Manipulation through Biased Product Reviews*

Kemal Kıvanç Aköz† Cemal Eren Arbatlı‡ Levent Çelik§

November 13, 2017

Abstract
We analyze an information manipulation game in which consumers estimate the quality of a product using product reviews as a noisy source of information. A firm, whose product quality is either ‘high’ or ‘low’, exerts costly and hidden efforts to insert bias into product reviews. We first consider a case in which the price is uninformative about quality. We show that, in equilibrium, there is wasteful spending on manipulation by both firm types. Depending on the price level, either one type or both types enjoy higher sales compared to a situation in which there is no manipulation. Thus, the price level governs whether the bias in product reviews benefits or harms the consumers. A regulator can increase consumer surplus by charging a fixed fee to the firm for access to product review platforms. When we allow for informative pricing, we find that there are partially separating equilibria in which the low-type firm randomizes between choosing a low price and the high price that the high-type firm would choose. We show that in these equilibria, biased product reviews still influence demand. However, unlike the high-type firm, which can increase its sales through manipulation only at sufficiently high levels of prices, the low-type firm enjoys higher sales through manipulation at any admissible price.

JEL: D8, K4, L1, L4.
Keywords: Information manipulation, Bayesian inference, product reviews, false advertising, price signaling, signal-jamming.

*Acknowledgements: We would like to thank Jacopo Bizzotto, Olivier Bochet, Mikhail Drugov, Anett Erdmann, Sergei Izmalkov, Maarten Janssen, Matthew Nagler, Regis Renault, Sandro Shelegia, Alex Smolin, Chris Wilson and John Wooders as well as the participants of the AUS Economics Seminar, NYUAD Theory Workshop, 3rd NES CSDSI workshop, and conferences EARIE 2017, IIOC 2017, XXXII Jornadas de Economía Industrial, 10th Workshop on the Economics of Advertising and Marketing and 2017 Lisbon Meetings in Game Theory and Applications for valuable discussions.

†New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates.
‡National Research University Higher School of Economics, Myasnitskaya 20, 101000, Moscow, Russia.
§National Research University Higher School of Economics, Myasnitskaya 20, 101000, Moscow, Russia.
1 Introduction

Information provision in the Internet era has become increasingly decentralized thanks to the expanding scale and scope of crowdsourcing and word-of-mouth. More and more consumers rely on product reviews on Amazon.com, hotel and restaurant reviews on TripAdvisor, or movie reviews on IMDb to help them make a better choice. As participation in online review platforms increases, firms are more tempted to make their online reviews look better. On the bright side, these online platforms have presumably induced higher competition among firms and thereby an overall improvement in the quality of products and services. However, they also have a potential dark-side as they allow greater anonymity to users. Although firms cannot fully control the content shared on these platforms, they can use various strategies to manipulate consumer opinions. Posting or funding fake reviews, incentivizing consumers to recommend a product, or selectively funding and disseminating research results that provide favorable information about a product are some of these strategies. Although consumers can rationally anticipate a positive bias in these reviews and recommendations, the real extent of this bias is hard to assess. Even if consumers were fully rational, the reviews they sample may not lead to an unbiased prediction about the quality of a product.

In this paper, we formalize these ideas through a signal-jamming model of biased product reviews and investigate when and how manipulation affects market demand and consumer welfare. We consider a single firm and a continuum of potential consumers with unit demand for an experience good (Nelson, 1970). The good is characterized by its inherent quality, which can be either high or low. The firm is privately informed about its type (henceforth, we treat the quality levels as the types of the firm). Consumers are uncertain about the quality, but draw on two sources of information before making a purchasing decision.

---

1According to a report in 2010, Amazon.com was the largest single source of consumer reviews on the internet. TripAdvisor, the largest travel site in the world, now has more than 315 million members and over 500 million reviews and opinions of travel-related businesses. IMDb, likewise, has 80 million registered users and over 4.5 million titles in its database.

2See, among others, Hu et al. (2011), Anderson and Simester (2014), and Mayzlin, Dover and Chevalier (2014). It is not difficult to find people who offer their services to write embellished product reviews in return for a small amount of money. According to a 2012 story in the New York Times, a Craigslist post proposed the following: “If you have an active Yelp account and would like to make very easy money please respond” (source: http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html). Bing Liu, a data-mining expert at the University of Illinois, Chicago, estimates that about a third of all online consumer reviews are fake.

3See, for example, Finucane and Boult (2004), and Sismondo (2008) for the case of medical research.

4See Mayzlin et al. (2014) for a discussion on the difficulty of distinguishing biased online reviews from unbiased ones.
making a purchase decision. First, by visiting a product review platform, they receive an individual-specific noisy signal about product quality. The firm can shift the mean of these quality signals by exerting costly effort. We assume that each firm type faces the same manipulation cost, which increases with the extent of the bias. Consumers cannot observe the firm’s efforts, but they do see the price set by the firm. This constitutes the second source of information. Depending on the type of the equilibrium, the price can be informative about quality. In particular, the low-type firm might use a mixed pricing strategy in which it sometimes charges the same price as the high-type firm so as to bias consumer beliefs upwards by mimicking the latter. Given the quality signal and the price, consumers form their posterior beliefs about quality and decide whether to purchase the product or not. This determines the final market demand. We say that manipulative reviews are ‘effective’ for a given firm type when they raise its demand above the level we would observe without any manipulation. The existing research typically focuses either on the signaling role of price or direct claims about quality. By presenting an analysis of information manipulation where the firm can use both of these channels concurrently, our paper not only nests and extends some of the results in the literature, but also reveals novel channels through which price can interact with equilibrium expectations of consumers.

We start our formal analysis with a benchmark model in which we treat price as exogenous and identical in both quality states. While this restriction makes the price uninformative, there are two reasons for starting the analysis with this case. First, this restriction allows us to highlight the individual role of manipulation via quality signals and helps us compare our results to some of the existing work. It turns out that the price level plays a key and non-trivial role in determining the extent of information manipulation. Second, and perhaps more importantly, each exogenous price level we consider in the benchmark model can in fact be supported as part of a Perfect Bayesian Equilibrium (henceforth, PBE) where both firm types pool on that price level (see Proposition 4). Thus, our findings in the benchmark model regarding the effects of the price level can be considered as a comparison across many different pooling price equilibria that arise when price is endogenous.

We first show that our benchmark model features a unique PBE in which both firm types pool on the exogenous price level.
types exert strictly positive effort to shift consumer signals. The intuition here is that consumers interpret higher signals as stronger evidence for higher quality. In response, both firm types engage in an implicit arms race where manipulation by one type induces further manipulation by the other. Consumers are aware of the fact that the signals they receive are biased upwards. In fact, they perfectly anticipate the equilibrium level of bias each type will insert into the product reviews. However, since they cannot observe the type realization and since the signals they receive contain random noise, they cannot pin down the underlying product quality based on these signals.

We show that the exact configuration of equilibrium bias levels depends crucially on the price level. We demonstrate in Theorem 2 that the low-quality [high-quality] firm type exerts higher effort to bias consumer beliefs than the other type when the price is lower [higher] than the ex-ante expected quality level. The firm type with higher bias can always raise its sales vis-a-vis the case without manipulation. Furthermore, at sufficiently high or low prices, manipulation increases sales for both firm types.

As equilibrium bias levels depend on the price level, so do the welfare consequences of biased reviews. In particular, the effect of manipulation through biased product reviews on ex-ante consumer surplus tends to be negative at low prices and positive at high prices. Firm profits, on the other hand, depend not only on sales but also the cost of manipulation. The implicit arms race we described above leads to some wasteful spending. We show that this wasteful spending may cause a reduction in the firm’s profits relative to a no-manipulation scenario, and that the firm would be able to avoid it if it could credibly pre-commit to an ex-ante bias schedule.

In Section 5, we relax the assumption that price is exogenously fixed and allow it to be potentially informative about quality. In this new setting, we first show that a wide range of pooling PBE exists where both types choose the same price, and hence the price remains uninformative. As a result, the equilibrium bias levels in these PBE coincide with those studied in our benchmark model with an exogenous price. Importantly, we then identify a range of partially separating PBE in which the low-quality firm uses a mixed pricing strategy. In these PBE, the high-quality type always charges an endogenously determined price $\bar{p}$, while the low-quality type randomizes between $\bar{p}$ – to mimic the high type – and a lower price. In the latter case, the low-type firm fully reveals itself to consumers and therefore spends no resources on review manipulation. When consumers observe the high price, they revise their beliefs by assigning a higher likelihood about the product being high-quality. They then use their individual signals to form their final posterior beliefs about the product. These PBE vary by $\bar{p}$ and the resulting equilibrium bias levels. We
show that the high-quality type exerts more manipulative effort than the low-quality type in these PBE when \( \bar{p} \) is relatively high, just like in our benchmark model.

