.

To evaluate the impact of fake reviews, we consider a series of counterfactual policy analyses 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 if 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 consumers are harmed on net by ratings manipulation. The net loss in consumer welfare is around \$0.11, which is a loss of 0.4% of the median product purchase price. Competitive responses are also meaningful. The presence of fake reviews allows the median fake review purchaser to raise prices by \$0.19, and the median honest product lowers their prices by \$0.06. As expected, fake reviews benefit the revenues and profits of manipulators, while harming honest products. Additionally, we find that the financial benefit of purchasing fake reviews tends to be higher for manipulators than honest products, consistent with financial incentives driving substantial variation in the decision to manipulate ratings.

The overall effects mask important differences in the impacts of misinformation and mistrust. When isolated, misinformation causes a much larger decrease in consumer welfare as consumers are led to buy lower-quality products. By contrast, when mistrust persists in the absence of fake reviews, consumers are actually slightly better off due to increased price competition between sellers. When both effects are present, mistrust partially offsets the

welfare harms from misinformation.

Finally, we find that Amazon is not strongly incentivized to combat fake reviews. While platform revenue does increase with consumer trust, it also increases with misinformation. As a result, Amazon would slightly lose revenue if it eliminated fake reviews and all attendant misinformation and mistrust. The platform's losses would be particularly large if enforcement was done quietly without consumers' updating their beliefs. Instead, if it were feasible, the platform would most prefer to improve trust without necessarily substantially reducing misinformation.

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 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,

[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).

[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.

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.

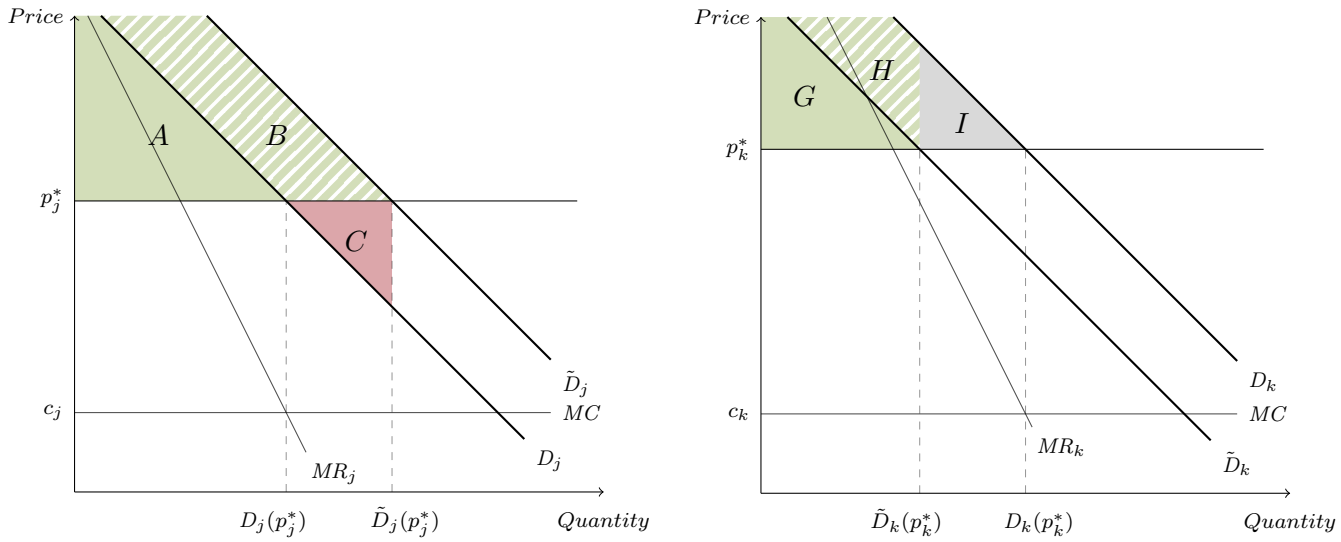
2 A Simple Model of Misinformation and Mistrust

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 of any individual product. 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

Figure 1: Effect of Misinformation (No Price Changes)



(a) Fake Review Purchaser

(b) Non-Purchaser

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 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

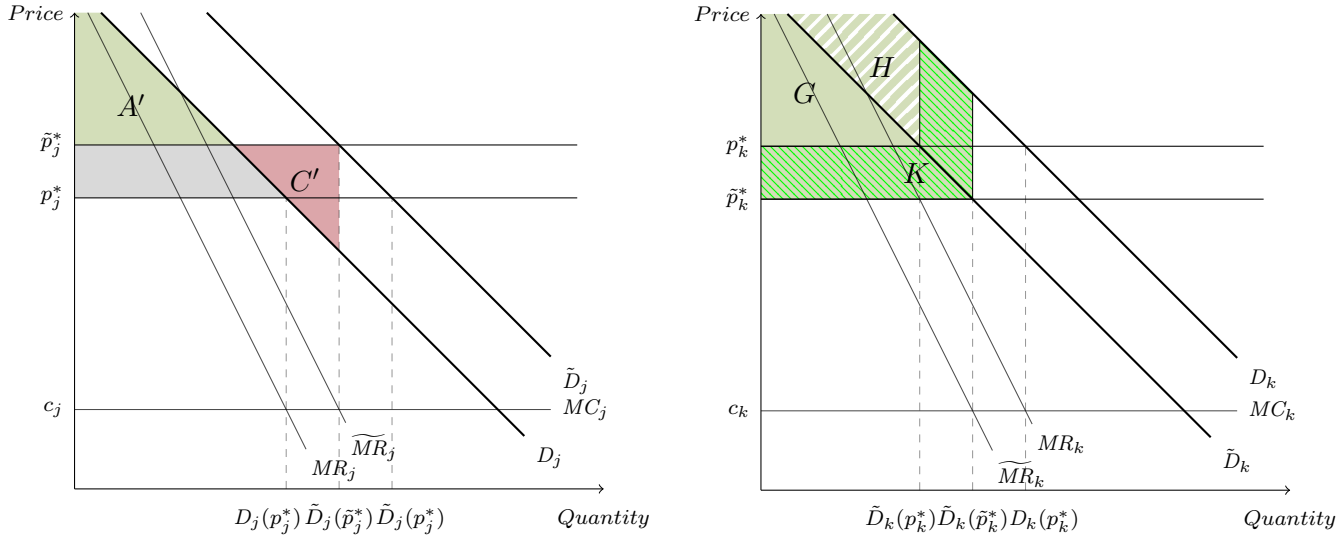
demand curve for product j and shifts in the demand curve for competitor product k. Figure 1 shows the effect of these demand shifts when holding prices fixed at the level that would have prevailed without fake reviews. The demand curves D_j and D_k are those that would occur absent fake reviews—i.e., 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 $\left(p_k^* - c_k\right)\left(D_k(p_k^*) - \tilde{D}_k(p_k^*)\right)$. 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.[5]

Figure 2: Competitive Responses to Misinformation



(a) Fake Review Purchaser

(b) Non-Purchaser

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 $\tilde{p}_j^*$.[6] 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 $A'-C'$.[7] 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^*

[5]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.

[6]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.

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

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

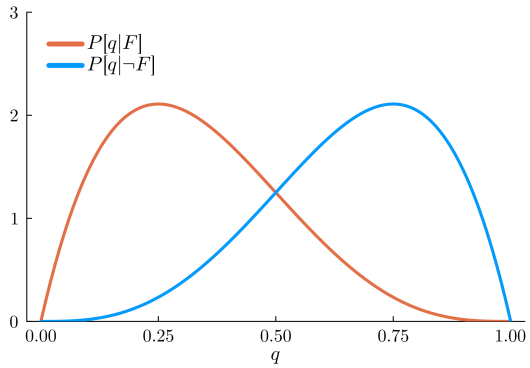
Thus far, we have modeled consumers as fully trusting reviews in order to isolate the effect of misinformation. Over time, however, consumers may learn from media, word of mouth, or personal experience that some products' ratings have been manipulated by fake reviews. In this section, we explore the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

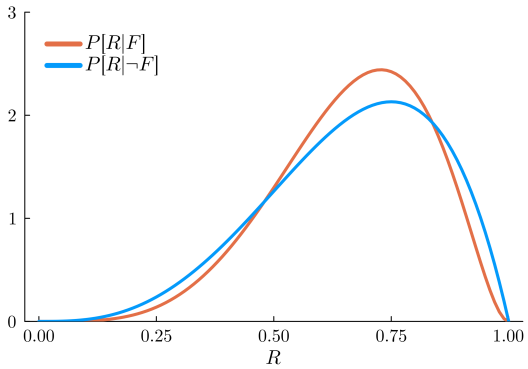
$E[q|R,F]$ and weight $1-P(F|R)$ on R:

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

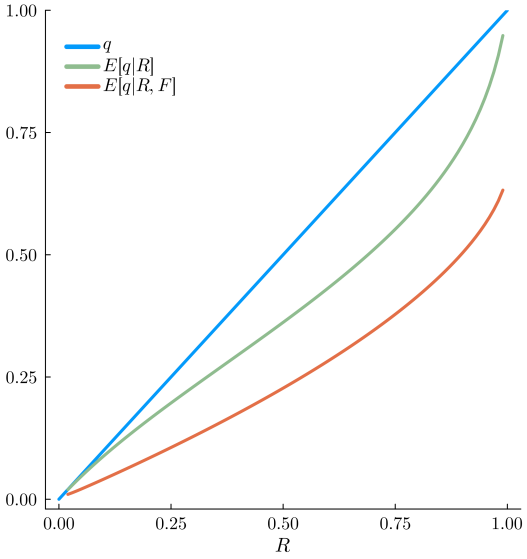
Figure 3: An Illustrative Example



(a) Quality



(b) Ratings



(c) Ratings

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 equation (1) characterizing 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

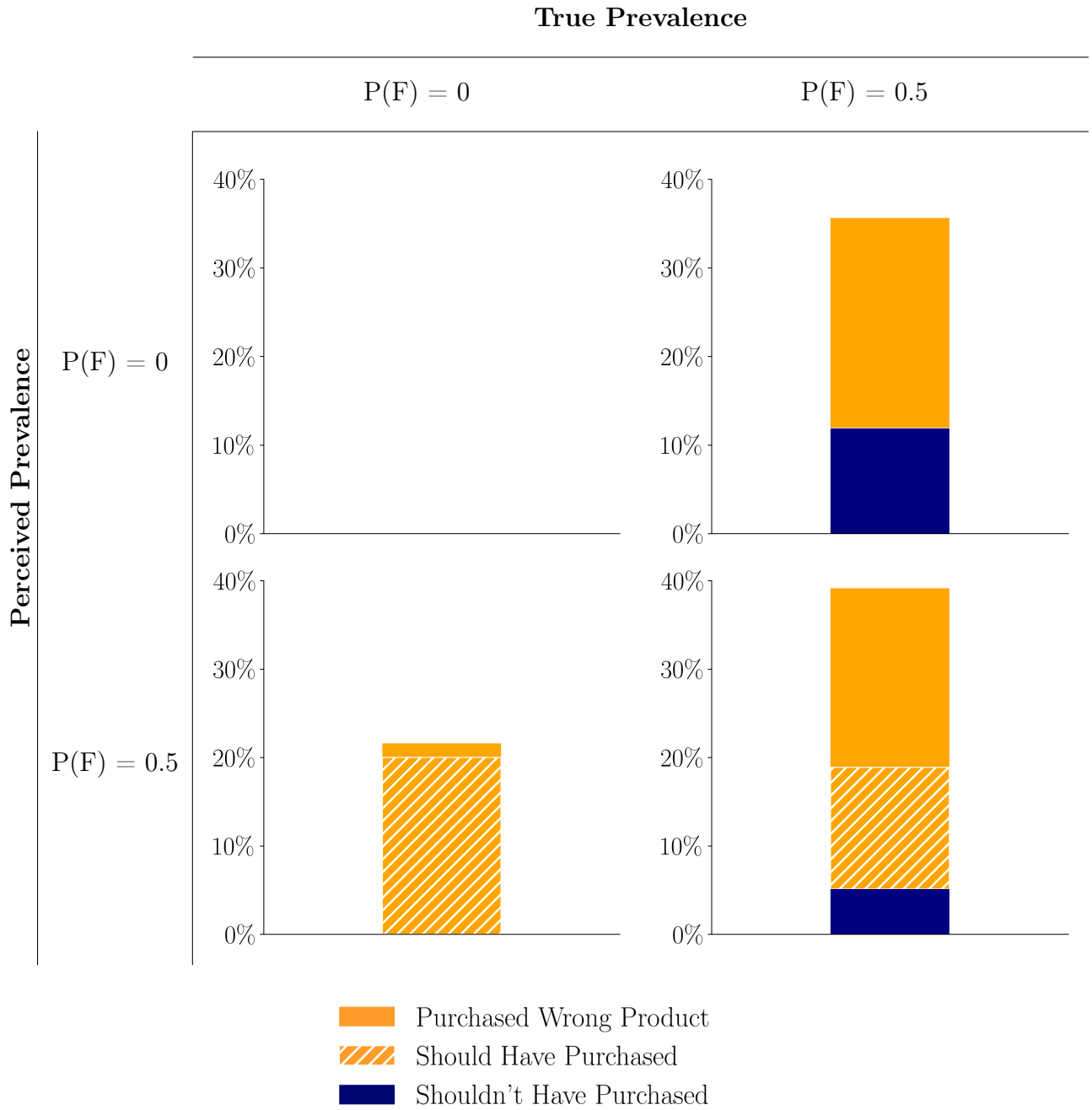
relationship between ratings and quality stems from the general prevalence of fake reviews and is not meaningfully shifted by the individual decisions of any single product.

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.

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.[8] 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.

[8]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.



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

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 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.[9] 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.[10] 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. 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.

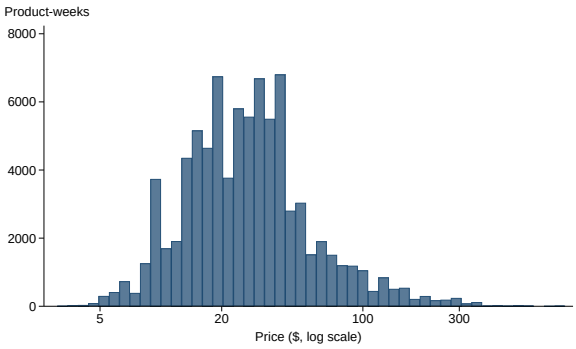
[9] In addition to relying on private groups, sellers often take additional steps to avoid detection by Amazon and authorities. First, sellers often use brokers as intermediaries. Additionally, sellers typically post a unique photo of the product rather than linking to the product's page to make algorithmic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

[10] 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.