When price is informative about quality, the expected consumer surplus depends both on the level of price and on how informative this price is. As \( \bar{p} \) increases, price becomes more informative about quality since the probability that the low-quality type mimics the high type goes down. Therefore, consumers assign a higher probability to the high-quality state when \( \bar{p} \) is higher. This in turn increases the sales for the product. As a result, the low-type firm expects a higher revenue at higher prices and is able to spend more resources on manipulation. The high-quality type follows the suit and also increases its manipulation effort as the implicit arms race between the types becomes fiercer. We find that this escalated arms race enables the low-quality type to increase its sales in all partially separating PBE compared to the corresponding no-manipulation scenario, whereas the high-quality type is able to do so when its equilibrium price is high enough.

This is in contrast to the benchmark model where, for a non-empty range of prices, only the high-quality firm was able to expand its demand – which was beneficial for consumers. Therefore, biased reviews are potentially more harmful in equilibria with informative pricing than under pooling equilibria.

For a regulator, minimizing possible harm to consumers from biased product reviews is not a straightforward task. Policy measures that can in principle increase the quality of information disseminated through online review platforms, such as filtering out fake or embellished reviews or providing more objective reviews, may be costly to implement and result in indirect burdens to consumers. Moreover, when consumers receive more precise product information, the dispersion among consumers’ information may also decrease. But when consumers’ expectations become more clustered, the firm’s manipulative effort can influence decisions of a larger number of consumers at the margin. Our numerical analysis in Section 6.3 provides a parametric example in which higher signal precision may actually increase manipulative effort and lead to higher bias in consumers’ expectations. In Section 6.1, we discuss a simpler and potentially more effective policy measure. The idea is to introduce a fixed cost to the firm for manipulative effort. This could be implemented through a flat fee charged by product review platforms if the firm wants to enable reviews for its product. Such a policy does not require any costly monitoring and therefore affects both types of firm. However, we show that the low-quality type is disproportionately more affected by such a policy because it earns a lower profit than the high-quality type. Therefore, there is an optimal level of fixed cost that completely discourages the low-type firm from exerting any manipulative effort while leaving some room for biased reviews by
the high-type firm.

The rest of the paper is organized as follows. Section 2 offers a brief review of the related literature. Section 3 lays out our benchmark model and characterizes the equilibrium. Section 4 presents the main analytical and numerical results regarding the effect of manipulation on total demand and consumer surplus. It also provides intuition for why and when manipulation can be effective in influencing aggregate consumer behavior. In Section 5, we introduce our extended model with endogenous price. Section 6 discusses the policy implications of our results and introduces two more extensions of the benchmark model. In the first extension, we analyze a modified model where the firm can pre-commit to a manipulation plan. In the second extension, we discuss the effect of signal precision on manipulation. Section 7 contains the conclusion.

2 Related literature

Our analysis relates to several papers in the literature that study information manipulation. Earlier papers such as Matthews and Mirman (1983), and Fudenberg and Tirole (1986) focus on the unobservable pricing decision of an incumbent firm to deter a potential entrant, while Holmström (1999) considers a career-concerns problem where the worker’s effort choice is unobservable. More recently, Edmond (2013), Caselli, Cunningham, Morelli and Moreno Barreda (2014), and Aköz and Arbatlı (2016) study information manipulation in a political context.

Regarding firms’ incentives to bias online product reviews, we are aware of two related papers, Mayzlin (2006) and Dellarocas (2006). In Mayzlin (2006), two firms compete by populating an online forum with costly messages about their products (promotional chat). As in our model, consumers cannot tell apart word-of-mouth from biased reviews posted by firms. In equilibrium, promotional chat is persuasive and the low-quality firm spends more resources on it than the high-quality firm. Dellarocas (2006) also considers strategic manipulation of online forums in a monopoly setting where consumers uniformly value quality but are heterogeneous with respect to the horizontal attribute towards the product. He shows that, under certain conditions, manipulation increases with quality, and when this is the case, it benefits consumers. In Mayzlin (2006), the price is exogenously fixed. In Dellarocas (2006) it is endogenous but completely uninformative about product quality by construction. In contrast, we endogenize the price in a way that allows it to provide partial information about product quality. Hence, the results in these two papers are related.

---

6For a more recent paper along this line, see Mirman, Salgueiro and Santugini (2014).
more closely to our benchmark model with an exogenous price. Differently from these papers though, we treat the price in our benchmark model as a free exogenous parameter and demonstrate a novel mechanism through which it modifies the relative advantage in manipulation. Also unlike Mayzlin (2006) and Dellarocas (2006), under sufficiently extreme price levels, our model features equilibria where manipulation increases sales for both types.

There are a few recent papers that feature manipulation via price signaling (Rhodes and Wilson, 2017, Piccolo, Tedeshi and Ursino, 2015 and 2016, and Janssen and Roy, 2017). In that sense, they are related to our extended model with informative pricing. Some of the equilibria in these papers allow deceptive advertising by the low-quality firm that affects consumers’ posterior beliefs about product quality. Although these models let the low-quality firm mimic the high-quality firm, by construction, the high type cannot respond with counter ads. In contrast, the signal-jamming framework we offer allows for manipulation by both types. Moreover, since the price is not fully informative in our model, consumers still utilize the signals from product reviews as they make their decisions. In this sense, our extended model combines partially informative price signals with partially informative product review signals.

Our paper is also related to the strand of advertising literature that focuses on false advertising or false advice. In some of these papers, false messages are taken as given instead of being derived as an equilibrium choice of firms, and consumers are assumed to take these messages at face value rather than rationally discounting them (e.g., Hattori and Higashida, 2012). Some other papers allow for false or unsubstantiated claims about product quality but they take the strength of such claims as exogenously given (e.g., Corts, 2013). False claims are supported in equilibrium only when firms are uncertain about their own product quality, i.e., when there is no intentional misinformation by firms (e.g., Corts, 2014). Kartik (2009) studies a strategic communication model where a privately-informed sender bears lying costs for misrepresenting his private information. In a signal-jamming framework, Drugov and Troya-Martinez (2015) offer a model of false advice by a seller. However, unlike in our model, the seller in their setting cannot condition the bias on quality. Therefore, false advice, albeit subject to punishment by regulators, does not affect total sales.

In our model, firms do not have full control over the information conveyed to consumers. The signals that consumers receive are informative about quality, and firms are able to increase the mean of these otherwise unbiased signals through hidden actions.

---

7See also Grunewald and Krakel (2017) for a related duopoly analysis.
Therefore, product review platforms – where participants’ identities are unverifiable – constitute the most natural setting to apply our signal-jamming approach. However, our setup can in principle be applied to any advertisement context where biases in messages are not perfectly observed by consumers. Such an interpretation connects this paper to the persuasive advertising literature. Part of this literature studies dissipative advertising where spending on advertising indirectly signals quality (e.g., Nelson, 1974, and Milgrom and Roberts, 1986). In our paper, bias is the result of hidden advertisement in the form of anonymous reviews. Therefore, consumers cannot observe the effort (or spending) by the firm and cannot make inferences based on that. Moreover, in contrast to some of the more recent papers in the advertising literature (e.g., Anderson and Renault, 2006, and Johnson and Myatt, 2006), we assume that the firm observes its product quality and conditions the bias level it will insert into product reviews on its quality.

3 Model

Consider a firm releasing a new product whose quality is unknown to consumers. There is a continuum of consumers with unit measure, and each consumer is indexed by $i \in [0, 1]$. The firm can be one of two types $j \in J = \{L, H\}$ based on the quality of the product it supplies. In particular, the $j$-type firm produces a product of quality $v_j$ such that $0 \leq v_L < v_H$. The marginal costs of production do not depend on product quality and are normalized to zero.

Consumers hold a common prior belief that they face an $L$-type firm with probability $Pr(v_L) = 1 - Pr(v_H) = g \in (0, 1)$. Before the decision to purchase, consumers visit various product review platforms to collect information about the product. Online shopping and product review websites such as Amazon, Yahoo, TripAdvisor or Yelp are good examples for such platforms. Users of online social media platforms such as Youtube, Instagram and Facebook are also exposed to product reviews. Finally, results from clinical trials funded and reported by companies or research conducted by independent institutions serve as

---

8Manipulation in our framework causes a rotation in the final demand curve as in Johnson and Myatt (2006). However, both the nature of the rotation and the channel through which it happens in our paper are quite different. In Johnson and Myatt (2006), the firm controls the precision of product information that is accessible to the consumers, but otherwise does not possess any more superior information than the consumers. In our framework, on the other hand, private information of the firm about product quality plays a key role. In particular, the implicit arms race induced by this private information is the main driver for demand rotation. See Figure 3 in subsection 4.2 and the discussion therein.

9See Renault (2016) for an overview of the recent literature on advertising.

10Our main results can be generalized –after an appropriate normalization– to the case where the high-quality firm has a higher marginal cost of production.
major information sources for consumers. We assume that the information that each consumer $i$ collects can be summarized by a private, noisy signal about product quality, which we denote by the random variable $x_i \in \mathbb{R}$.

The $j$-type firm observes the quality of its product and then exerts some effort to manipulate its reviews on the product review platforms. Costly activities such as hiring paid reviewers and online bots to create embellished product reviews, or funding research projects to produce favorable information about a product can all be part of this effort. The net outcome of these activities is summarized by a single non-negative number $b_j \geq 0$, which reflects the common bias in product review platforms that consumers are not able to detect. We model this bias as a uniform shift by $b_j$ in the mean of all signals that consumers receive. In particular, we assume that each private signal $x_i$ has three components: (i) the true value of the product, $v_j$; (ii) the bias inserted by the firm, $b_j$; (iii) the consumer-specific variation in the information collected, which we denote as $\varepsilon_i \in \mathbb{R}$ for each consumer $i$. We assume that the noisy signal $x_i$ is additively separable in these three components as follows:

$$x_i = v_j + b_j + \varepsilon_i.$$ 

Consumer-specific variation $\varepsilon_i$ is distributed independently across consumers and generated by a known cumulative distribution function $F$ and a corresponding density function $f$. Assumption 1 below states the restrictions we place on the noise distribution.

**Assumption 1** Density function $f$ for the idiosyncratic noise $\varepsilon_i$ is continuous, log-concave, symmetric around zero and unimodal with unbounded support and finite moments.

Upon observing her private signal, each consumer decides whether to buy one unit of the firm’s product or not. Ex-post utility of each consumer who purchases the product offered by the $j$-type firm is given by

$$\text{Utility} = v_j + \psi_i - p.$$ 

11Assuming idiosyncratic noise generates an informational heterogeneity among consumers. If we assume that all consumers face a common noise $\varepsilon$ in the information they receive so that all consumers receive the same signal, the market demand, in terms of share of consumers, would be either 0 or 1. From the firm’s perspective, each manipulative action would then correspond to a different distribution of signals that consumers may receive, therefore a different probability of full demand. All of our results with this probabilistic interpretation of demand would carry through in this alternative setup. An alternative way to introduce heterogeneity among consumers is to assume that preferences differ among consumers. Specifically, one can assume that consumer $i$’s payoff is $v_L + \psi_i - p$ when she purchases the good at price $p$. Here, $\psi_i$ is the individual match quality between the consumer and the product, and is generated by some c.d.f. over the support $(v_L, v_H)$. This specification would generate a downward-sloping demand curve for every public signal $x$. 

9
\[ u = v_j - p_j \]  

(1)

where \( p_j \) is the unit price announced by the firm and \( v_j \) is the product quality. If the product is not purchased, then \( u = 0 \). We assume in the benchmark model that both firm types charge the same price which is exogenously fixed at \( \bar{p} \in (v_L, v_H) \).\(^{12}\) Therefore, in this setting, price is not informative about quality. We discuss the model with endogenous and informative pricing in Section 5.

The noisy signals are the only source of information for consumers when they make a purchase decision. We denote the binary purchase decision of the consumer by a function \( s(x_i) \in \{0, 1\} \) such that \( s(x_i) = 1 \) if \( i \) buys the product. Therefore, from the perspective of the type-\( j \) firm, the total amount of sales can be written as

\[ S(v_j, b_j) = \int_0^1 s(x_i)di = \int_0^1 s(v_j + b_j + \varepsilon_i)di. \]  

(2)

The profit to the firm is given by

\[ \pi_j = \bar{p}S(v_j, b_j) - C(b_j), \]  

(3)

where \( C(.) \) is the cost associated with the bias \( b_j \). The source of this cost could be: (i) direct costs associated with hiring employees or fake reviewers as well as incentivizing consumers to promote the product; (ii) opportunity cost of time spent on online product review forums; (iii) cost of strategic research expenditures (either through funding research projects or directly conducting research in an R&D department of the firm); or (iv) expected fines or reputation costs if the firm is caught manipulating its product reviews (see Drugov and Troya-Martinez, 2015). If the firm spent little resources on manipulation, both the number of embellished reviews and the probability of someone detecting them would be small. Therefore, we assume that a small amount of bias \( b_j \) does not cost too much to the firm. However, as the firm exerts more manipulative effort, since both the direct costs of manipulation and the probability of detection increase at the same time, we assume that the incremental cost of increasing manipulative effort rises relatively fast. Assumptions 2 and 3 state the restrictions we place on the cost function.

\(^{12}\)We do not consider prices below \( v_L \) or above \( v_H \) since these prices lead to trivial demand levels such that either all or none of the consumers purchase the product.
Assumption 2 The cost function $C(.)$ satisfies $C'(0) = 0$, and $C'(b), C''(b) > 0$ for all $b > 0$.

Assumption 2 guarantees that whenever manipulation has a positive benefit in terms of higher sales, the firm exerts some effort to insert bias. We assume that manipulative effort does not involve any fixed costs. In Section 6.1 we discuss what happens if a regulating agency requires the firm to pay a fixed fee.

Assumption 3 $\min_{b \geq 0} C''(b) > p \max_{x \in \mathbb{R}} f'(x)$.

Assumption 3 imposes a lower bound on the convexity of the cost of manipulative effort, which guarantees that the profit function is concave in the level of bias. This assumption is useful in ruling out multiple equilibria which are qualitatively similar but impose different levels of manipulative effort. The following weaker Assumption 4 imposes an upper bound on the manipulative effort that the $L$-type firm can exert.

Assumption 4 The cost function and the quality levels satisfy the following inequality

$$C'(v_H - v_L) \geq v_H f(0).$$

The timeline of the game is illustrated in Figure 1.

![Figure 1: Timeline](image)

We employ perfect Bayesian equilibrium (PBE) as the solution concept for our analysis. Intuitively, PBE requires sequential rationality and Bayesian updating for posterior beliefs whenever possible. More formally, a strategy profile consists of the bias level chosen by each type of the firm $(b_L, b_H)$ and the purchasing decision $s(x)$ of each consumer after observing the noisy signal $x$. Then, a strategy profile

$$\langle b^*_L, b^*_H, s^*(\cdot) \rangle$$

accompanied with a posterior belief $\mu(v_j|x)$ that the quality is $v_j$ upon observing signal realization $x$ is a PBE if and only if
\[ b^*_L \in \arg \max_{b_L \in \mathbb{R}^+} [\bar{p} S(b_L, b^*_H, v_L) - C(b_L)] \]
\[ b^*_H \in \arg \max_{b_H \in \mathbb{R}^+} [\bar{p} S(b^*_L, b_H, v_H) - C(b_H)] \]

\[ S(b^*_L, b^*_H, v_j) = \int_{-\infty}^{\infty} s^*(x) f(x - v_j - b^*_j) dx \]
\[ s^*(x) = \begin{cases} 1 & \text{if } \sum_j v_j \mu(v_j|x) \geq \bar{p} \\ 0 & \text{otherwise} \end{cases} \]

3.1 Equilibrium analysis

First, note that consumer \( i \) buys the product if and only if her posterior expectation regarding the quality of the product is greater than the price, \( \mathbb{E}(v_j|x_i) \geq \bar{p} \). Therefore, market demand is determined by the distribution of signals and consumers’ interpretation of them. Assumption 1 imposes some regularity conditions on the posterior beliefs that consumers could have. In particular, consumers’ posterior expectation of quality is strictly increasing with their signal. Therefore, the purchasing decision admits a simple monotonic threshold structure. To show this, we first suppose that consumers follow a monotonic strategy and derive the manipulative effort choice of the firm. We then confirm that consumers’ decisions indeed satisfy our supposition.