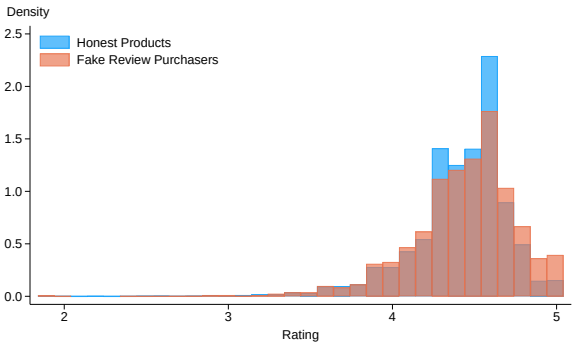
We obtain data on fake review activity by collecting information directly from the private Facebook groups where fake reviews are purchased. As scraping Facebook is technically infeasible, this required using a team of research assistants to monitor the top Facebook groups for these transactions and hand-collect data on a random sample of posting products and 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.

In addition, we 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 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. These close competitors are defined 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 is an FRP or an "Honest Product" (HP). Most products have average ratings between 4 and 5 stars, with the FRPs' ratings being inflated partially by fake reviews. Table 1 shows a full set of descriptive statistics on the focal fake review purchasers and their close competitor honest products. See Appendix B.1 for more information on the data.

Sales Data For the demand model, it is necessary to have a measure of product-level market shares. Amazon does not report sales data directly, instead reporting a measure called Best Seller Ranking or sales rank, which ranks products based on their rate of sales relative to other products in the same broad category. We collect sales rank for all products in our sample on a daily basis.

To calculate actual sales quantities, we exploit a feature of Amazon that makes product inventories observable for products with fewer than 1000 units in inventory. We collect this inventory data every 2 days for every focal fake review purchaser and competitor product. Following He and Hollenbeck (2020), we use the changes over time in inventories to construct

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

a measure of daily quantities sold. For observations where this data is not available, we estimate a model relating sales to sales rank that 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.

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 Bayesian consumers observe the number of positive and negative reviews for each product and infer the product's quality under the assumption that reviews are independent and the probability that a given review is positive increases with the product's quality and when the seller purchases fake 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. We estimate these in Section 4.2. The second is consumers' perceptions about the prevalence of fake reviews, 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 distributed binomial $B(N, q)$. 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. Therefore, we map Amazon ratings onto our binary framework by modeling consumers as viewing reviews as being entirely one and five stars.[11]

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 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 split of N reviews between N^+ positive and N^- negative reviews is binomial $B(N, p_{Fq})$:

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

[11] 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 $\frac{5N^+ + 1N^-}{N} \approx \bar{r}$.

Given this, a Bayesian consumer's posterior belief about the quality of a product with N^+ positive and N^- negative ratings 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 $P(N^+|q, N, F)$, the probability of receiving N^+ positive reviews out of N reviews conditional on the product's quality and whether the seller purchases fake reviews. This term is simply the binomial from (3). 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 $P\left(F \mid N^+, N\right)$, the consumer's perceived probability that a seller whose product has N^+ positive reviews out of N reviews is purchasing fake reviews. The last is θ^F, the consumer's perceived fraction of reviews that are fake for products that purchase fake reviews. These final two specifically regard consumer's *perceptions* on the prevalence of fake reviews, which need not align with the true prevalence. Therefore, we estimate consumers' perceptions of prevalence through a survey experiment described in Section 4.3.

We use the posterior from equation (4) to compute the expected quality the consumer anticipates after viewing a product's rating:

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

This expectation is how consumers' beliefs ultimately factor into the indirect utility function used to characterize demand in Section 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 allows us to represent consumers' priors with an econometric estimate of the distributions of quality. The latter is that consumers implicitly assume $P(q \mid N, F) = P(q \mid F)$, which substantially reduces the dimensionality of the priors. 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. 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. In our primary specification, we let q be Beta distributed conditional on F. In other words, $\gamma = \{(\alpha_F, \beta_F)\}_F$ and $q|F \sim \text{Beta}(\alpha_F, \beta_F)$. See Appendix A.2 for additional details.

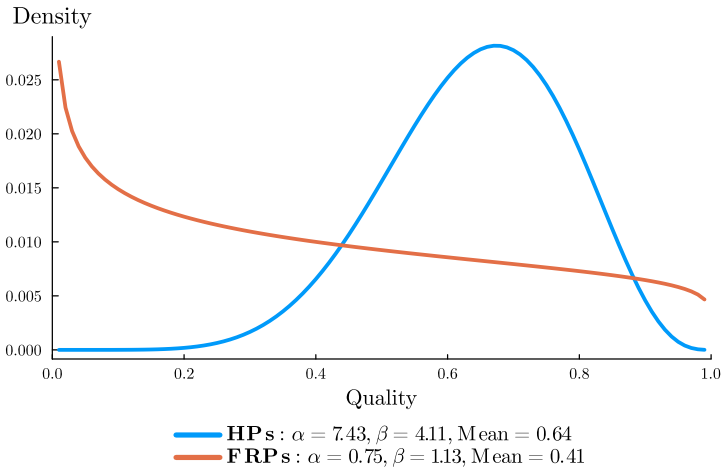
Using this, the likelihood of N^{o+} organic positive ratings out of N^o organic ratings is:

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

We estimate γ to be the maximizer of the log-likelihood of the organic reviews in the data:

$$\hat{\gamma} := \arg\max_{\gamma} \sum_j LL\left(N_j^{o+}, N_j^o; \gamma\right), \tag{7}$$

Figure 6: Estimated Priors



where j indexes products in the data.

The estimated distributions for $P(q|F;\hat{\gamma})$ are shown in Figure 6. The estimates imply that products purchasing fake reviews tend to be of substantially lower quality than products that do not.[12] The average quality of a product that purchases fake reviews is 0.41, while the average quality of a product that does not is 0.64.

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. The first is $P(F|N^+, N)$, the perceived probability that a product with a given rating purchases fake reviews. The second is consumers' perception of θ^F, the fraction of reviews that are fake for products that do purchase fake reviews. Since these represent consumers' perceptions, they are not directly observable in market data.

In this section, we describe an incentivized survey experiments that we use to characterize consumers' perceptions. The principal survey task that we describe in detail below aims to experimentally elicit consumers' perceptions of the prevalence of fake reviews and determine how beliefs vary with product characteristics, including whether a product actually purchased

[12]This finding is robust to alternative specifications, such as discretizing the unit interval and parameterizing $q|F$ to have a constant value on each sub-interval.

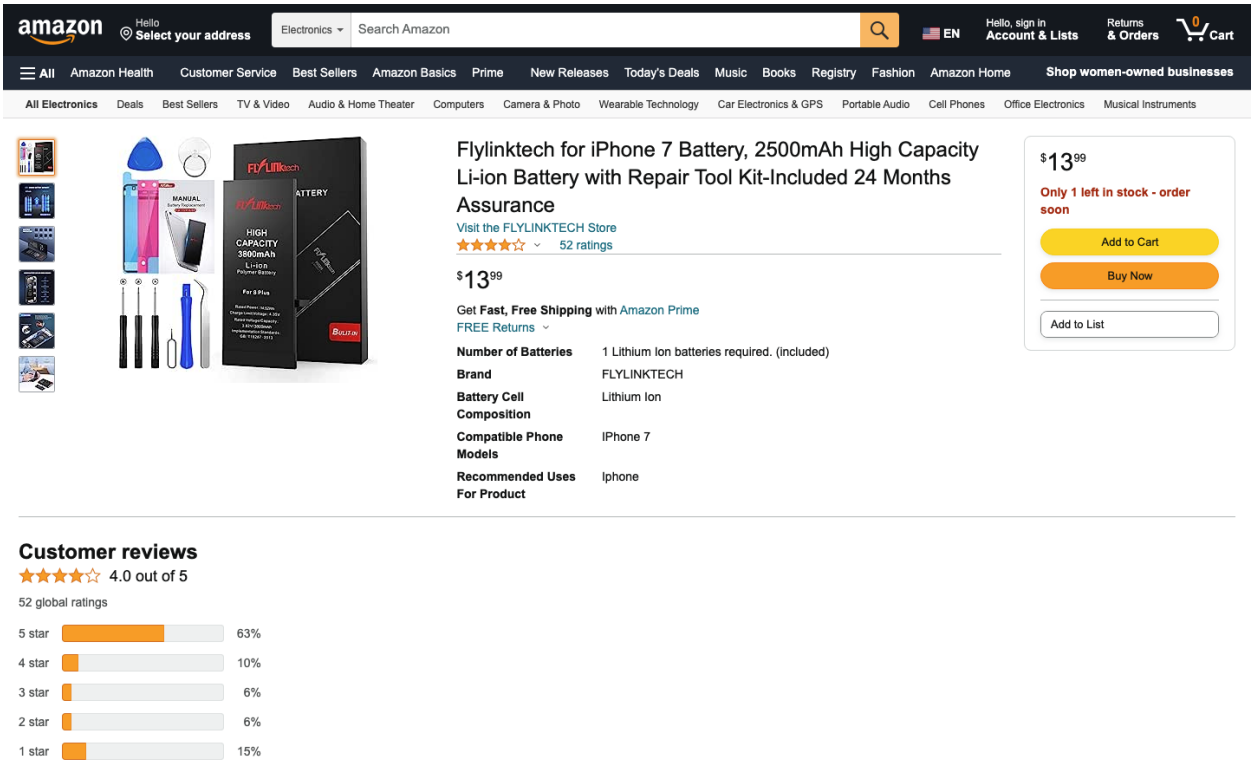
fake reviews. We use the fact that we observe the ground truth about which products use fake reviews to incentivize respondents by paying them more for selecting probabilities that better align with the truth. The survey implementation clearly communicates these payoffs to participants. We also leverage randomization assess how participants' responses vary with different product observables.

This main survey task takes place after the participant has completed a reading comprehension check, answered a host 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 also incorporate a host of best practices, including incorporating an initial reading comprehension check, a mid-survey attention check, and an additional comprehension check for the main component of the survey in order to screen out bots and humans who are not fully engaged with 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?"

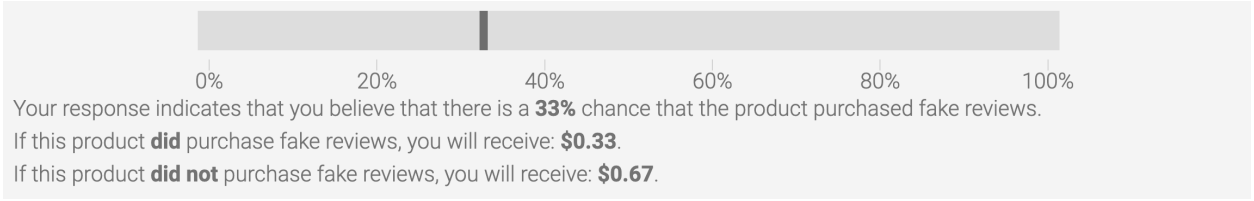
Main Prediction Task The principal survey task asks participants to view an Amazon product page and provide their best guess of the probability that that product uses fake reviews. The pages participants view are constructed using the HTML code from the pages of products in our sample. As illustrated in Figure 7 Panel (a), 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 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 engage the slider, it automatically updates and provides a full description of how their conditional payouts that updates with

Figure 7: Example Survey Page and Slider



(a) Example product page shown as in survey



(b) Slider showing respondents' beliefs and payoffs.

each selected probability, 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 how to use the slider, they must demonstrate use its use prior to starting the prediction exercises and are asked to select 100% in the middle of a sequence of predictions as an attention check.

Respondents repeat the prediction task for 10 different products, with two drawn randomly from each of the five Amazon categories the participant indicated they are most likely to purchase online. Thus, they can earn a maximum of $10 in addition to the base payment

of $1 that is paid to all respondents. The pool candidate products includes two randomly selected fake review purchasers from each of Amazon's 19 primary product categories and the single closest competitor product for each. For each question, respondents see a fake review purchaser with 32% probability and an honest product with 68% probability. For some respondents, randomly replace one product with an Amazon gift card as sanity check that they place approximately zero probability on the seller using fake reviews.

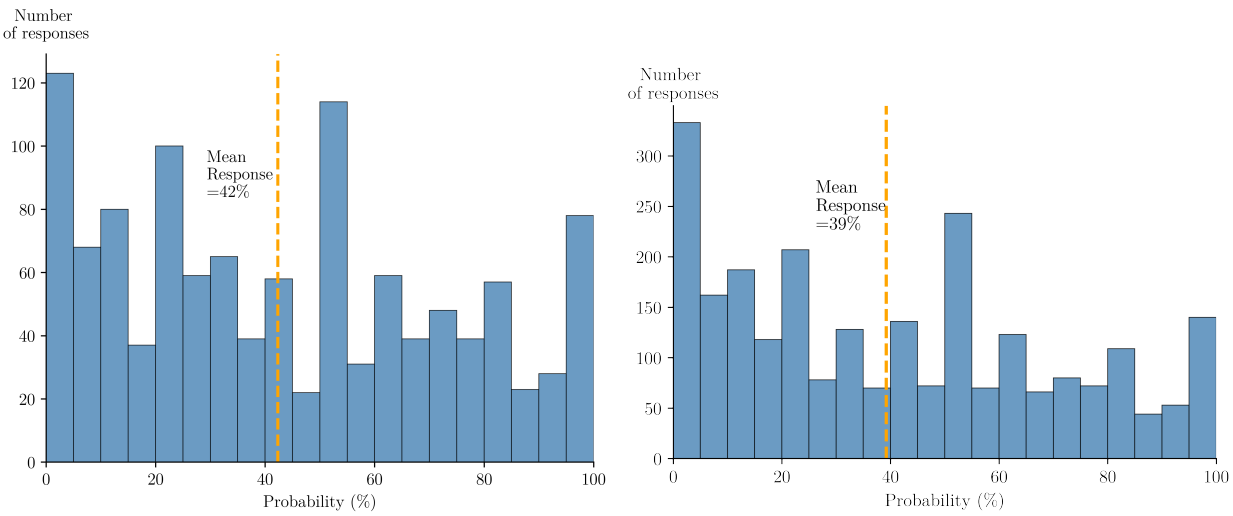
In addition to randomizing products, we also modify the HTML code to independently randomize the number of reviews and average rating that each participant sees. This randomization allows us to identify how ratings affect consumers' perceptions about the likelihood a product purchasing fake reviews—i.e., $P(F|N^+, N)$—using experimental variation that is decoupled the product itself. In implementation, we modify not only the number of reviews and the average rating, but also the histogram of ratings. For most participants, we show them the modal histogram from products with that rating and number of reviews. We show 40% of respondent more uniform or extreme ratings distributions (5th or 95th percentiles of rating variance) to test whether consumers use the extremity of a histogram as a signal of purchasing fake reviews. See Appendix B.3.6 for details.