Suppose that the firm expects consumers to adopt a monotonic purchase strategy such that consumer \( i \) purchases the product, \( s(x_i) = 1 \), if and only if her signal is higher than or equal to some threshold signal \( \bar{x} \), i.e., \( x_i \geq \bar{x} \). When the firm expects a purchase threshold \( \bar{x} \), the optimal bias level \( b_j \) solves

\[ \max_{b_j \geq 0} \bar{p}[1 - F(\bar{x} - v_j - b_j)] - C(b_j) \quad \text{for each } j \in \{L, H\}. \]  

The first-order condition to this problem is given by

\[ \mu(v_j|x) = \frac{Pr(x|v = v_j) Pr(v = v_j)}{Pr(x|v = v_L) g + Pr(x|v = v_H)(1 - g)}, \quad (4) \]
\[
\bar{p} f(x - v_j - b_j) = C'(b_j) \quad \text{for each } j \in \{L, H\}. \tag{6}
\]

Assumption 1 and Assumption 2 together ensure that an interior solution \( b_j > 0 \) to this problem exists for each \( j \). Moreover, if we further impose Assumption 3 we can guarantee that the solution is unique. Figure 2 illustrates how the bias levels \( b_L \) and \( b_H \) are set by the low and high-quality firm types when the price \( \bar{p} \) is below (left panel) and above the prior expected product quality (right panel), i.e., \( \bar{p} < \mathbb{E}(v) \) and \( \bar{p} > \mathbb{E}(v) \), respectively.\(^\text{13}\)

Given the manipulative effort level by each type of the firm, consumers form their equilibrium beliefs using Bayesian updating. When a consumer receives a signal \( x \), her posterior expectation of the product quality will be

\[
\mathbb{E}(v|x) = \frac{\sum_j v_j f(x - v_j - b_j) P(v = v_j)}{\sum_j f(x - v_j - b_j) P(v = v_j)}. \tag{7}
\]

Therefore, at a price \( \bar{p} \), a consumer will be indifferent between purchasing and not purchasing the product if and only if she receives a signal \( \bar{x} \) that satisfies

\[
\begin{align*}
\mathbb{E}(v|\bar{x}) &= \frac{\sum_j v_j f(\bar{x} - v_j - b_j) P(v = v_j)}{\sum_j f(\bar{x} - v_j - b_j) P(v = v_j)} = \bar{p} \iff \\
\sum_j (v_j - \bar{p}) f(\bar{x} - v_j - b_j) P(v = v_j) &= 0. \tag{8}
\end{align*}
\]

\(^\text{13}\)We assume that the noise follows a standard normal distribution, the cost function is quadratic and \( v_L = 0 \) for the calculations in Figure 2.
The three equations given in (6) and (8) determine the equilibrium with biased product reviews. In the absence of manipulation, consumers’ posterior expectation of the product quality conditional on signal $x$ would be

$$\mathbb{E}(v|x) = \frac{\sum_j v_j f(x - v_j)P(v = v_j)}{\sum_j f(x - v_j)P(v = v_j)}.$$

The corresponding purchase threshold signal $\bar{x}$ in the no-manipulation scenario is then implicitly determined by the following equality:

$$\mathbb{E}(v|x) = \frac{\sum_j v_j f(x - v_j)P(v = v_j)}{\sum_j f(x - v_j)P(v = v_j)} = \bar{p} \iff \sum_j (v_j - \bar{p}) f(x - v_j)P(v = v_j) = 0. \quad (9)$$

Consumers’ problem of estimating product quality is not trivial. A Bayesian consumer knows that the product reviews are manipulated by the firm. Thus, she has to adjust her posterior belief accordingly. By Assumption 1 if there were no bias in the signals, a higher signal would directly translate into a higher likelihood for the $H$-type firm. However, in the presence of bias, a higher signal could be driven either by a higher product quality or a higher manipulative effort by the firm. For the posterior expectation to be monotonic in the value of the observed signal, the quality difference should dominate the difference in manipulation, which requires a sufficient increase in the cost of bias compared to the response of consumers to higher signals. This way, a given value $\varepsilon$ of the noise would lead to a higher signal $x$ when the underlying quality is high, i.e., $v_H + b_H + \varepsilon > v_L + b_L + \varepsilon$. The following lemma states that both Assumption 3 and Assumption 4 are sufficient for that.\footnote{Note that Assumption 4 is a weaker condition than Assumption 3 as simple integration shows that Assumption 3 implies Assumption 4. The reason for introducing the stronger Assumption 3 is that some of the comparative statics results rely on the uniqueness of equilibrium bias levels for each price level.}

**Lemma 1** Suppose that Assumptions 1 and 2 hold, and that consumers use a unique, finite purchase threshold $\bar{x}$. Then, Assumption 3 as well as Assumption 4 implies that the mean signal about product quality is always higher when the firm is of high type than when it is of low type, i.e., $b_H + v_H > b_L + v_L$.

We present the proof for Lemma 1 and all other omitted proofs in Appendix A. The following lemma states that the posterior expectation of product quality is monotonically...
increasing with the value of the signal.

**Lemma 2** Suppose that Assumptions 1 and 2 and Assumption 4 hold and that \( b_L + v_L < b_H + v_H \). Then, \( \mathbb{E}(v|x) \) is strictly increasing in \( x \).

An immediate consequence of Lemma 2 is the existence as well as the uniqueness of a purchase threshold \( \bar{x} \) for every bias pair \((b_L, b_H)\) by the firm.

**Corollary 1** Suppose that Assumptions 1, 2 and 4 hold. Then, for each price level \( \bar{p} \in (v_L, v_H) \), there is a unique threshold \( \bar{x} \) with \( \mathbb{E}(v|x) = \bar{p} \) such that only those consumers with \( x_i \geq \bar{x} \) purchase the product.

Combining the first-order conditions in (6) and Corollary 1, we can establish the existence of an equilibrium with positive bias levels conditional on the price \( \bar{p} \). Theorem 1 below lays out the conditions for existence and uniqueness of such an equilibrium.

**Theorem 1** Suppose that Assumptions 1, 2 and 4 hold. Then, for each price level \( \bar{p} \in (v_L, v_H) \), there exists a PBE, characterized by strictly positive bias levels \( b_L > 0 \) and \( b_H > 0 \) and a threshold \( \bar{x} \), such that consumer \( i \) with signal \( x_i \) purchases the product if and only if \( x_i \geq \bar{x} \). Moreover, if Assumption 3 holds, the PBE is unique for each price level \( \bar{p} \in (v_L, v_H) \).

One of the important implications of Theorem 1 is that there is always some manipulative effort by the firm. The firm chooses to spend some of its resources to insert bias into product reviews regardless of its product quality. This result distinguishes our model from the advertising literature and most variants of false advertising models where usually only the lower types of the firm engage in advertising (see, for example, Rhodes and Wilson, 2017).

The following proposition states that the high-quality firm earns a higher profit in every equilibrium than the low-quality type even when the former exerts more effort to bias product reviews and thus faces higher manipulation costs.

**Proposition 1** The profit of the \( H \)-type firm is always higher than the profit of the \( L \)-type firm, i.e., \( \pi_H > \pi_L \).

This result stems from the quality advantage of the high-type firm. Given that consumers use a unique purchase threshold \( \bar{x} \) and face the same price under both quality realizations, it follows from Lemma 1 that the high-quality firm type will always make
more sales and generate more revenue than its low-quality counterpart. This is true because the mean signal under high quality lies above the mean signal under low quality regardless of the effort to bias the product reviews. But then the \(H\)-type firm should always earn a higher profit than the \(L\)-type because whenever this is not true, the former could raise its profit by exerting the same manipulative effort as the latter.

4 The effect of biased product reviews

In this section, we present and discuss our central results on the effects of biased product reviews. We start by illustrating how the bias levels chosen by the two types of the firm interact through consumer beliefs. This interaction exhibits an arms race that leads, from an ex-ante perspective, to inefficiently high amounts of manipulation. We next argue that the price plays a crucial role in determining the outcome of this arms race. The price governs the relative marginal benefits (competitive advantage) that accrue to each type from manipulation. As we demonstrate later in the section, without this price effect on competitive advantage, manipulative efforts would not differ across types and they would have no effect on sales and consumer surplus.

4.1 The implicit competition between firm types

Theorem 1 shows that both firm types exert manipulative effort. However, from an ex-ante perspective, this behavior may not be optimal for the firm. To see this, suppose that the equilibrium bias levels are \(b_L > b_H > 0\). Consider a consumer who was initially indifferent with a signal \(x = \bar{x}\). Now imagine for a moment that each type is forced to lower its bias by the same amount \(\min\{b_j\} = b_H\) so that the new bias levels are \(b_L - b_H\) and 0 for the \(L\)-type and \(H\)-type firm, respectively. Notice that a signal of \(\bar{x} - b_H\) would keep this consumer’s expected quality unchanged since, for any \(k\), \(f(\bar{x} - v_j - b_j) = f(x - k - v_j - (b_j - k))\) for \(j \in \{H, L\}\). But since price remains unchanged, the new purchase threshold \(\bar{x}\) must also go down by the same amount, i.e., \(\bar{x} = \bar{x} - b_H\). Formally,

\[
\frac{\sum_j v_j f(\bar{x} - \min\{b_j\} - v_j - (b_j - \min\{b_j\})) P(v = v_j)}{\sum_j f(\bar{x} - \min\{b_j\} - v_j - (b_j - \min\{b_j\})) P(v = v_j)} = \frac{\sum_j v_j f(\bar{x} - v_j - b_j) P(v = v_j)}{\sum_j f(\bar{x} - v_j - b_j) P(v = v_j)} = \bar{p}.
\]
Since the relative position of the purchase threshold vis-a-vis signal means remains unaltered, each type of the firm can make the same amount of sales with less manipulative effort, and hence at a lower cost. If such an adjustment in manipulation levels makes both firm types unambiguously better off, then why does the firm engage in excessive levels of manipulation?

The answer lies in the implicit competition between the two firm types. When the firm observes its quality level, it takes the effort choice of the other firm type as given. If the other type increases its manipulative effort, any given signal realization \( x \) will translate into a lower expected quality by consumers. Then the former quality type will need to increase its manipulative effort too in order to offset this adverse effect on consumer beliefs. In other words, the two types engage in an implicit arms race due to the negative externality their manipulative efforts cause for each other. Through more aggressive manipulation, each firm type can gain an advantage at the expense of the other type because consumers cannot make state-dependent adjustments in their expectations.

The price level determines the relative advantage of each type in the arms race. In particular, when the price is sufficiently low, the low-quality type enjoys a competitive advantage over the high-quality type in swaying consumer beliefs. As a result, it generates greater bias than the \( H \)-type. The opposite is true when the price is sufficiently high. The following proposition formalizes this result and specifies the price cutoff where the relative advantage switches from one type to the other.

**Proposition 2** The bias levels \( b_H = b_L \) if and only if \( \bar{p} = (1 - g)v_H + gv_L = \mathbb{E}(v) \). For every price level \( \bar{p} > (\leq) \mathbb{E}(v) \), \( b_H > (\leq) b_L \).

To understand the intuition behind this result, let us for a moment ignore the strategy of the firm and focus on consumers’ response to changes in the price. Keeping the firm’s actions fixed, the purchase threshold \( \bar{x} \) decreases as the price goes down. This is because the utility cost of purchasing the low-quality product goes down while the surplus to buying from the high-quality firm goes up. Specifically, when the price is lower than the prior expected quality \( \mathbb{E}(v) \), signal density for the low-quality state will be higher than the signal density for the high-quality state at the resulting purchase threshold. To see why, consider the following modification to the consumer indifference condition in (8):

\[
(v_H - \bar{p})(1 - g)f(\bar{x} - b_H - v_H) = (\bar{p} - v_L)gf(\bar{x} - b_L - v_L).
\]

When \((\bar{p} - v_L)g < (v_H - \bar{p})(1 - g)\) or equivalently \( \bar{p} < (1 - g)v_H + gv_L \), the threshold signal \( \bar{x} \) that makes consumers indifferent between purchasing and not purchasing the
product is such that

\[ f(\bar{x} - b_H - v_H) < f(\bar{x} - b_L - v_L), \]

which immediately implies by the first-order conditions in (6) that \( b_L > b_H \). This result reflects the fact that at the margin, \( L \)-type has more to gain from manipulation because it faces a greater mass of indifferent consumers. We reach the opposite conclusion when \( \bar{p} > \mathbb{E}(v) \).

While the direct price effect on revenues and hence the incentives to manipulate go in the same direction for both types, Proposition 2 demonstrates that the indirect effect, mediated through consumer beliefs, goes in opposite directions for each type.

4.2 The effect of biased reviews on sales and consumer surplus

Theorem 1 has established that both firm types exert some manipulative effort in every equilibrium, and Proposition 2 has shown how the price level influences manipulative efforts. However, neither of these results guarantees any effect on total sales. For example, by Proposition 2, both firm types insert the same level of bias into product reviews when the price coincides with the prior expected quality \( \mathbb{E}(v) \). Then, by equation (10), consumers’ posterior beliefs will be the same as when \( b_L = b_H = 0 \). That is, at this price, both types exert manipulative effort to bias the reviews, but these efforts have no effect on consumer beliefs nor on the market demand.

One important question our model sheds light on is when biased reviews help a given firm type increase its sales. To isolate the impact of information manipulation on the resulting sales, we fix the price and compare consumer demands under two equilibria: one where manipulative efforts of firms are unrestricted (manipulation equilibrium) and one in which manipulative effort is restricted to be 0 for both types (no-manipulation equilibrium). Specifically, we will say that type \( j \) does effective manipulation at price \( \bar{p} \) if

\[ 1 - F(\bar{x} - b_j - v_j) > 1 - F(\bar{x} - v_j) \iff b_j > \bar{x} - \bar{x}. \] (12)

Recall that threshold \( \bar{x} \) is the signal that makes a consumer indifferent between purchasing and not purchasing when there is no manipulation. It is uniquely defined by equation (9). Therefore, \( \bar{x} - \bar{x} \) is the average adjustment in consumers’ purchase threshold signal in response to the bias levels engaged by the two firm types. If type \( j \)'s bias level exceeds the average belief adjustment by consumers, then it means that consumers are not able to fully filter out this bias.
How biased reviews affect total sales by each type depends on the price level. This dependence in turn is related to how bias levels $b_L$ and $b_H$ compare to each other. Theorem 2 lays out the two central results of our paper. The first result is that the firm type that exerts more manipulative effort in equilibrium always improves its sales under manipulation relative to the no-manipulation case. Whether the firm with lower manipulation can also enjoy higher sales hinges on the price level. Our second result is that at sufficiently high or sufficiently low prices, both types can simultaneously increase sales relative to the no-manipulation benchmark. Otherwise, biased reviews push sales by each type in opposite directions, and the type with a lower manipulative effort experiences a decline in sales.

**Theorem 2** Suppose that Assumptions 1, 2 and 3 hold. Then there exist uniquely determined prices $p_L < (1-g)v_H + gv_L = \mathbb{E}(v)$ and $p_H > \mathbb{E}(v)$ such that $\bar{x} = v_L + (b_L + b_H)/2$ if and only if $\bar{p} = p_L$ and $\bar{x} = v_L + (b_L + b_H)/2$ if and only if $\bar{p} = p_H$. Furthermore, the effect of manipulation on the equilibrium demand can be summarized as follows:

i. Manipulation raises the demand for both types at the same time if and only if $\bar{p} < p_L$ or $\bar{p} > p_H$.

ii. Manipulation raises the demand for the $L$-type and lowers the demand for the $H$-type if and only if $p_L < \bar{p} < \mathbb{E}(v)$.

iii. Manipulation raises the demand for the $H$-type and lowers the demand for the $L$-type if and only if $\mathbb{E}(v) < \bar{p} < p_H$.

By Proposition 2 we know that at prices lower than $\mathbb{E}(v)$, the $L$-type firm exerts more manipulative effort. Since consumers cannot adjust their beliefs separately for each type of the firm, their adjustment in the purchase threshold signal, $\bar{x} - x$, falls short of the bias inserted by the $L$-type. As a result, $L$-type has a higher equilibrium demand in this region compared to the no-manipulation case. When the price level is smaller than $\mathbb{E}(v)$, the $H$-type firm loses its advantage in affecting the consumers more than the $L$-type, which implies that the sales for the $H$-type is less compared to the no-manipulation case. Theorem 2 states that for every price $p_L < \bar{p} < (1-g)v_H + gv_L$, only the $L$-type can achieve a higher demand.