Finally, for a randomly selected product, we ask respondents a follow-up question: "For this question, please assume that this product has purchased fake reviews. Guess the fraction of fake reviews among all its reviews." This is meant to elicit beliefs about θ^F, the proportion of fake reviews for products known to be using them. To incentivize respondents, we pay them based on how close they get to our measure of θ^F.

4.3.1 Survey Results

We ran the online survey experiment on Prolific in July 2023, which produced a sample of 401 qualified respondents, out of an initial 711, who passed the reading comprehension and attention checks. Appendix 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.

We also examine how these probabilities vary with the product's average rating and number of reviews. Figure B.3 shows this for each separately. There is a clear upward relation with rating (Panel a) but little apparent relationship with number of reviews (Panel b). Figure 9 shows how these vary together. Consumers appear generally suspicious of products with high ratings, especially those with fewer reviews. In contrast, products with few reviews and low ratings are perceived as particularly unlikely to have purchased fake reviews.

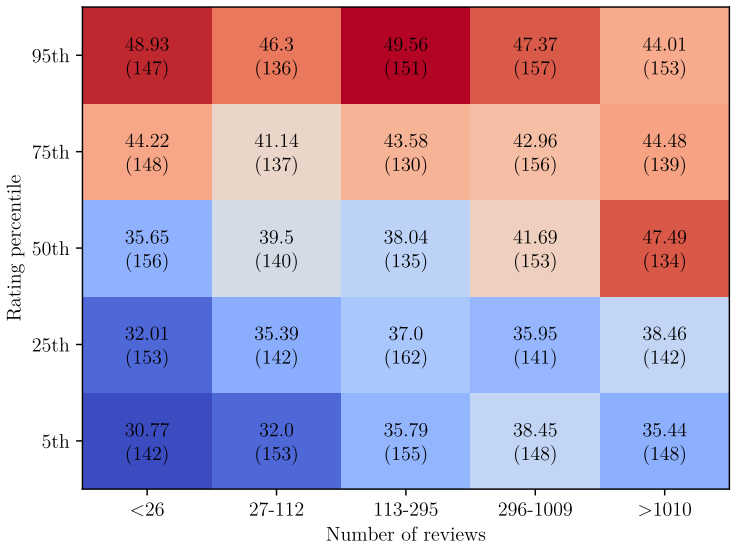


Figure 9: Beliefs by Rating and Number of Reviews

Importantly, 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 these values further, 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.4 shows participants' responses when asked to estimate θ^F. The mean and median response of 38% and 31% are both lower than we measure in Section 3.1. Therefore, we let $\theta^F = 38\%$ when modeling consumers' perceptions.

For additional analyses of responses by product category, differing histograms, 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 to view a sample of reviews during each of the prediction tasks.[13]

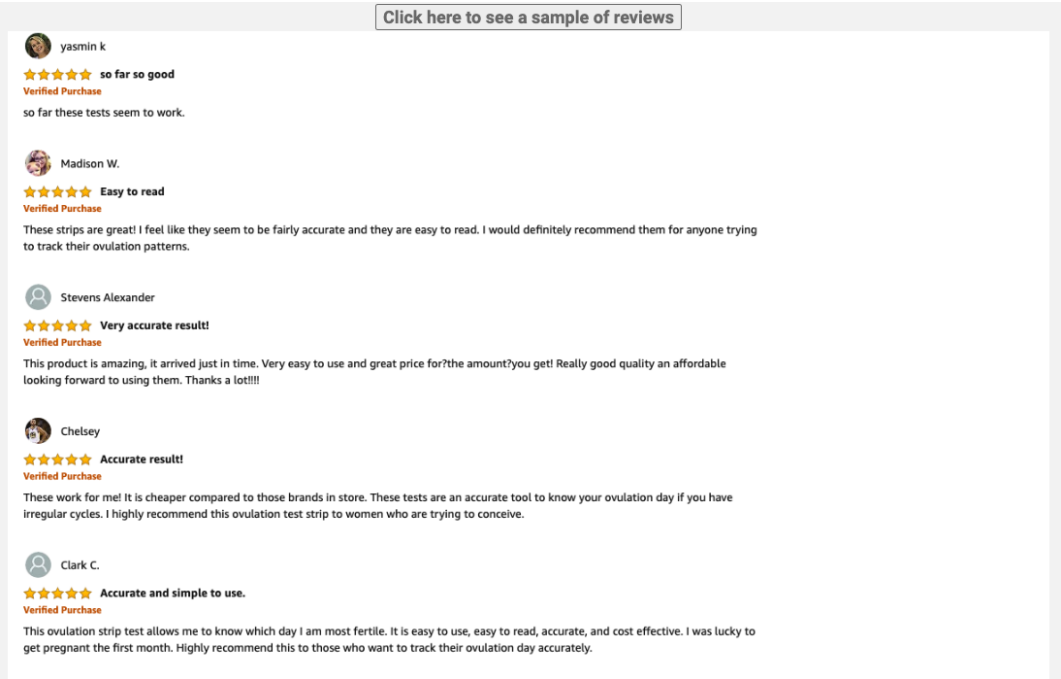


Figure 10: Example expanded reviews.

Survey participants clicked to view the review page 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 8 shows the distribution of responses after viewing the text of reviews. Reviewing the text of reviews 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.[14]

[13]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.

[14]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.

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 on their indirect utility function:

$$u_{ijt} = \beta_i \mathbb{E}\left[q_{jt}|N_{jt}^{+}, N_{jt}\right] - \alpha_i p_{jt} + \beta^X X_{jt} + \lambda_t + \zeta_j + \xi_{jt} + \epsilon_{ijt} \tag{8}$$

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. 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 error term of a trivial supply side.

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 simple supply specification we use only the product-level intercept. 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 the elasticity of demand with respect to expected product quality is fairly high at roughly 1.5. 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 -1.9 with a median of -

1.5. This suggests somewhat inelastic demand, consistent with prior estimates of Amazon product demand (Reimers and Waldfogel, 2017, 2021). We find a negative coefficient on the product age and a negative coefficient on the listing rank, which is consistent with greater demand for newer and better-ranked products.

Table 2: Results of Demand Estimation

Age	-0.036
	(0.018)
Listing Rank	-0.029
	(0.00096)
μ_{-p}	-2.8
	(0.065)
σ_{-p}	0.16
	(0.011)
μ_q	0.76
	(0.032)
σ_q	0.023
	(0.0074)
Product FEs	Yes
Week FEs	Yes
Optimal IVs	Yes
Median Own-Price Elast.	-1.4
Mean Own-Price Elast.	-1.9
Median Own-Quality Elast.	1.5
Mean Own-Quality Elast.	1.5
Observations	81,364

Notes: The random coefficients are parameterized as $\binom{\alpha_i}{\beta_i} \sim log\mathcal{N}(\mu, \Sigma)$ where $\mu = \binom{\mu_{-p}}{\mu_q}$ and $\Sigma = \begin{pmatrix} \sigma_{-p} & 0 \\ 0 & \sigma_q \end{pmatrix}$.

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. Implementing this analysis consists of several parts. First, we compare consumer beliefs about product quality, as well as prices and quantities sold and seller profits, between the factual world where fake

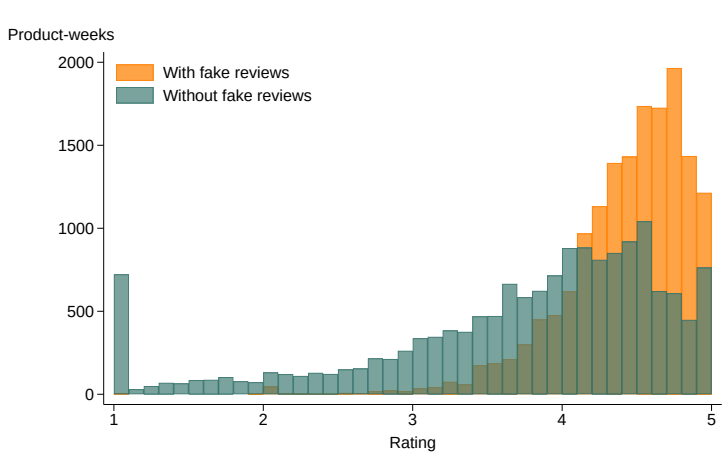
reviews are present and consumers are mistrustful of reviews to the counterfactual world in which no fake reviews are present and consumers are fully trusting of reviews. Second, to isolate the misinformation and mistrust channels we evaluate separate counterfactuals in which Amazon eliminates fake reviews but consumers remain mistrustful and in which consumer mistrust is eliminated but fake reviews remain. In each case, we consider separately the role of competition in these changes by holding prices fixed vs allowing firms to react by changing prices. Finally, we examine 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, with a simulated counterfactual without fake reviews or the attendant misinformation and mistrust. Note that 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.

Figure 11 uses our Section 3.1 estimates of fake reviews for each product in our data to recompute product ratings and consumer beliefs after removing fake reviews. Panel (a) shows that fake reviews dramatically inflate ratings of fake review purchasers by an average of 0.7 stars. Panel (b) shows how perceived product qualities (Equations 4 and 5) change with and without fake reviews. Fake reviews result in consumers typically overestimating the quality of fake review purchasers—particularly for products with poor organic ratings—and underestimating the quality of honest products. Appendix Figure C.1 decomposes this change in perception by misinformation (Panel a) and mistrust (Panel b). Absent mistrust, fake reviews simply misinform consumers by increasing the perceived quality of products that purchase fake reviews. Together, Figures C.1 and 11 imply that the level of consumer mistrust estimated from our survey noticeably lowers consumers' perceptions about product quality. Importantly, this mistrust affects honest products as well resulting in their quality being systematically underestimated.

Figure 11: Effect of Removing Fake Reviews on Product Ratings and Perceived Qualities



(a) Ratings of Fake Review Purchasers

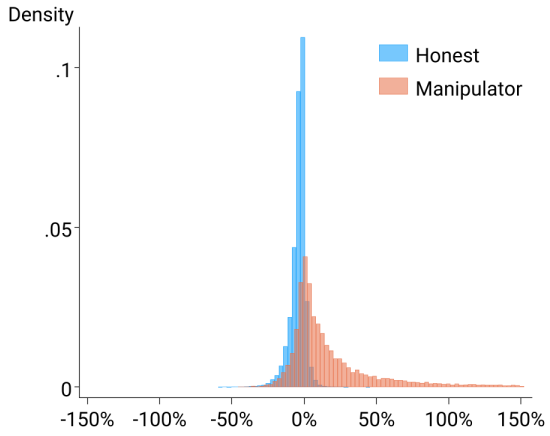
(b) Perceived Qualities

As perceptions of quality change with the inclusion or exclusion of fake reviews, demand changes as well. We study the implications of these changes by comparing market outcomes under the factual to the counterfactual without fake reviews creating either misinformation or mistrust. In simulating the market under alternative demand, we allow sellers to adjust their prices to reach a Bertrand Nash equilibrium.

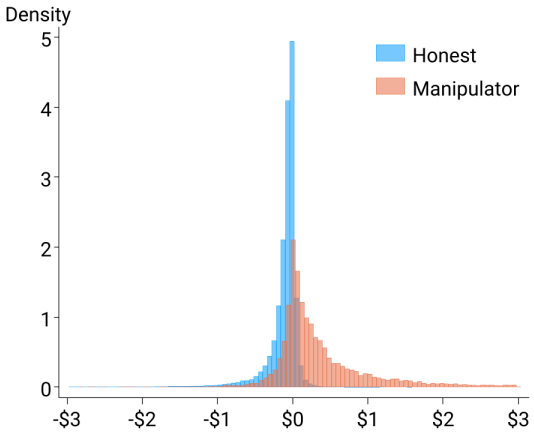
We simulate outcomes sequentially over time in order to incorporate path-dependence in search rankings, sales, and the frequency of organic reviews. Consider, for example, a product that sees its sales in period 1 reduced due to its inability to the absence of fake reviews in the counterfactual. Reduced sales in period 1 could reduce the number of new organic reviews, as well as harm the product's search rank in the following period. Both the change in organic reviews and search rank will affect demand in period 2.

We account for this by estimating hedonic timeseries models of both search rank and the arrival rate of new organic reviews. Our hedonic multinomial model of search position allows the previous two weeks of sales and reviews determine products' search rankings. Likewise, our hedonic Poisson model allows the previous two weeks of sales to determine the number of new organic reviews. Details and estimates for the hedonic search rank and organic review models are respectively reported in Appendix C.1.1 and C.1.2.

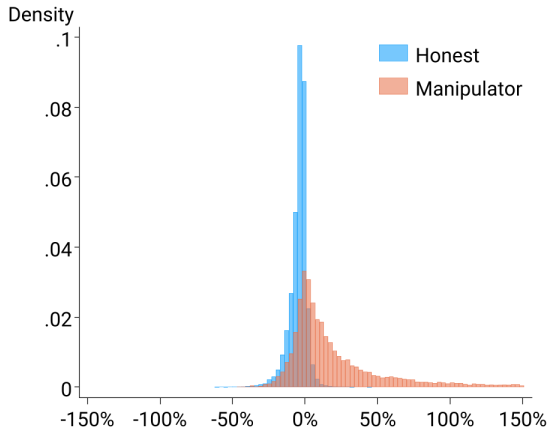
Figure 12: Counterfactual Quantities, Prices, Revenues, and Profits



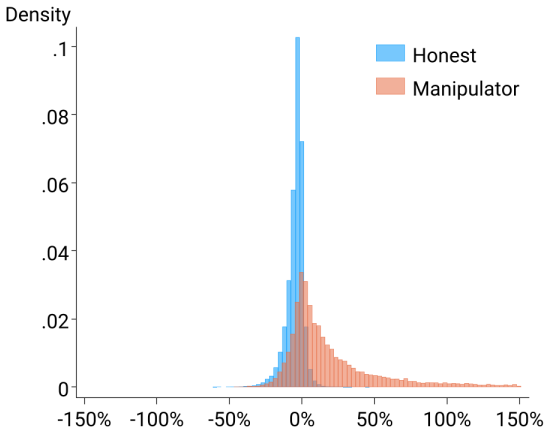
(a) Change in Quantities



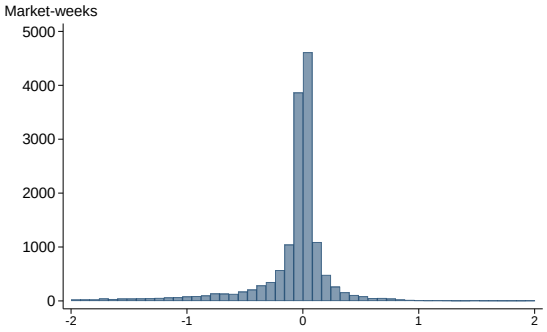
(b) Change in Prices



(c) Change in Revenues



(d) Change in Profits



(e) Change in Welfare

Figure 12 depicts our principal findings on the effect of fake reviews, which are also summarized in Table C.5. 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 18.3% more units, while the average honest products sells 3.5% less. Overall, fake reviews result in a reduction of 0.5% in units sold on the marketplace. Panel (b) indicates that the increase in demand allows manipulators to profitably increase their prices (median increase of \$0.19), while honest products lower their prices to compete with manipulators (median decrease of \$0.06).

These shifts in prices and quantities translate to substantially greater revenues (Panel c) and profits (Panel d) for manipulators at the expense of honest products. Fake reviews increase the total profitability of manipulators by a dramatic 14% while reducing the profits of honest products by 4.7%. The total revenue and profits for all products on the platform decrease by 1.1% and 0.3%, respectively. Since Amazon typically receives a fixed percent of sales revenue, fake reviews actually result in slightly less revenue for Amazon. This implies that Amazon does have some financial incentive to combat fake reviews. However, if enforcement is costly, Amazon may be better off simply allowing the level of misinformation and mistrust in our sample. Moreover, we show in Section 6.2 that Amazon benefits from improving trust and not from reducing misinformation. If doing enforcement in a way that is credible and improves trust is difficult or costly, it will further weaken Amazon's incentives.

There is substantial variation in the impact of fake reviews on consumers (Panel e) with the average effect being a harm of \$0.11.[15] Welfare changes have two principal drivers. The first are mistakes: consumers induced to purchase the manipulator's product due to misinformation or dissuaded from purchasing a preferable product due to mistrust are made

[15]See Appendix C.1.3 for details on how we compute consumer welfare under misinformation and mistrust.

worse off. We explore these further in the the next section. The second are price changes: consumers who still purchase honest products tend to be better off due to discounting, while those that purchase manipulators tend to be worse off because of price increases.

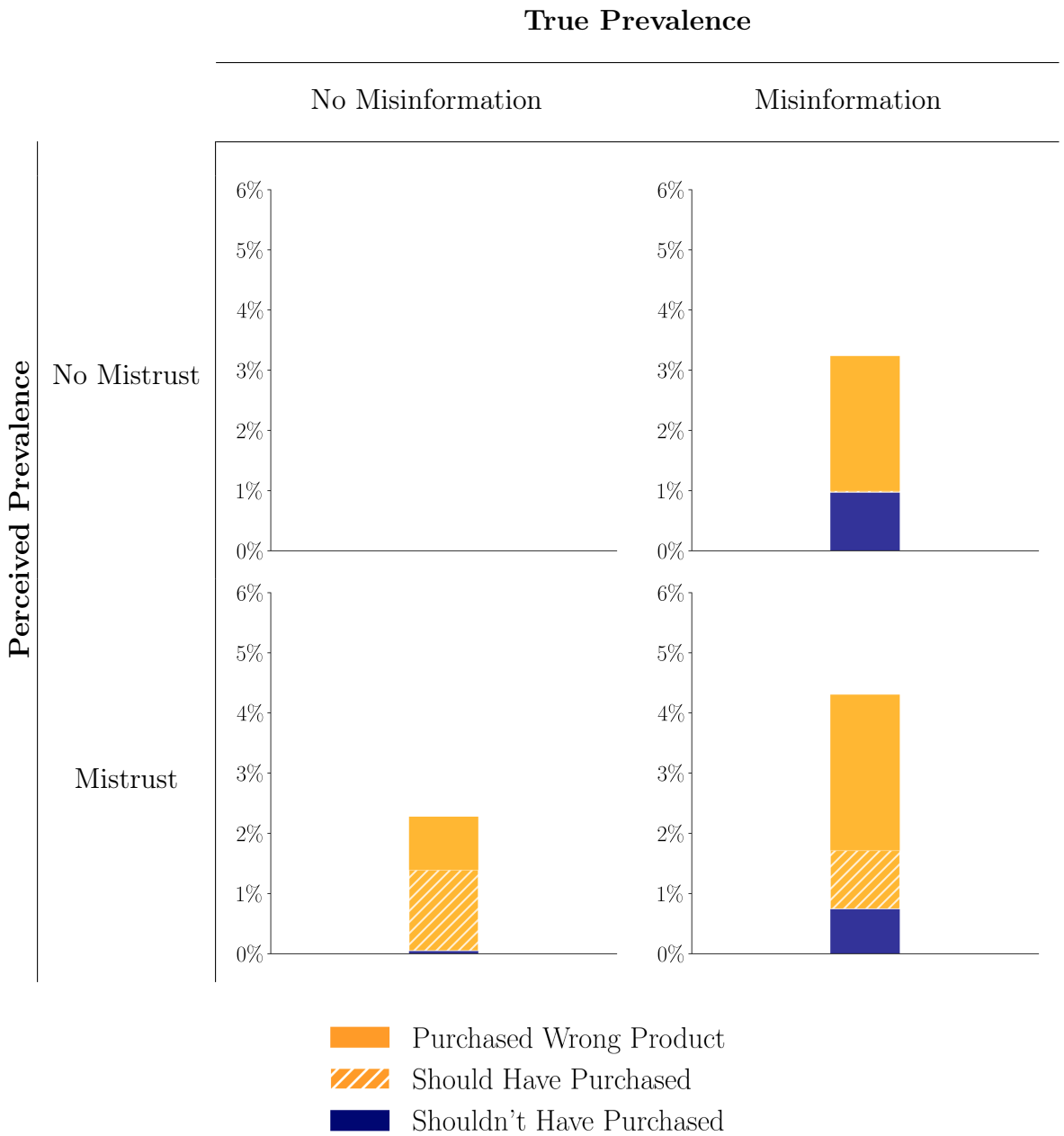
6.2 Isolating the Effects of Misinformation and Mistrust

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
Welfare ($)	365,689,662	363,482,194	363,297,846	365,544,835	366,193,535	363,549,049	363,930,515
Welfare change (%)	+0.00	-0.60	-0.65	-0.04	+0.14	-0.59	-0.48
FRP average prices ($)	30.89	30.89	31.57	30.89	30.78	30.89	31.47
HP average prices ($)	37.96	37.96	37.87	37.96	37.92	37.96	37.83
FRP sales (units)	1,588,863	1,987,557	1,921,590	1,532,692	1,540,026	1,937,184	1,879,868
HP sales (units)	9,872,529	9,613,339	9,668,638	9,730,095	9,735,115	9,472,060	9,529,524
FRP revenue ($)	49,077,636	60,150,023	60,037,435	47,188,502	47,187,572	58,703,492	58,544,840
HP revenue ($)	367,773,495	358,324,755	359,599,850	363,164,765	362,838,858	353,633,070	354,548,209
Platform revenue ($)	41,685,113	41,847,478	41,963,729	41,035,327	41,002,643	41,233,656	41,309,305
FRP profits ($)	34,124,501	41,602,756	41,880,040	32,945,910	32,926,567	40,584,878	40,821,504
HP profits ($)	191,289,163	186,034,234	186,348,557	188,570,709	188,346,763	183,359,531	183,393,347

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 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 also 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. Correspondingly, contrasting the last column with the first column summarizes the results from Figure 12. Additionally, Figure 13 depicts the composition of purchasing mistakes under each counterfactual.



Note: Figure tabulates the number of consumers who make each type of mistakes made under combinations of misinformation and mistrust. Here, we allow for re-pricing in equilibrium.

Figure 13: Mistakes Under Misinformation and Mistrust

Misinformation misleads consumers to substantially overestimate manipulators' quality (Figure C.1), especially for low-quality manipulators masquerading as highly rated products. This results in many consumers mistakenly purchasing manipulator products instead of honest products or the outside good (Figure 13). Without price adjustments, manipulators induce consumers to purchase 25.1% more units and earn 19.0% more profit, while honest products sell 2.6% fewer units and earn 3.6% less profit. On average, consumers experience \$0.14 (0.6%) of harms. When prices can adjust, manipulators raise their prices by an average of \$0.58 (11.3%), and honest products reduce theirs by an average of \$0.14 (0.6%). Adjusting prices increases profits for both manipulators and honest products. For consumers, the competitive responses slightly reduce the harms from misinformation. Allowing consumers to also mistrust reviews negates a fraction of the harms of misinformation.

Mistrust in isolation causes consumers to slightly underestimate the quality of all products and especially those with mediocre ratings (Figure C.1). However, in stark contrast to misinformation, mistrust does not make low-quality products with poor organic reviews appear more attractive than high-quality products with better organic reviews. Therefore, while mistrust does cause consumers to mistakenly purchase the wrong product or not purchase at all (Figure 13), these mistakes tend to be minor in that they shift consumers to alternative choices with only slightly lower utility (Figure C.3). Holding prices fixed, consumers' mistrust of ratings leads to 269018 mistakes in our sample, but these mistakes are sufficiently small that the average harm over all consumers in the sample is just \$0.01. Furthermore, when prices can adjust, mistrust actually slightly *benefits* consumers on average (\$0.03) because products must compete more aggressively on price when it is more difficult to differentiate quality through ratings.

These counterfactuals also shed light on the incentives of the platform, which receives a fixed share of revenue. Our estimates suggest the platform benefits from misinformation—which causes consumers to overestimate quality and spend more on the platform—and 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.[16] 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. Inconspicuously identifying and removing fake reviews is the most harmful policy to Amazon in the short-run. Our estimates suggest that 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 removing fake reviews to the extent required for credibility.

6.3 Incentives to Purchase Fake Reviews

In this section, we explore the financial incentives to purchasing fake reviews. 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:

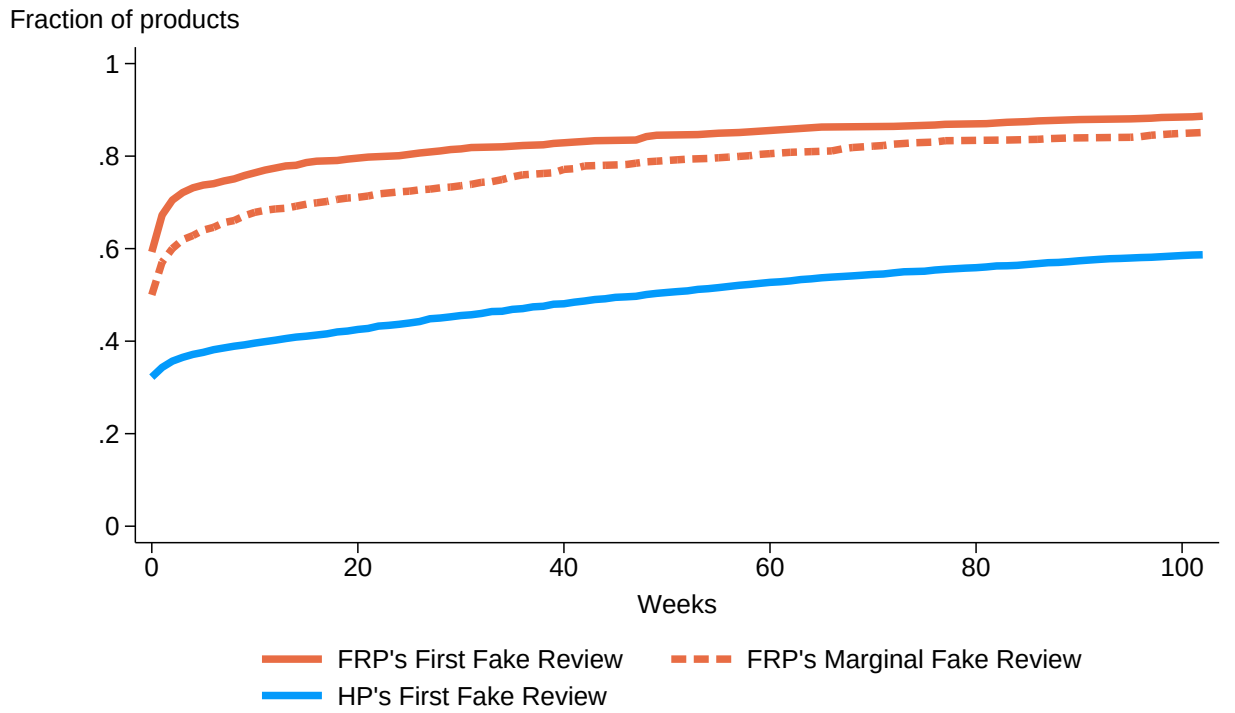
$$\begin{aligned} c^{FR} =& mc + \left(1+\tau^{PP}\right)\left(1+\tau^{G}\right)p + c^R - \left(1-c^A\right)p \\ & mc + \left(\left(1+\tau^{PP}\right)\left(1+\tau^{G}\right) - \left(1-c^A\right)\right)p + c^R \\ & mc + \left(\tau^{PP} + \tau^{G} + \tau^{PP}\tau^{G} + c^A\right)p + c^R \end{aligned}$$

[16] 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.