To see the mechanism behind effective manipulation, consider the indifferent consumer in the no-manipulation scenario. This is the consumer who observes a signal realization $x = x$ and whose resulting posterior belief about quality is equal to $\bar{p}$. Now let us allow for
manipulation. Given the anticipated equilibrium bias levels $b_L$ and $b_H$, if the signal $\bar{x} + b_L$ induces this very consumer to purchase the product at the same price $\bar{p}$, then it means the $L$-type firm has become successful in effectively manipulating consumers’ beliefs in a way that increases its sales. This is so because if $\mathbb{E}(v|x + b_L) > \bar{p}$, then the new indifferent consumer must have observed a signal realization $\bar{x}$ that is strictly less than $x + b_L$. Note that a signal realization $x = \bar{x} + b_L$ under manipulation and $x = \bar{x}$ under no-manipulation are equally likely when the true quality is low. Hence, manipulation will induce a more favorable posterior belief if and only if it raises the likelihood of the former signal above the latter one when the true quality is high, i.e., $f(\bar{x} + b_L - b_H - v_H) > f(\bar{x} - v_H)$. When $\bar{p} < \mathbb{E}(v)$, this result follows easily because $b_L > b_H$ and $\bar{x} < (v_L + v_H)/2$. When $\bar{p} > \mathbb{E}(v)$, on the other hand, it holds only if $\bar{x}$ is sufficiently high because $b_H > b_L$. To be more precise, we need $\bar{x} > v_H + (b_H + b_L)/2$ for the result. Since $\bar{x}$ diverges to $\infty$ as $\bar{p}$ converges to $v_H$,$^{15}$ it then follows that this is true when $\bar{p}$ is sufficiently high, or more precisely when $\bar{p} > p_H$, where $p_H$ is as defined in Theorem 2.$^{16}$

An alternative way to interpret Theorem 2 is to visualize it in terms of how the demand curve changes after manipulation. The situation is depicted in Figure 3 where we plot the demand curves before and after manipulation for each firm type. In the figure, $D^0$ represents the demand curve before manipulation (i.e., in the no-manipulation case), so $D^0_j = 1 - F(\bar{x} - v_j)$. Similarly, $D^1$ represents the demand curve after manipulation, so $D^1_j = 1 - F(\bar{x} - v_j - a_j)$. As we can see from the figure, manipulation causes a rotation in the demand curve for both firm types. However, the particular way rotation happens and the reasons behind it are quite different than Johnson and Myatt (2006). Most importantly, the firm holds private information about product quality in our framework and this plays a key role.

If manipulation is effective (i.e., demand-increasing) for the $L$-type firm (case $ii$ in Theorem 2), biased reviews make consumers worse off since the share of consumers who end up with a negative surplus in the low-quality state increases while the share with a positive surplus in the high-quality state goes down. If manipulation is effective only for the $H$-type firm (case $iii$), then consumers are better off by a similar reasoning. To be more precise, we define the aggregate ex-ante consumer surplus under manipulation equilibrium as follows:

---

$^{15}$See Lemma 3 in Appendix A.

$^{16}$An alternative but equivalent way is to consider the indifferent consumer in the manipulation equilibrium and compare her behavior to when she receives the signal $\bar{x} - b_L$ in the no-manipulation case. Such an argument would give us the conditions for effective manipulation in terms of $\bar{x}$, which we do in Theorem 2.
CS_{ea}^M = (1 - g)(v_H - \bar{p})[1 - F(\bar{x} - b_H - v_H)] - g(\bar{p} - v_L)[1 - F(\bar{x} - b_L - v_L)]. \quad (13)

Similarly, let the ex-ante consumer surplus under no-manipulation case be defined as

CS_{ea}^N = (1 - g)(v_H - \bar{p})[1 - F(\bar{x} - v_H)] - g(\bar{p} - v_L)[1 - F(\bar{x} - v_L)]. \quad (14)

We define the net effect of manipulation on consumer surplus as the difference between $CS_{ea}^M$ and $CS_{ea}^N$. The following corollary to Theorem 2 shows that relatively lower prices are associated with a negative overall effect of biased reviews on ex-ante consumer surplus whereas higher prices are associated with a positive effect.

**Corollary 2** The net effect of manipulation on consumer surplus is negative when $p_L < \bar{p} < (1 - g)v_H + gv_L = \mathbb{E}(v)$ and positive when $\mathbb{E}(v) < \bar{p} < p_H$.

The ranges of the price level where the welfare effects are ambiguous are the ones in which manipulation is effective for both types; i.e., when $\bar{p} < p_L$ or $\bar{p} > p_H$ as in Theorem 2. In such cases, the net effect of biased reviews depends on the relative amount of manipulative effort by each type as well as the relative responsiveness of consumers to private signals. We provide a numerical example to show exactly how ex-ante consumer surplus changes with biased reviews. We assume normally distributed noise in product reviews, a quadratic cost function for manipulation and a uniform distribution for the...
prior beliefs such that the ex-ante expected quality is equal to 0.5. Figure 4 confirms that the net effect of manipulation on ex-ante consumer surplus is negative when the price lies below the ex-ante expected quality and positive when it lies above. However, the welfare loss in the former case ($\bar{p} < E(v)$) as well as the welfare gain in the latter case ($\bar{p} > E(v)$) diminishes as the price approaches its respective lower and upper bounds.

4.3 The trade-off between information and bias in reviews

Corollary 2 establishes that consumers may sometimes be better off if they simply received a noisy signal about product quality that is not biased through manipulated product reviews. However, it may often be infeasible—or too costly—to acquire unbiased information about the quality of a product. With the expansion of social media usage, popular social media platforms have become appealing channels of advertisement, both direct and hidden. As a result, social media users are exposed to many (embellished or sincere) product reviews and promotions on a daily basis. Therefore, it is certainly hard, if not impossible, to insulate oneself from fake product reviews that firms try to insert into the information flow going through these platforms. Yet even if a consumer somehow could ignore all the reviews and choose not to collect any information at all, it is still not
straightforward if she would want to do it. Translated into our framework, the question is if a consumer would ever prefer to merely rely upon her prior beliefs when making a purchasing decision instead of acting upon her posterior beliefs after a manipulated but nonetheless informative signal. The following proposition addresses this question.

**Proposition 3** Suppose that Assumptions 1, 2 and 4 hold. Then, a consumer would never ignore the only available signal $x_i = v_j + b_j + \varepsilon_i$ (where $j \in \{H, L\}$) and base her purchasing decision entirely on her prior beliefs.

Proposition 3 establishes that no matter how high manipulative effort by each type is, consumers are better off by visiting product review platforms. Hence, in our model, when a consumer chooses to browse an online review website to get information about a newly released product, exposure to the firm-induced bias is typically a price she rationally agrees to pay. Without any signal, a consumer always purchases the product when $\bar{p} \leq gv_L + (1-g)v_H = \mathbb{E}(v)$ and never purchases otherwise. Note that the ex-ante expected quality of the product $\mathbb{E}(v)$ is also the price threshold below which the $L$-type firm inserts greater bias than the $H$-type ($b_L > b_H$). Below this price, consumers overestimate quality when the firm is of low type, and so manipulation worsens consumer surplus. In contrast, when $\bar{p} > \mathbb{E}(v)$, biased reviews improve consumer surplus, but this is also when no consumer (acting merely on their prior beliefs) would purchase the product in the absence of any signal.

5 Endogenous price

In our benchmark model, we have concentrated on the indirect impact of the price level on equilibrium expectations since we assumed that price was exogenous and uninformative about quality. In that case, the price level has a critical role in the determination of equilibrium expectations and the ability of the firm (of either type) in manipulating consumer expectations. In this section, we drop the assumption that price is exogenous and allow the firm to strategically choose its price as well as manipulative effort level. We discuss below the main implications of such an extension for the effectiveness of manipulation and the equilibrium level of bias in product reviews. We present additional results on equilibrium refinements in Appendix B.

We employ perfect Bayesian equilibrium (PBE) as the solution concept in this section as well. We allow for mixed pricing strategies whose support lies in the interval $[v_L, v_H]$ to analyze any strategic information transmission from the firm to the consumers through
pricing. To simplify the notation throughout the analysis, we will assume that consumers’ prior beliefs assign equal probabilities to both firm types and that their purchasing decisions are symmetric\footnote{In particular, we assume that beliefs off the equilibrium path are symmetric across consumers. Off-equilibrium beliefs become important when the firm sets a price that was unanticipated by consumers. The concept of PBE does not impose any restrictions on how these beliefs are formed. Following Fudenberg and Tirole (1991), PBE that satisfy this requirement are sometimes referred to as “strong” PBE.}. A strategy profile consists of the bias level chosen by each type of the firm \((b_L, b_H)\), (possibly mixed) pricing decision \((\beta_L, \beta_H)\), the purchasing decision \(s(x, p)\) of each consumer after observing the noisy signal \(x\) and the price \(p\). Then, a strategy profile

\[
\langle b_L^*, b_H^*, \beta_L^*, \beta_H^*, s^*(\cdot, \cdot) \rangle
\]

accompanied with a posterior belief \(\mu(v_j|x, p)\) upon observing signal realization \(x\) and price \(p\) is a symmetric PBE if and only if