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.

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.

Figure 14: Weeks to Break Even on Purchasing a Fake Review



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 only

paid 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. Therefore, to assess the benefits of purchasing fake reviews relative to their costs, we simulate the length of time required for the additional profits generated from a fake review to exceed the cost of purchase.

Our findings suggest that purchasing fake reviews is often financially beneficial, especially for the products we do observe purchasing fake reviews. Figure 14 shows that the typical ratings manipulator is able to quickly break even. The majority do so on their first fake review—or even their marginal fake review—within a few quarters. Purchasing fake reviews appears to be much less attractive for honest products, who generally have better and more numerous organic ratings. Only about half of honest products would break even on purchasing a fake review after two years. That our model implies substantially greater return on investment from purchasing fake reviews for products that empirically do so than those that empirically don't suggests that financial incentives do drive substantial variation in the decision to manipulate ratings. However, that our model implies a non-trivial fraction of honest products could benefit from purchasing fake reviews suggests that unmodeled factors, such as moral costs and risk of enforcement actions, also play an important role in the decision to purchase fake reviews.

Given the large potential benefits to positively manipulating one's own rating, a natural question is whether firms could also benefit from purchasing negative fake reviews for competitors' products. We explore this in Appendix C.2.2 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 fake review purchaser does not capture all demanders diverted from its competitor. Finally, purchasing fake reviews may also be particularly risky because the competitor is strongly incentivized to identify the behavior and report it to Amazon or

regulators.

7 Conclusion

A core mission of consumer protection regulators is to prevent firms from engaging in deceptive practices. A form of deception of growing importance 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.

There are two channels by which rating manipulation impacts consumer welfare. The first is the direct effect of the deception, which we refer to as misinformation. Fake reviews harm consumers by misleading them into purchasing low-quality products with inflated ratings and further by allowing these products to raise prices. On the other hand, they benefit consumers by increasing competitive pressure on honest sellers. The second is the indirect effect on consumers' trust in ratings. These effects are also ambiguous, as mistrust limits consumers' ability to benefit from the information ratings provide but can also increases price competition as firms find it more difficult to differentiate on quality.

We formalize these effects with an explicit model of how consumers formulate beliefs about quality from ratings and make purchases based on these beliefs. There are a few key inputs required for our analysis. The first are novel data on several thousand products on Amazon for which can directly observe fake review activity. The second are consumers' perceptions of fake reviews in the market place, which we elicit using a survey experiment. The last are preference parameters characterizing consumer demand, which we estimate using substitution patterns from scraped Amazon data.

We leverage our model simulate equilibrium effects of fake reviews on the Amazon marketplace and to quantify the different channels by which consumers are impacted. We find that the presence of fake reviews harms consumers. Harms principally occur through the

misinformation channel. In contrast, mistrust actually makes consumers slightly better off, both by offsetting misinformation and by increasing 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.

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.

FTC (2019). Ftc brings first case challenging fake paid reviews on an independent retail website. `https://www.ftc.gov/news-events/press-releases/2019/02/ftc-brings-first-case-challenging-fake-paid-reviews-independent`. Accessed: 2024-03-18.

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

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.

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.

Reimers, I. and Waldfogel, J. (2017). Throwing the Books at Them: Amazon's Puzzling Long Run Pricing Strategy. *Southern Economic Journal*, 83(4):869–885.

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.

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.

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).

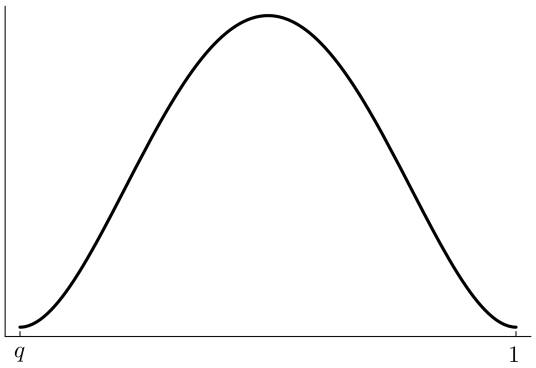


Figure A.1: Distribution of R_j with fake reviews.

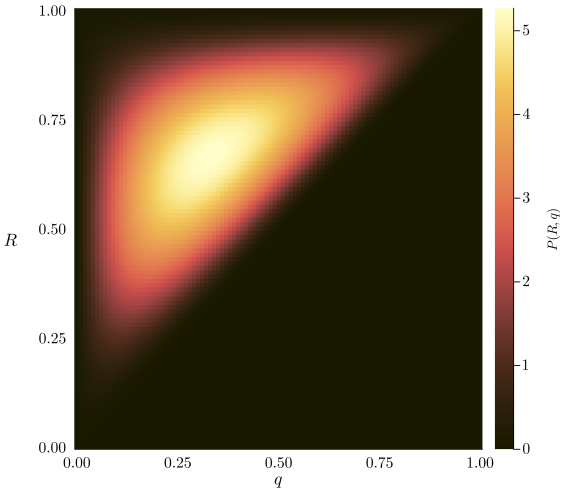


Figure A.2: Joint distribution of quality and R

A.2 Posteriors under beta-distributed priors

The consumer's prior beliefs of quality are distributed beta with parameters α_F, β_F for $F \in \{0, 1\}$, with the probability density function:

$$\mu_{Fq} = \frac{(p_{Fq})^{\alpha_F - 1}(1 - p_{Fq})^{\beta_F - 1}}{B(\alpha_F, \beta_F)}$$

For a given product with quality q, the probability that its first N reviews include N^+ good reviews and N^- bad reviews is

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

Given that a product has N reviews split into N^+ positive and N^- negative, the consumer's posterior probability distribution of the product's quality is a beta distribution with parameters $(\alpha_F + N^+, \beta_F + N^-)$:

$$\begin{aligned}
P(q|N, N^+, F) &= \frac{P(N, N^+|q, F)\mu_{Fq}}{P(N^-, N^+|F)} \\
&= \frac{\binom{N}{N^+} p_{Fq}^{N^+}(1 - p_{Fq})^{N^-}\mu_{Fq}}{\sum_{\tilde{q} \in \mathcal{Q}} \binom{N}{N^+} p_{F\tilde{q}}^{N^+}(1 - p_{F\tilde{q}})^{N^-}\mu_{F\tilde{q}}} \\
&\approx \frac{p_{Fq}^{N^+}(1 - p_{Fq})^{N^-}\mu_{Fq}}{\int_{\tilde{q}=0}^{1} p_{F\tilde{q}}^{N^+}(1 - p_{F\tilde{q}})^{N^-}\mu_{F\tilde{q}}d\tilde{q}} \\
&= \frac{p_{Fq}^{N^+}(1 - p_{Fq})^{N^-} p_{Fq}^{\alpha_F - 1}(1 - p_{Fq})^{\beta_F - 1}B(\alpha_F, \beta_F)^{-1}}{\int_{\tilde{q}=0}^{1} p_{F\tilde{q}}^{N^+}(1 - p_{F\tilde{q}})^{N^-} p_{F\tilde{q}}^{\alpha_F - 1}(1 - p_{F\tilde{q}})^{\beta_F - 1}B(\alpha_F, \beta_F)^{-1}d\tilde{q}} \\
&= \frac{p_{Fq}^{N^+ + \alpha_F - 1}(1 - p_{Fq})^{N^- + \beta_F - 1}}{\int_{\tilde{q}=0}^{1} p_{F\tilde{q}}^{N^+ + \alpha_F - 1}(1 - p_{F\tilde{q}})^{N^- + \beta_F - 1}d\tilde{q}} \\
&= \frac{p_{Fq}^{N^+ + \alpha_F - 1}(1 - p_{Fq})^{N^- + \beta_F - 1}}{B(N^+ + \alpha_F, N^- + \beta_F)}.
\end{aligned}$$

The consumer's unconditional posterior distribution is:

$$P(q|N, N^+) = \sum_{F \in \{0,1\}} \frac{p_{Fq}^{N^+ + \alpha_F - 1}(1 - p_{Fq})^{N^- + \beta_F - 1}}{B(N^+ + \alpha_F, N^- + \beta_F)} P(F).$$

B Data Appendix

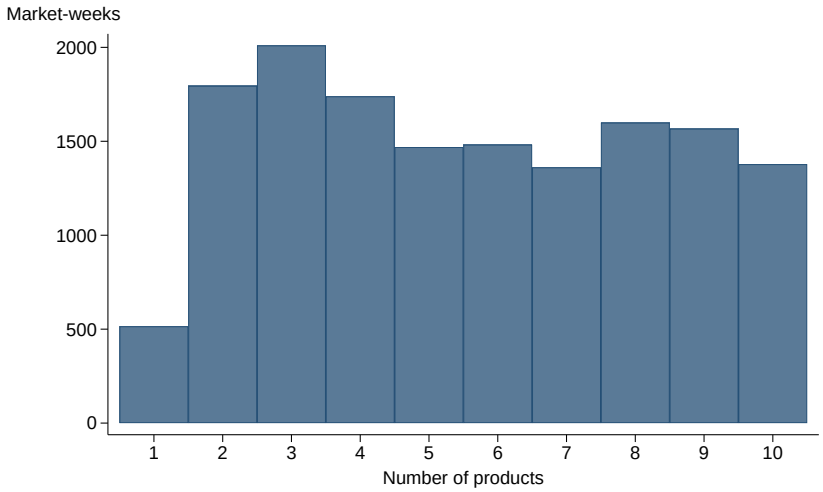
B.1 Data Description

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%.[17]

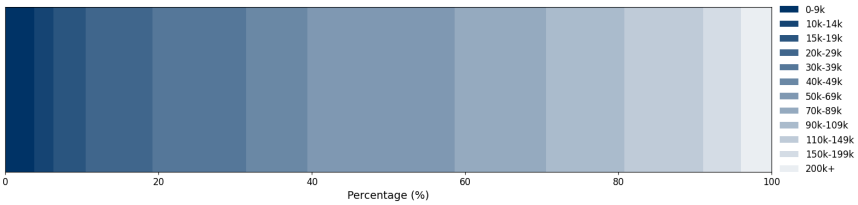
B.3 Survey Experiment

B.3.1 Demographics

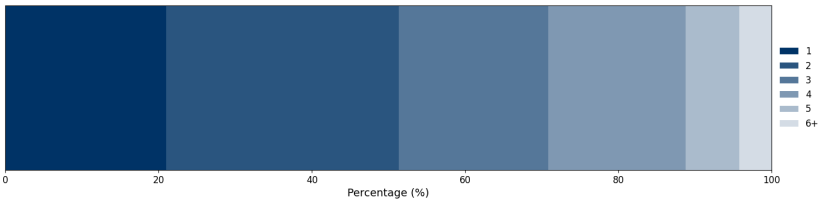
[17]By contrast, among honest products, we observe only .6% of their reviews are left by these fake reviewers.

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

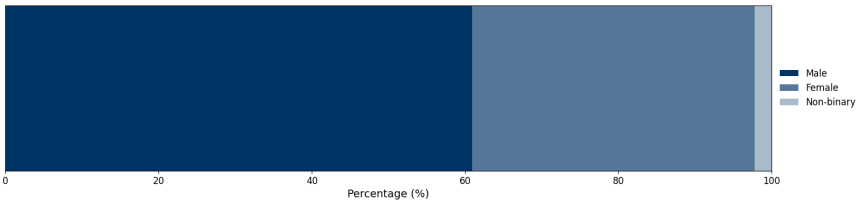
(a) Income



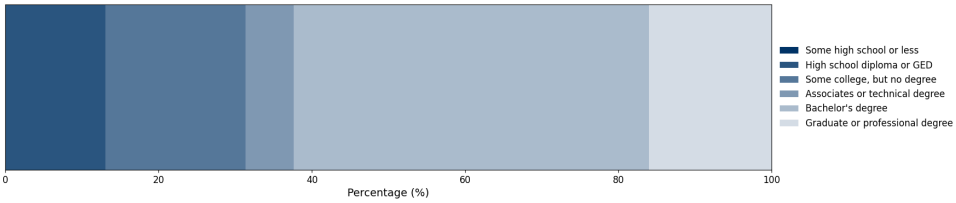
(b) Household Size



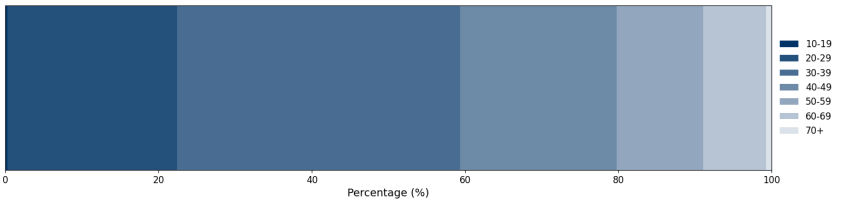
(c) Gender



(d) Education



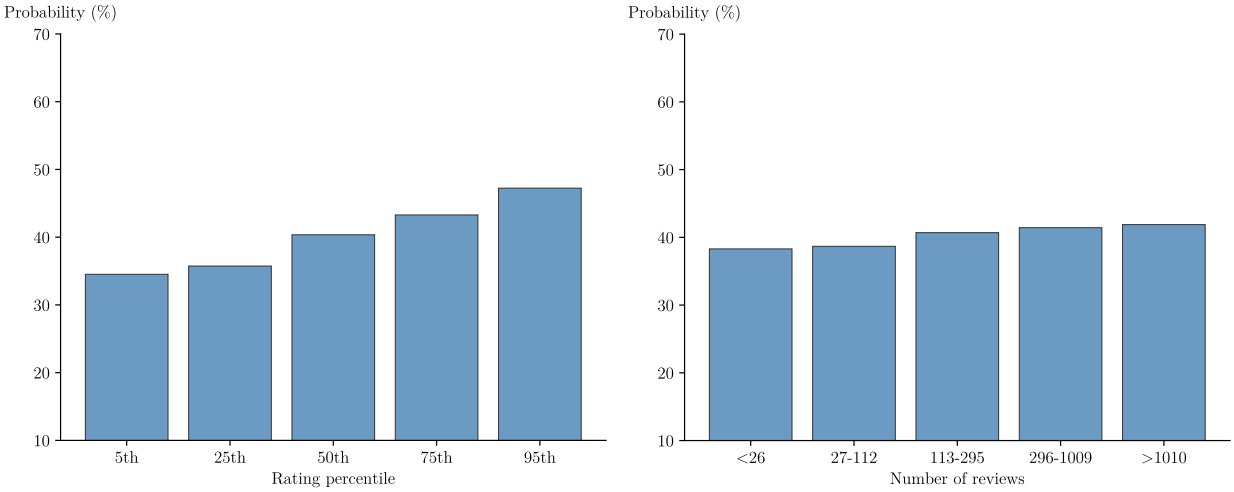
(e) Age



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

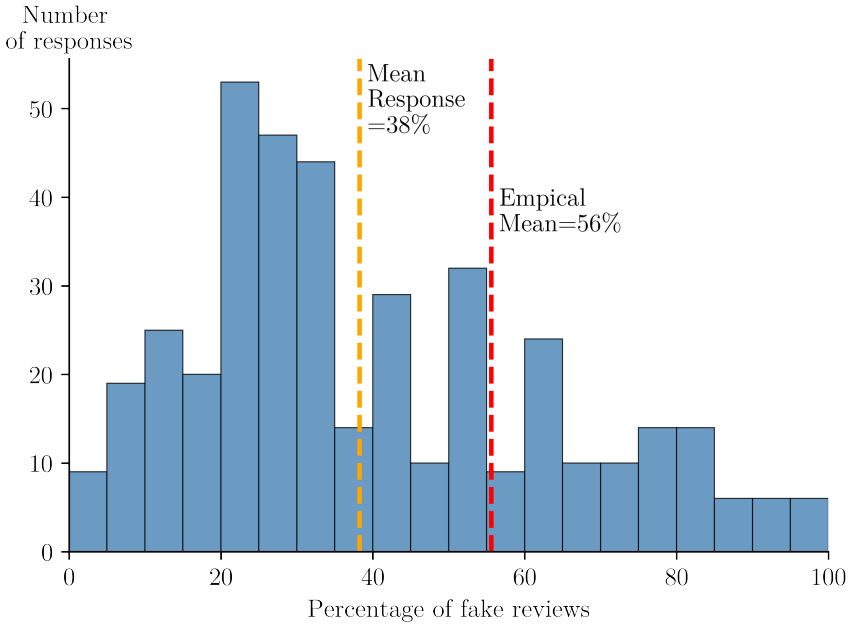


(a) By Rating

(b) By # Reviews

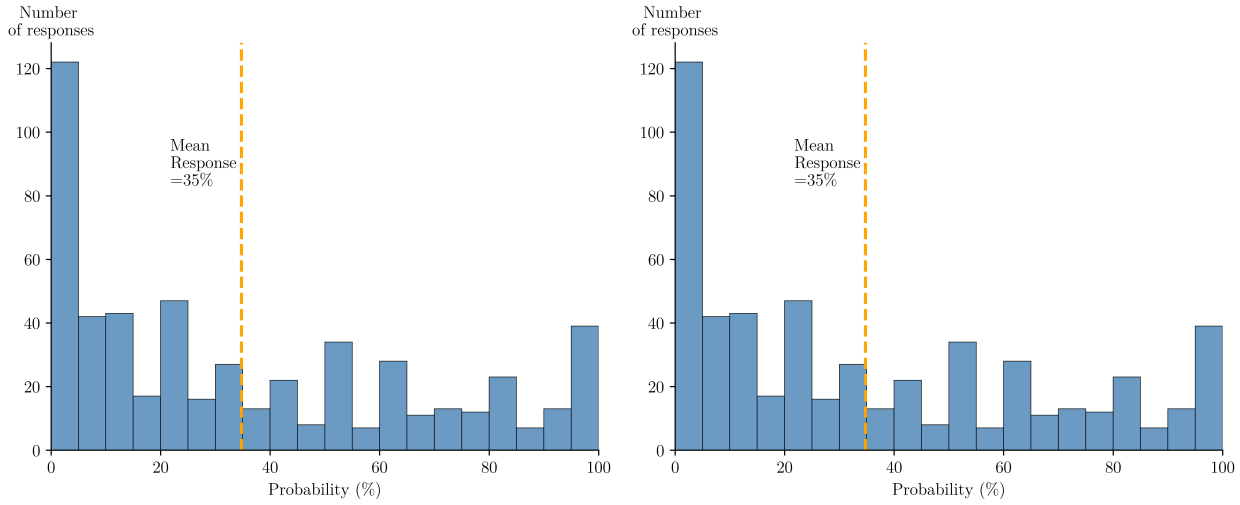
B.3.3 Beliefs about the Frequency of Fake Reviews

Figure B.4: Surveyed Perceptions of θ^F



B.3.4 Supplemental Survey with Review Text

Figure B.5: Perceptions of Fake Review After Viewing Review Text



(a) Fake Review Purchasers

(b) Honest Products

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

To investigate the impact of review histograms, we show 60% of our sample the modal histogram for number of reviews and rating they observe (Figure B.6) and randomize the remaining 40% between highly bimodal (Figure B.7) and unimodal (Figure B.8) histograms (95th and 5th percentiles of variance). Figures B.9 and B.10 show how the results compare for these histograms for each bin of ratings and number of reviews. Respondents appear to report slightly higher probabilities when shown ratings histograms with higher variances. However, the difference is not generally dramatic and are driven by the product pages with very few reviews or low ratings.