\[
\forall p \in \text{supp}(\beta^*_L) \quad b_L^*(p) \in \arg\max_{b_L \in [0, \infty)} pS(b_L, b_H, \beta_L^*, \beta_H^*, p, v_L) - C(b_L)
\]

\[
\forall p \in \text{supp}(\beta^*_H) \quad b_H^*(p) \in \arg\max_{b_H \in [0, \infty)} pS(b_L, b_H, \beta_L^*, \beta_H^*, p, v_H) - C(b_H)
\]

\[
\beta_L^* \in \arg\max_{\beta_L \in \Delta([v_L, v_H])} \int_{p \in \text{supp}(\beta_L)} pS(b_L^*, b_H^*, \beta_L, \beta_H^*, p, v_L) \beta_L(p) dp - C(b_L^*)
\]

\[
\beta_H^* \in \arg\max_{\beta_H \in \Delta([v_L, v_H])} \int_{p \in \text{supp}(\beta_H)} pS(b_L^*, b_H^*, \beta_L^*, \beta_H, p, v_H) \beta_H(p) dp - C(b_H^*),
\]

where the strategy and beliefs of consumers are defined as

\[
S(b_L^*, b_H^*, \beta_L^*, \beta_H^*, p, v_j) = \int_{-\infty}^{\infty} s^*(x, p) f(x - v_j - b_j^*) dx
\]

\[
s^*(x, p) = \begin{cases} 
1 & \text{if } \sum_j (v_j - p) \mu(v_j|x, p) \geq 0 \\
0 & \text{otherwise},
\end{cases}
\]

and the posterior belief \(\mu(v_j|x, p)\) is formed by Bayesian updating whenever possible.

When the firm gets to choose not only the bias level but also the price, it may set a price that conveys information about the quality of the product. The informativeness of the price depends on the equilibrium coordination of expectations. If consumers do not expect to learn anything from the price, they consult only to their private signals to
make inferences. Then, the firm does not have any incentives to signal quality through its price choice unless consumers interpret deviations from the prescribed equilibrium price as a deviation by the high-quality firm type. Such a strategy profile would resemble the exogenous price case that we have analyzed above. Indeed, Proposition 4 below shows that when $v_L = 0$, every price between 0 and $v_H$ can be supported as part of a pooling equilibrium in which price is uninformative.

**Proposition 4** Suppose that Assumptions 1, 2 and 4 hold. If $v_L = 0$, for every price $\bar{p} \in (0, v_H)$ there is a pooling PBE that supports $\bar{p}$ as the equilibrium price. The equilibrium strategies, except for the price, are determined by the equilibrium conditions laid out in Theorem 1. If $v_L > 0$, prices that are close enough to $v_H$ cannot be supported by a pooling PBE.

Proposition 4 implies that the analysis we did with the exogenous price assumption can be extended to the endogenous price specification. In particular, our findings in the benchmark model regarding the effects of the price level can be considered as a comparison across different pooling PBE that arise when price is endogenous. When $v_L > 0$, the firm has an “outside option” of setting $\bar{p} = v_L$ and enjoying full market demand that yields a positive profit. Therefore, the firm prefers to take this outside option when the pooling price is close enough to $v_H$ such that it causes significantly reduced sales and a lower profit.

Allowing for strategic pricing by the firm adds another layer on the impact of price on equilibrium expectations that may not be captured by pooling equilibria. The firm may directly use price level as an additional tool for managing the expectations of consumers. For example, the high-type firm might consider using the price as a separating signal to convey its quality level to the consumers. However, as long as there is such a price that is believed by consumers to signal high quality, the low-type firm would simply charge the same price and pretend to be a high-type, thereby rendering separation infeasible. This result is stated in Proposition 5 below.

**Proposition 5** If $v_L > 0$, there are no pure strategy separating PBE. If $v_L = 0$, there are pure strategy separating PBE, but in all of them the revenue of the high quality type is 0.\(^{18}\)

Non-existence of a separating equilibrium implies that a pure pricing strategy cannot be fully informative. However, when we allow the firm to choose a probabilistic pricing...
strategy, the low-type firm may adopt a mixed pricing rule that mimics the hypothetical behavior of the high-type with some probability. Theorem 3 below describes such equilibria and shows that relatively higher prices can be supported by partially-separating mixed equilibria.

**Theorem 3** Suppose that Assumptions 1, 2 and 4 hold. For each price level $\bar{p}$ above some threshold $\bar{p}$, there exists a partially separating PBE. In this PBE, the high-quality firm type chooses with pure strategy the price-bias pair $(p, b) = (\bar{p}, b_H)$ where $b_H > 0$. The low-quality firm type plays a binary mixed strategy that assigns probability $\bar{\alpha}(\bar{p})$ to $(p, b) = (\bar{p}, b_L)$ where $b_L > 0$, and probability $1 - \bar{\alpha}$ to $(p, b) = (v_L, 0)$. Each consumer $i$ with a signal $x_i$ purchases the product either when the observed price is $v_L$ or when the signal $x_i \geq \bar{x}$ for some signal threshold $\bar{x}$.

In this PBE, the mixing probability $\bar{\alpha}$, the bias levels $b_L$ and $b_H$, and the signal threshold $\bar{x}$ satisfy the following system of equations:

$$v_L = \bar{p}(1 - F(\bar{x} - b_L - v_L)) - C(b_L)$$  
(17)

$$\bar{p}f(\bar{x} - b_L - v_L) = C'(b_L)$$  
(18)

$$\bar{p}f(\bar{x} - b_H - v_H) = C'(b_H)$$  
(19)

$$(v_H - \bar{p})f(\bar{x} - b_H - v_H) = \bar{\alpha}(\bar{p} - v_L)f(\bar{x} - b_L - v_L).$$  
(20)

In a partially separating pricing equilibrium, the $L$-type firm engages in “false advertising” to convince the consumers to purchase the product. Such a strategy makes pricing partially informative for consumers. When they observe a price that the $H$-type firm would normally choose, they consider the possibility that the $L$-type might have mimicked the $H$-type with some probability, and therefore reach an updated belief about the firm’s type. Rhodes and Wilson (2017) consider a similar false-advertising framework, where the low-type firm uses a misleading mixed strategy. In our model, consumers do not constrain themselves merely to the information they learn from the price since they also have access to biased information through product review platforms. Therefore, the channels through which the price level indirectly affects the equilibrium behavior continue to exist and moreover interact in interesting ways with the directly informative role that price takes on in this type of equilibria.

In each partially separating equilibrium, the price level has a different level of informativeness. Figure 5 shows that the mixing probability $\bar{\alpha}$ declines as price $\bar{p}$ increases,
meaning that price becomes more informative in equilibria with higher prices. Indeed, inspecting equation (20) reveals that a higher equilibrium price $\bar{p}$, all else equal, reduces the probability with which the L-type firm chooses $\bar{p}$. Lower $\bar{\alpha}$ makes it easier for consumers to be convinced that the firm is of high type, and this in turn increases the demand. Therefore, a higher price along with the resulting higher demand increases the revenue for both types of the firm. Since, as implied by equation (17), the L-type firm’s profit stays constant over all partially separating equilibria, a higher revenue for the L-type firm results in higher spending on manipulation. Proposition 6 below confirms this intuition by showing that the equilibrium bias level by the L-type firm increases with the price level.

**Proposition 6** Consider the set of partially separating PBE indexed by prices $\bar{p}$ over the interval $(\bar{p}, v_H)$, as described in Theorem 3. In these PBE, the purchase threshold signal $\bar{x}$, bias level $b_L$ by the L-type firm and the profit earned by the H-type firm increase with $\bar{p}$, i.e., $\partial \bar{x}/\partial \bar{p}$, $\partial b_L/\partial \bar{p}$ and $\partial \pi_H/\partial \bar{p} > 0$.

In addition to its direct effect on manipulation through signaling, the price level also governs the terms of the implicit competition between the two types of the firm. The
Figure 6: Bias levels as price increases

mechanism behind this role of price is similar to the one in the uninformative price case. As
the price level approaches $v_H$, the $H$-type firm is more effective in influencing consumers
with manipulative effort. Like in the benchmark model, the $H$-type firm exerts more
manipulative effort than the $L$-type at sufficiently high prices. Figure 6 illustrates the
relationship between manipulation levels and the price level over partially separating PBE.