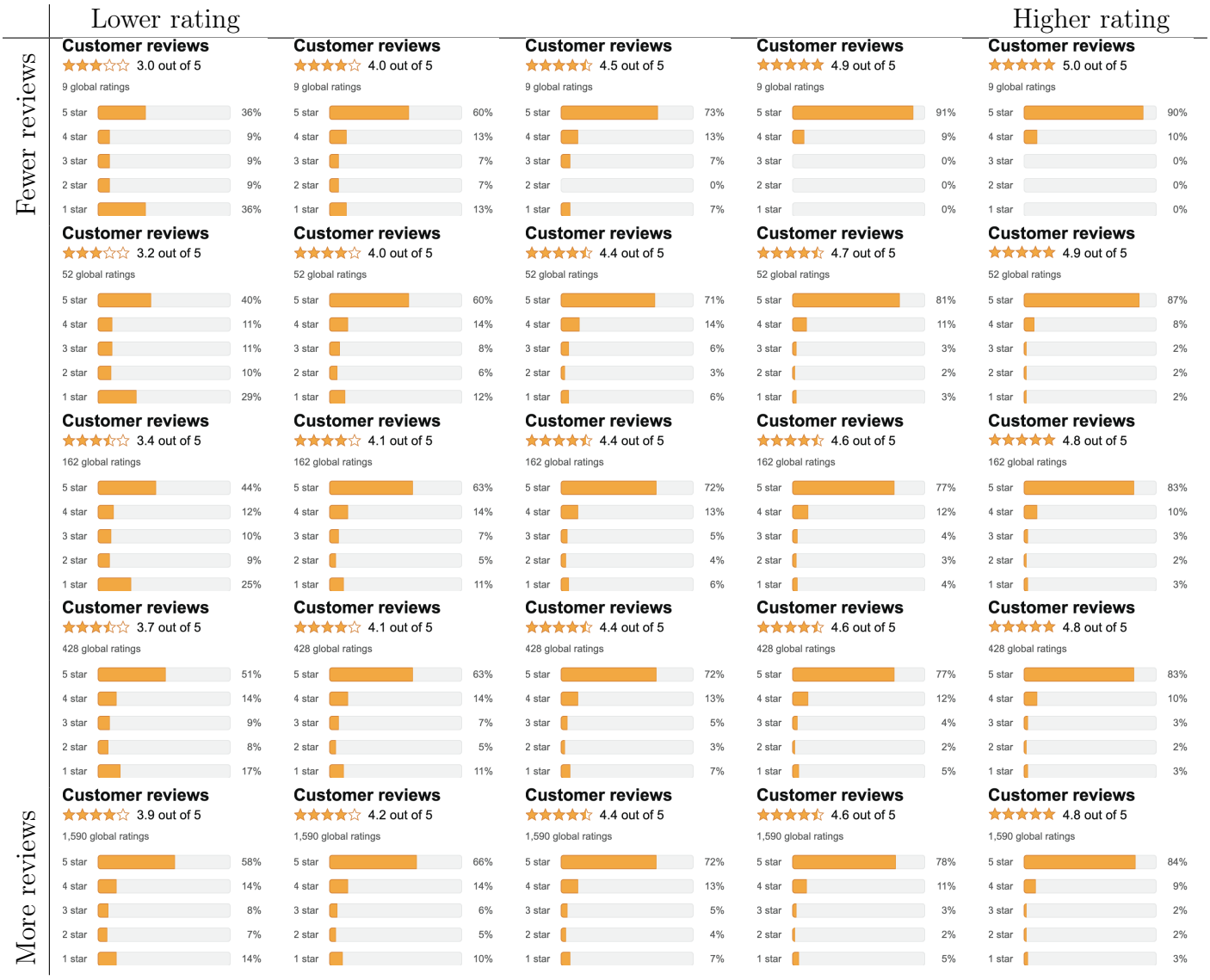


Figure B.6: Modal histograms

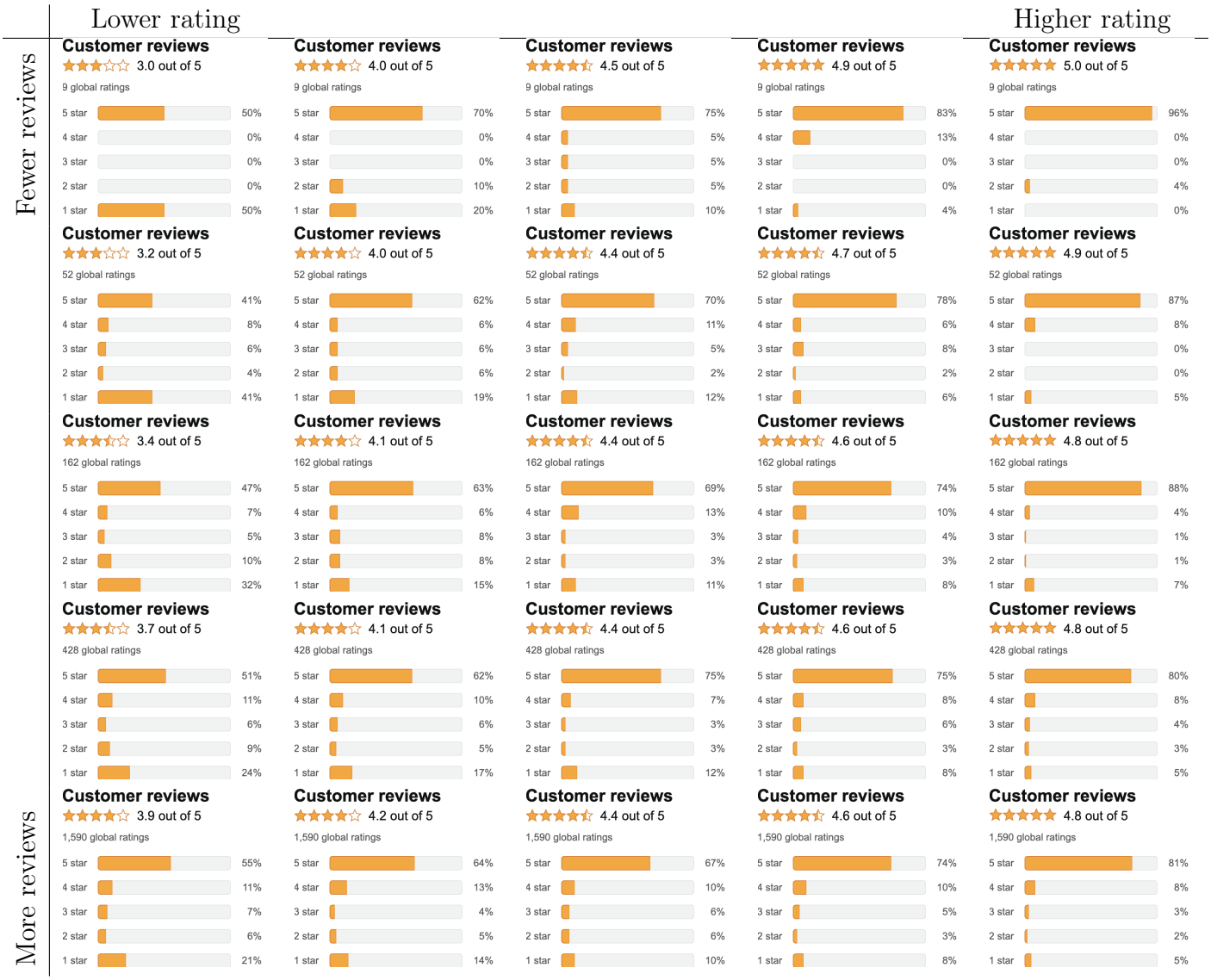


Figure B.7: Bimodal histograms

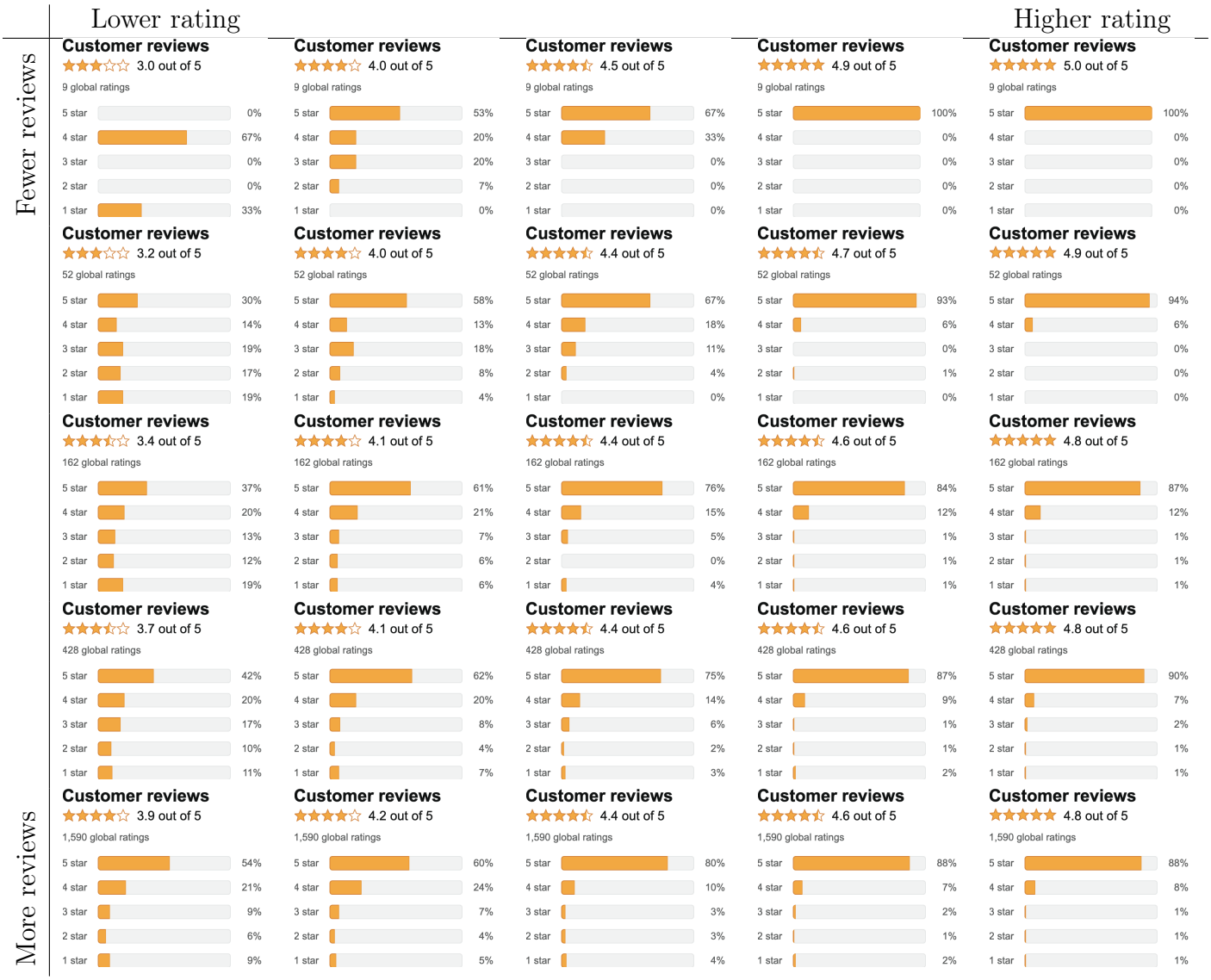


Figure B.8: Unimodal histograms

Figure B.9: Comparing Across Histograms by Rating

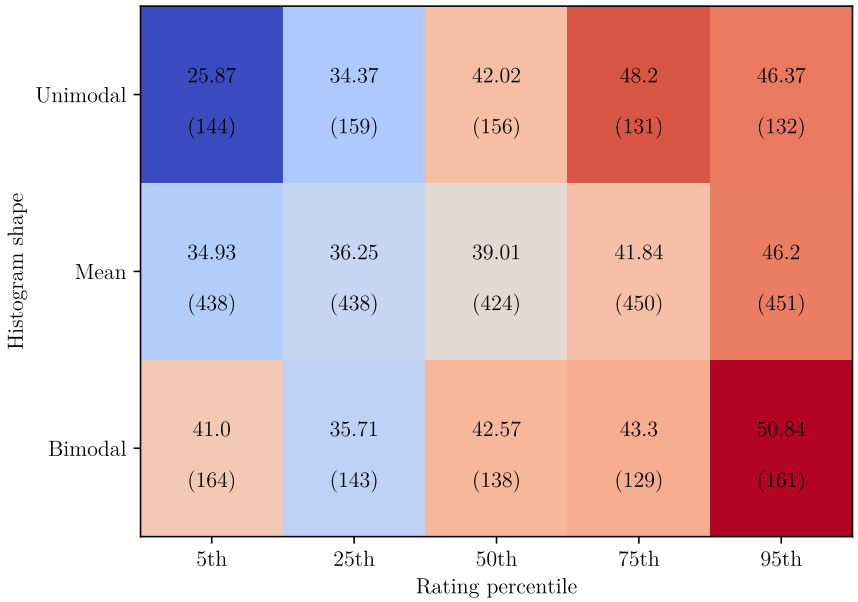
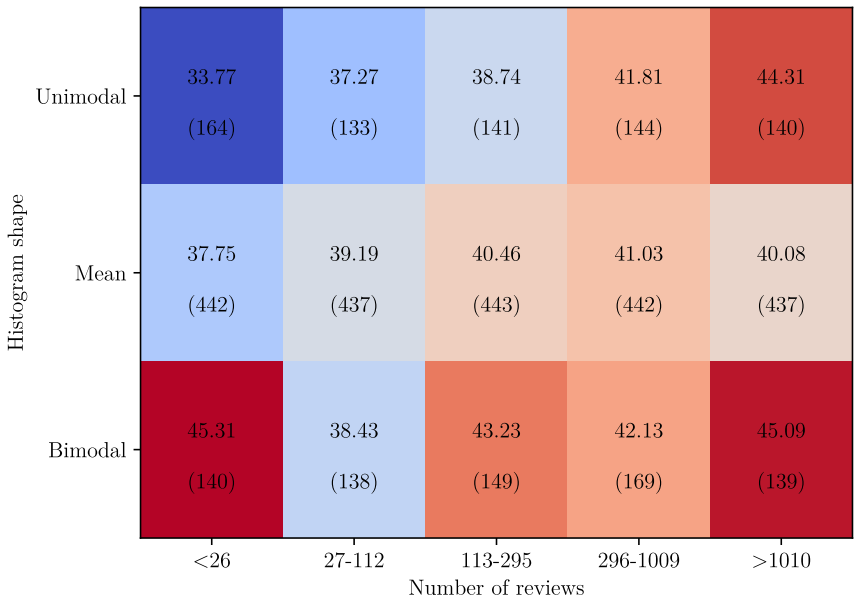


Figure B.10: Comparing Across Histograms by Number of Reviews



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.

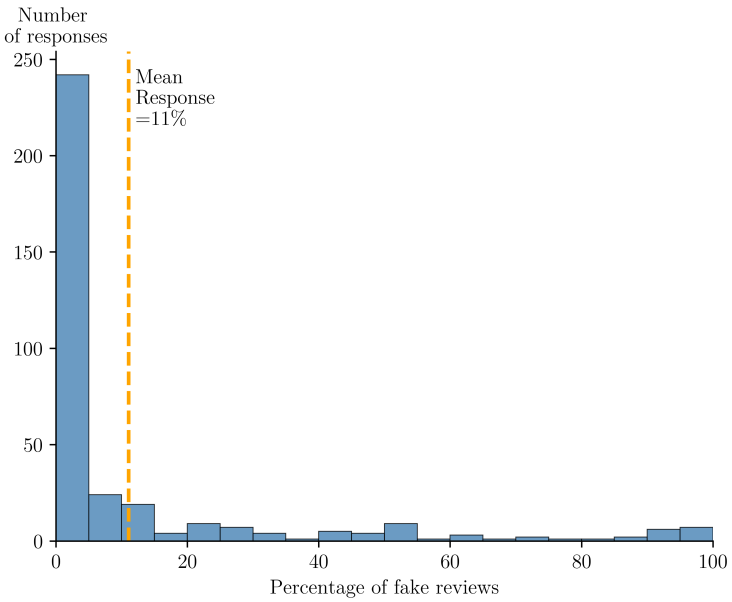


Figure B.11: Responses for Amazon Gift Card

C Counterfactual Simulations

C.1 Additional Details

C.1.1 Hedonic Model of Product Search Rank

A product listing's rank on Amazon is affected by the sales of the product and its competitors. Accounting for this is important in estimating the full impact of counterfactual policies, as the counterfactual changes in perceived quality affects not just current shares but also future demand through the changes in ranks. To capture this feedback mechanism, we conduct dynamic simulations that estimate the demand in each period using counterfactual product ranks, which are predicted using past-period counterfactual shares. The counterfactual ranks are predicted using estimates from a hedonic model of product ranks based on past shares, past reviews, the age of the product on the market, and current sponsorship status.

We break down the product ranking decision into a discrete choice problems of which product to rank first in a series of descending sets of products. That is, the ranking of a set of J products in a market is treated as $J-1$ observations of the ranking algorithm deciding to rank the j^{th} product first in the choice set that is the products ranked j through J. We

estimate a multinomial logit model of the choice of product to rank first in each possible choice problem for each market.

Table C.1 shows the estimation results. Among the lagged variables, the most significant predictors were the market shares and number of good reviews in the past two weeks.

Table C.1: Hedonic model of product rank

Log Shares: Lag 1	0.262***	0.281***	0.206***	0.276***
	(18.21)	(18.84)	(15.35)	(18.16)
Log Shares: Lag 2	0.160***	0.177***	0.142***	0.177***
	(11.96)	(12.84)	(10.99)	(12.48)
Log N. Good Reviews: Lag 1	0.100***			
	(14.90)			
Log N. Good Reviews: Lag 2	0.0717***			
	(11.08)			
Cumulative rating: Lag 1		0.0745***		0.0687***
		(9.09)		(8.23)
Cumulative rating: Lag 2		0.0686***		0.0703***
		(8.76)		(8.27)
Weekly rating: Lag 1			0.0232***	0.0200***
			(4.88)	(4.00)
Weekly rating: Lag 2			0.0133**	0.00175
			(2.91)	(0.37)
Log Cumulative N. Reviews: Lag 1		0.105***		0.0900***
		(15.76)		(12.02)
Log Cumulative N. Reviews: Lag 2		0.0758***		0.0595***
		(11.82)		(8.41)
Log Weekly N. Reviews: Lag 1			0.0592***	0.0137
			(8.28)	(1.64)
Log Weekly N. Reviews: Lag 2			0.0304***	0.0221**
			(4.27)	(2.89)
Sponsored	0.476***	0.469***	0.489***	0.471***
	(13.17)	(12.99)	(13.54)	(13.06)
Constant	-1.438***	-1.296***	-1.364***	-1.383***
	(-6.57)	(-5.90)	(-6.19)	(-6.28)
Product FEs	Yes	Yes	Yes	Yes
Observations	317472	317472	317472	317472

t statistics in parentheses

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

C.1.2 Modeling Organic Reviews

To model the generation of organic reviews, we estimate a Poisson model in which the number of organic reviews arriving each week depends on the logged number of units sold in the two weeks prior. The results of the estimation are reported in Table C.2. We find that a 1% increase in previous-week sales increases the number of organic reviews by just under 1%. The coefficient on the second lag is also significant, and we include both lags when simulating counterfactuals.

Table C.2: Poisson model of organic reviews arrival

Log Sales: Lag 1	0.586***	0.550***
	(19.98)	(16.95)
Log Sales: Lag 2		0.0441***
		(3.60)
Constant	1.410***	1.396***
	(7.79)	(7.68)
N. Obs.	56499	56499
Mean Dep. Var.	57.41	57.41
SD	300.6	300.6

t statistics in parentheses

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

Table C.3: Counterfactuals with endogenous organic reviews

	No FR	Misinfo		Mistrust		Misinfo+Mistrust	
		Fixed prices	Floating prices	Fixed prices	Floating prices	Fixed prices	Floating prices
Welfare ($)	365,689,662	363,482,194	363,297,846	365,544,835	366,193,535	363,549,049	363,930,515
Welfare change (%)	+0.00	-0.60	-0.65	-0.04	+0.14	-0.59	-0.48
FRP average prices ($)	30.89	30.89	31.57	30.89	30.78	30.89	31.47
HP average prices ($)	37.96	37.96	37.87	37.96	37.92	37.96	37.83
FRP sales (units)	1,588,863	1,987,557	1,921,590	1,532,692	1,540,026	1,937,184	1,879,868
HP sales (units)	9,872,529	9,613,339	9,668,638	9,730,095	9,735,115	9,472,060	9,529,524
FRP revenue ($)	49,077,636	60,150,023	60,037,435	47,188,502	47,187,572	58,703,492	58,544,840
HP revenue ($)	367,773,495	358,324,755	359,599,850	363,164,765	362,838,858	353,633,070	354,548,209
Platform revenue ($)	41,685,113	41,847,478	41,963,729	41,035,327	41,002,643	41,233,656	41,309,305
FRP profits ($)	34,124,501	41,602,756	41,880,040	32,945,910	32,926,567	40,584,878	40,821,504
HP profits ($)	191,289,163	186,034,234	186,348,557	188,570,709	188,346,763	183,359,531	183,393,347

Table C.4: Counterfactuals with fixed organic reviews

	No FR	Misinfo		Mistrust		Misinfo+Mistrust	
		Fixed prices	Floating prices	Fixed prices	Floating prices	Fixed prices	Floating prices
Welfare ($)	365,921,547	362,231,885	362,358,739	365,769,000	366,426,685	362,253,385	362,967,294
Welfare change (%)	+0.00	-1.01	-0.97	-0.04	+0.14	-1.00	-0.81
FRP average prices ($)	30.93	30.93	31.75	30.93	30.80	30.93	31.66
HP average prices ($)	37.96	37.96	37.85	37.96	37.93	37.96	37.81
FRP sales (units)	1,608,187	2,126,500	2,027,958	1,535,860	1,544,701	2,088,211	1,994,048
HP sales (units)	9,885,461	9,545,877	9,625,224	9,753,101	9,756,511	9,398,225	9,483,198
FRP revenue ($)	49,659,307	64,006,259	63,449,937	47,384,010	47,366,230	62,920,388	62,231,974
HP revenue ($)	367,933,270	355,650,247	357,499,346	363,648,174	363,298,656	350,684,508	352,293,041
Platform revenue ($)	41,759,258	41,965,651	42,094,928	41,103,218	41,066,489	41,360,490	41,452,501
FRP profits ($)	34,528,424	44,242,450	44,279,768	33,047,540	33,011,750	43,458,591	43,393,579
HP profits ($)	191,481,560	184,585,524	185,158,158	188,971,025	188,731,618	181,776,853	182,126,846

C.1.3 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:[18]

$$\begin{aligned} P(q|N^{o+}, N^o, F; \hat{\gamma}) &= \frac{P(N^{o+}|q, N^o, F)P(q|F; \hat{\gamma})}{P(N^{o+}|N^o, F)} \\ &= \frac{q^{N^{o+}}(1-q)^{N^o - N^{o+}}P(q|F; \hat{\gamma})}{\int q^{N^{o+}}(1-q)^{N^o - N^{o+}}dP(q|F; \hat{\gamma})}. \end{aligned} \tag{9}$$

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:

$$\Delta\mathbb{E}q := \int q dP(q|N^+, N; \hat{\gamma}) - \int q dP(q|N^{o+}, N^o, F; \hat{\gamma}). \tag{10}$$

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.1.4 Summary of changes with Misinformation and Mistrust

[18] The first equality applies the assumption that $P(q|N^o, F) = P(q|F)$.

Table C.5: Changes in outcomes with misinformation and mistrust

	Mean	P25	Median	P75
Welfare (%)	-2.7	-1.3	-0.1	0.1
FRP prices (%)	4.4	0.2	1.4	4.5
HP prices (%)	-0.6	-0.6	-0.2	-0.0
FRP sales (%)	43.5	2.2	20.3	66.7
HP sales (%)	-3.6	-5.4	-2.5	-0.6
FRP revenues (%)	49.9	2.9	25.9	77.6
HP revenues (%)	-4.1	-6.2	-2.9	-0.6
FRP profits (%)	50.4	3.0	25.4	78.1
HP profits (%)	-4.2	-6.3	-2.9	-0.7

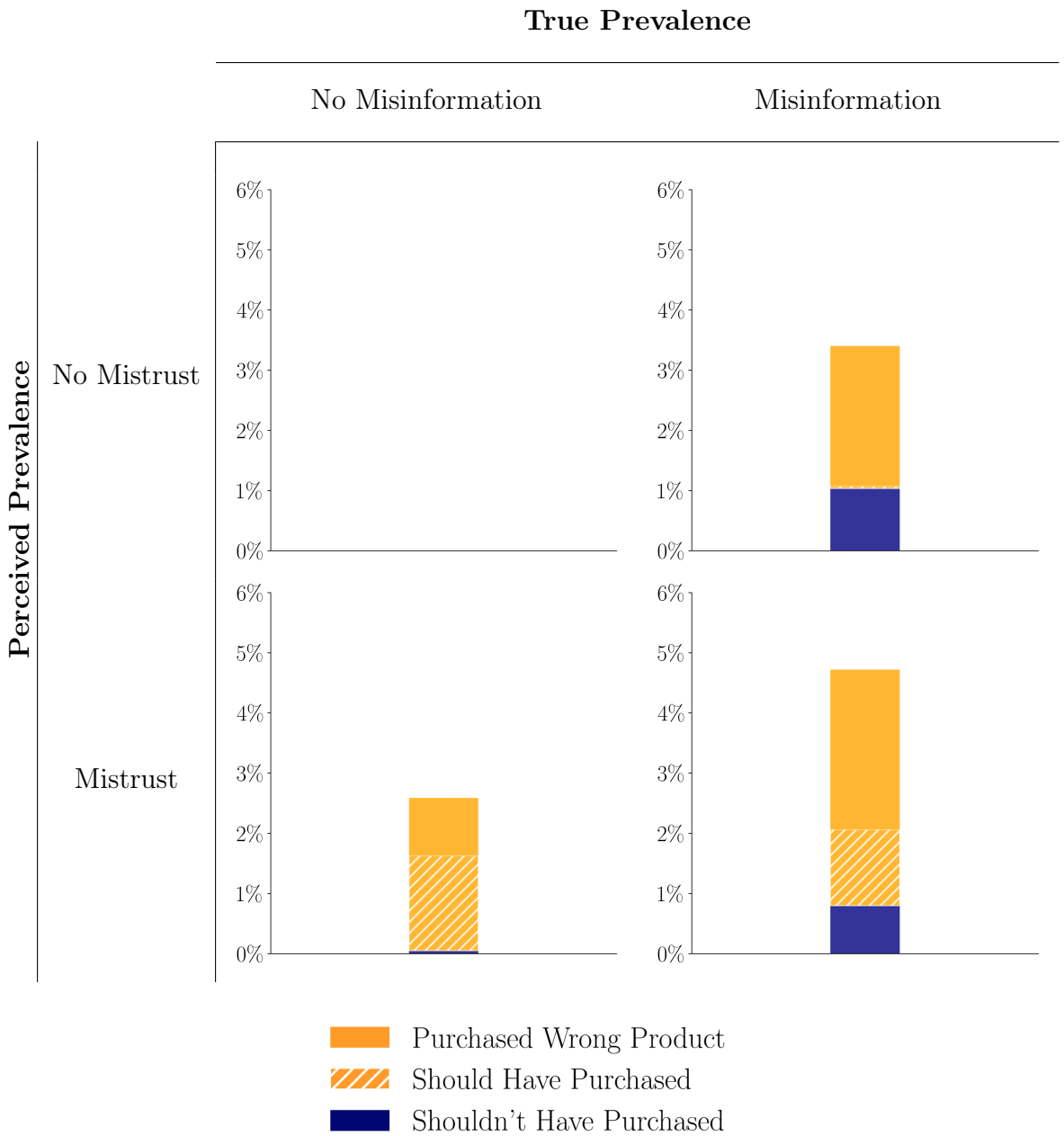
C.2 Additional Counterfactuals

C.2.1 Additional Results for Estimating the Effect of Fake Reviews

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

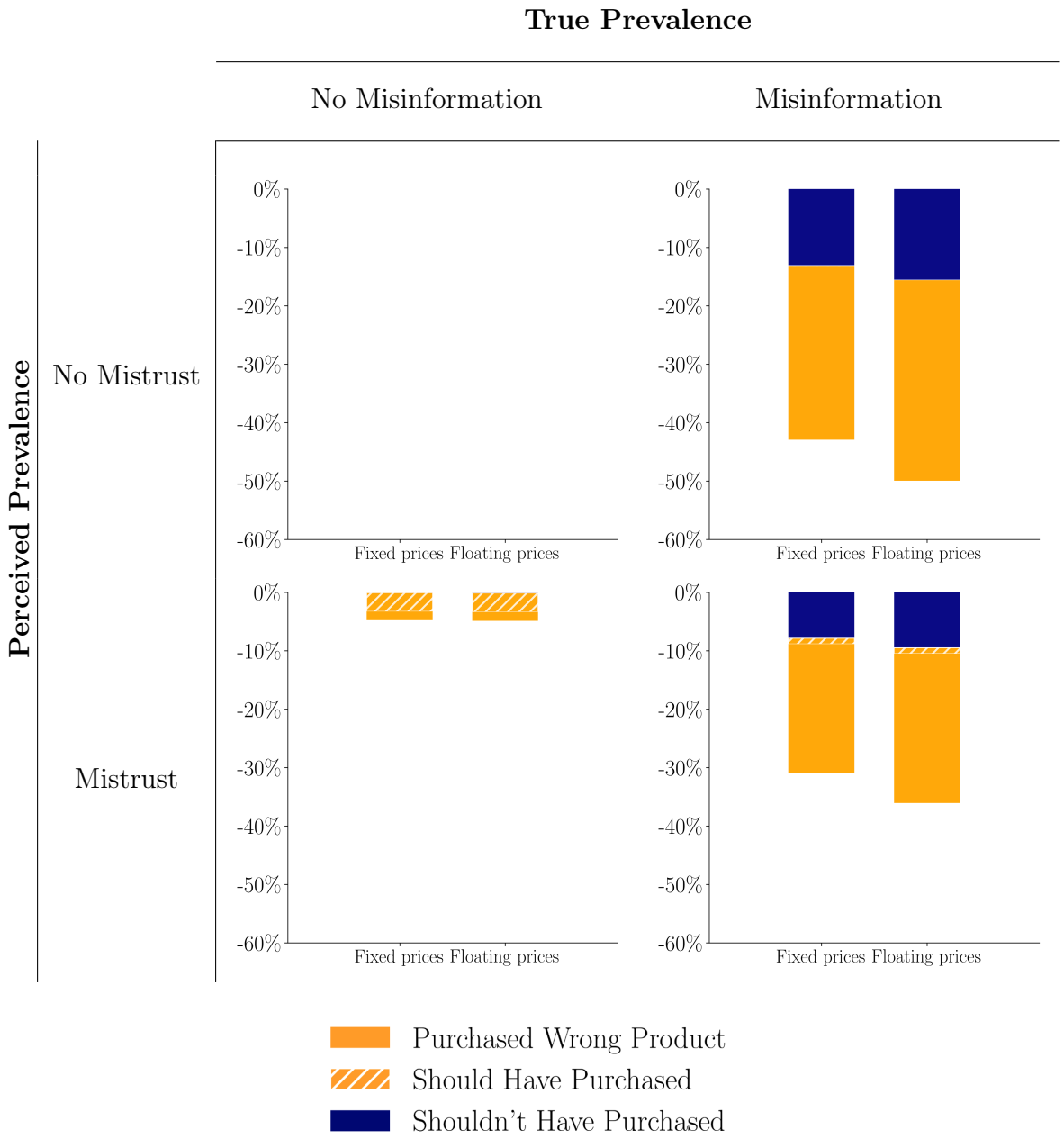
(a) Misinformation

(b) Mistrust



Note: Figure tabulates the number of consumers who make each type of mistakes made under combinations of misinformation and mistrust. Here, we fix prices at the equilibrium levels with no fake reviews.

Figure C.2: Mistakes Under Misinformation and Mistrust



Note: Figure tabulates the proportion of welfare harms done by each type of mistake made under each counterfactual scenario. The denominator is the total welfare of all consumers who made mistakes in the given counterfactual scenario, had they chosen correctly.

Figure C.3: Welfare Harms Under Misinformation and Mistrust

C.2.2 Benefits and Costs of Purchasing Negative Fake Reviews

In this section, we compare the benefits from purchasing a fake review to that of purchasing a fake negative review for one's competitor. 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.

Figure C.4: Weeks to Break Even from Purchasing One Negative Review for a Close Competitor

The median net cost of a negative fake review for a close competitor is \$26.48, which is much greater than the median net cost of a positive fake review of \$11.70. Additionally, the benefits of a negative fake review tend to be much lower since diverted consumers do not fully shift 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 \$5.21 compared to \$55.15 for a positive fake review. Figure C.4 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.2.3 Deleting Fake Review Purchasers

In this section, we consider the counterfactual policy of deleting Fake Review Purchasers 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.5. 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 1.5% higher and gain profits that are around 15% higher than the factual equilibrium.

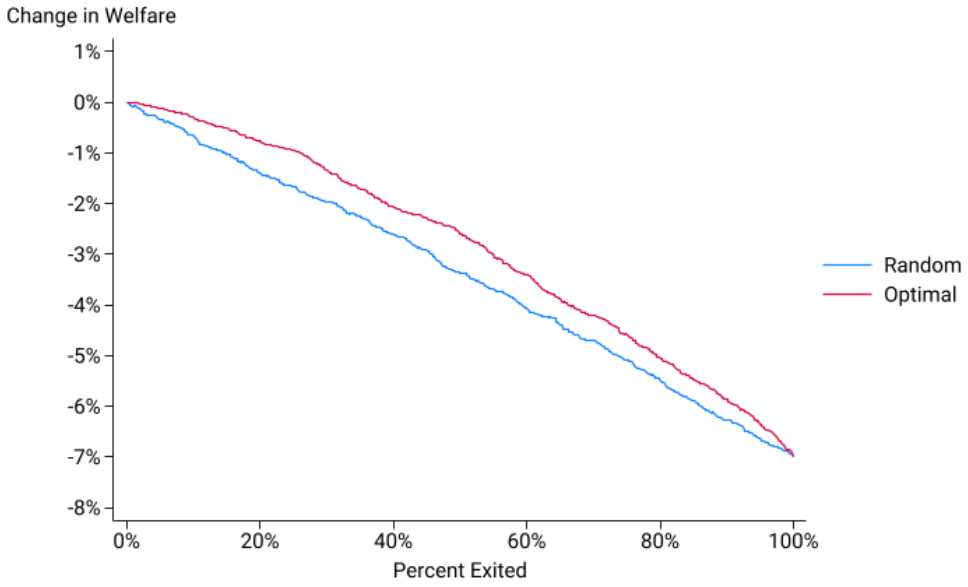


Figure C.5: Change in welfare from deletion of Fake Review Purchasers

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 above, as is the magnitude of the welfare changes.

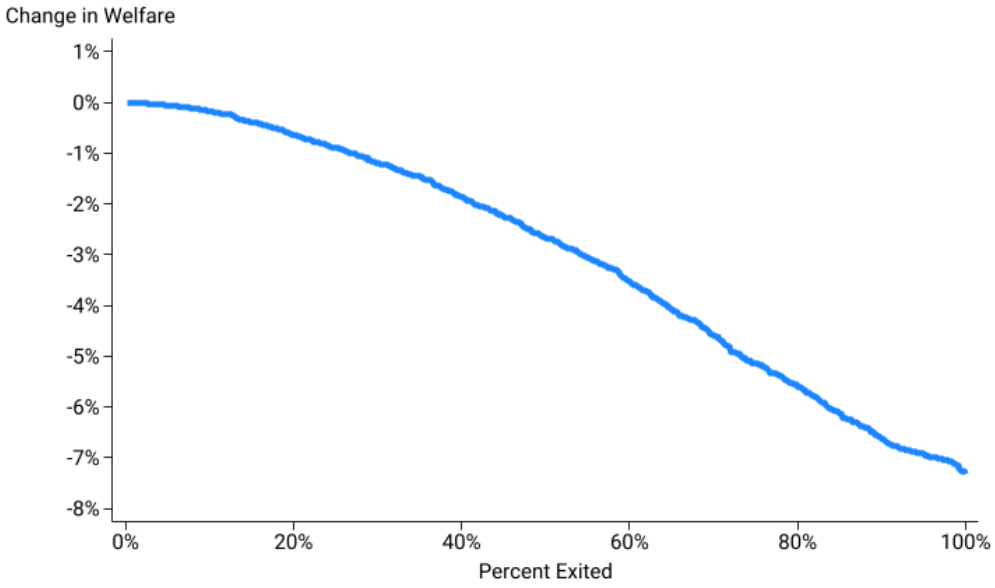


Figure C.6: Change in welfare from exit of Fake Review Purchasers