In our benchmark model, we have measured effectiveness of manipulation at each
fixed price by comparing the equilibrium demand the firm faces when it can bias product
reviews to the demand it faces under no manipulation. Since biasing the product reviews
is the only viable channel of manipulation for the firm in the benchmark model, this
comparison gave us a natural measure for the effectiveness of manipulation. However, in
the endogenous price specification that we consider in this section, there are two channels
through which the firm can manipulate consumer beliefs, namely through biased product
reviews and the price. Moreover, the mixing probability $\bar{\alpha}$ is an endogenous variable
that is determined in equilibrium together with $\bar{p}$ and the bias levels $b_L$ and $b_H$. As can
be seen from equation (20), when $b_L$ and $b_H$ change, $\bar{\alpha}$ endogenously adjusts even if we
hold $\bar{p}$ constant. In other words, if we exogenously set the bias levels equal to zero while
holding the price level fixed at the same value, not only the consumers will use a different
purchase threshold signal \( x \neq \bar{x} \), but also the mixing probability \( \bar{\alpha} \) that the \( L \)-type firm uses in its pricing strategy will change to a different value, which we can denote as \( \alpha \). Therefore, when we say \( j \)-type firm’s product review manipulation is effective, we mean that the marginal contribution of \( j \)-type firm’s manipulative effort to its ex-post demand is positive compared to the corresponding partially separating equilibrium at the same price where the bias levels are set to be zero.

When the firm is not allowed to do manipulation, consumers update their beliefs based on their (unbiased) private signals and the price. If they observe a price \( p = v_L \), they are certain that the product quality is low. When they observe \( p = \bar{p} \), they remain uncertain about the quality. The equilibrium conditions for a partially separating PBE when manipulation is not allowed are as follows:

\[
v_L = \bar{p}(1 - F(\bar{x} - v_L)) \quad (21)
\]

\[
(v_H - \bar{p})f(\bar{x} - v_H) = \alpha(\bar{p} - v_L)f(\bar{x} - v_L), \quad (22)
\]

where \( \alpha \) is the mixing probability with which the \( L \)-type firm chooses \( \bar{p} \). It is straightforward to show that for relatively high prices that are close enough to \( v_H \), a partially separating PBE without manipulation described by equations (21) and (22) exists. This is an assumption we maintain in this section when we measure the effectiveness of manipulation.

To show the impact of the firm’s manipulative effort, we take some given price \( \bar{p} \) and compare the firm’s sales when there is manipulation and when there is not. In particular, we say that type-\( j \) firm’s manipulation is effective if

\[
1 - F(\bar{x} - b_j - v_j) > 1 - F(\bar{x} - v_j) \iff b_j > \bar{x} - \bar{x}.
\]

The following proposition characterizes when the \( H \)-type firm exerts more manipulative effort than the \( L \)-type firm in equilibrium, and when manipulation is effective for each type.

**Proposition 7** The bias levels \( b_H = b_L \) if and only if

\[
\bar{p} = \frac{v_H + \bar{\alpha} v_L}{1 + \bar{\alpha}} =: \hat{p}_{\alpha}.
\]

For every price level \( \bar{p} > (\leq) \hat{p}_{\alpha} \), \( b_H > (\leq) b_L \). Manipulation is always effective for the \( L \)-type firm and is effective for the \( H \)-type when \( \bar{p} \geq \hat{p}_{\alpha} \).
Thus, as in the benchmark model, the firm type that exerts more manipulative effort is able to achieve effective manipulation in the endogenous price case. However, the $L$-type firm’s equilibrium manipulation is always effective now, whereas the $H$-type firm achieves effective manipulation only when $\bar{p} > \hat{p}_\alpha$. With endogenous pricing, the $L$-type firm can influence consumer beliefs through not only its manipulation level but also the probability it assigns to the higher price. Therefore, even for extremely high prices, the $L$-type is able to use a combination of these two channels to tilt the market demand.

The fact that $L$-type can always do effective manipulation in equilibrium makes biased reviews potentially more harmful under partially separating pricing equilibria than under pooling equilibria. Indeed, our numerical calculations show that, after fixing the price level, manipulation of product reviews reduces ex-ante consumer surplus compared to the case in which the product reviews are not biased. This is illustrated in Figure 7.

6 Extensions

In this section, we present three extensions to the benchmark model. First, we discuss how a regulating agency could intervene to reduce harmful manipulation. We show that
a rather simple policy, a fixed cost for manipulation, can work effectively in increasing consumer welfare. Second, we modify our original model to see how a firm would behave if it could commit to a manipulation plan prior to observing its type. The goal of this extension is to see how our qualitative results hinge on the absence of commitment in the benchmark model. The final extension is a slight modification of the benchmark model to demonstrate the impact of an exogenous variation in signal precision on the equilibrium levels of manipulative effort.

6.1 Regulating manipulation via fixed cost

There are various ways a regulator might want to intervene in order to increase consumer surplus or any other measure of welfare. There are two broad categories of policy intervention based on the information requirements of the regulator.

If it is feasible for the regulator to collect information about the quality of the product, then it can target the low-quality firm directly. One example of such a policy is sampling consumer complaints and devising a fine schedule for the firm based on some summary statistics of the complaints (see Drugov and Troya-Martinez, 2015, for an analysis of such a policy). In such a case, the low-quality firm type might face additional costs if it engages in manipulation. This leads to a reduced level of manipulative effort engaged by the low-type firm, which in turn induces an increase in the consumer surplus. A formal analysis of such a policy could be embedded into our analysis by introducing an additional marginal cost of manipulation for the low-type firm.

A more crude intervention is to monitor product quality and impose quality standards that drives the low-type firm out of the market. Such a policy would solve the asymmetric information problem in the market because the only firm type remaining in the market would be the high-type. Then the game would turn into a simple bargaining game between the firm and the consumers. If the price remains at the same level as it was prior to the regulation, the consumer surplus would increase. However, if the firm gets to choose its price after such a regulation (as we allow in Section 5), sequential rationality would imply that the firm would be able to expropriate all of the consumer surplus by setting the highest possible price, \( v_H \). Therefore, such a policy does not necessarily benefit the consumers especially when embellished product reviews raise consumer surplus, as illustrated by Figure 4.

One disadvantage of policies that require information acquisition by the regulatory agency is that information acquisition is potentially costly and therefore might cause additional burdens to consumers through taxation. Any policy that does not require
costly information acquisition by the regulatory agency will have an effect on both types of the firm. The challenge for the regulatory agency is that any policy that reduces manipulative effort by the low-quality type might also harm the high-quality firm. The welfare implications of such an intervention is not straightforward then. We discuss below the policy of introducing a fixed toll for producing embellished reviews. We argue that even if such a policy affects both types, it has a more profound effect on the low-quality type.

**Introducing a fixed cost for manipulation**

Suppose the regulator introduces a fixed fee for producing biased product reviews. The implementation of such a policy is not straightforward since manipulative efforts are not observed – which is a prerequisite for effective manipulation in the first place. One possible solution is to require the review platforms to charge a fixed fee to those firms who want the reviews of their products to appear on the platform. As an example for such a business practice, online shopping sites typically charge a fixed entry fee to sellers for a space on their website. Since we expect in equilibrium that all sellers, who offer vertically differentiated products, will exert some manipulative effort, such a policy would have direct effects on the incentives to manipulate. Alternatively, popular social media platforms, such as Facebook and Twitter, might ban all of the content related to a firm’s product unless the firm pays a fixed fee to allow such content.

To formally assess the merits of a fixed cost policy, let \( \bar{c} \) be the fixed fee that a product review platform charges the firm for enabling product reviews. Then the firm would pay \( \bar{c} \) in addition to the variable cost \( C(b_j) \) for any positive manipulative effort. Addition of \( \bar{c} \) would not alter the firm’s preference ordering over positive manipulative effort levels but rather potentially deter the firm from exerting any effort at all. If the \( j \)-type firm chooses to exert effort to bias reviews, its net profit would be

\[
\bar{p}(1 - F(\bar{x} - b_j - v_j)) - \bar{c} - C(b_j).
\]

If the firm opts out and chooses not to bias reviews, then it will earn

\[
\bar{p}(1 - F(\bar{x} - v_j)).
\]

Note that the first-order conditions for an interior solution may not identify the optimal
bias decision of the firm in all of the cases. When the fixed cost $\bar{c}$ is exceedingly high compared to the profit gains by manipulative effort, the optimal decision of the firm will be to exert no manipulative effort – which is a corner solution. By Proposition 1 we know that the $H$-type firm always earns a higher profit than the $L$-type firm. Therefore, if the $H$-type firm chooses not to exert any manipulative effort at some given price because $\bar{c}$ is too high, then the $L$-type would not exert any manipulative effort at that price either. The reverse is not true. There could be price levels for which the $H$-type exerts some manipulative effort but the $L$-type does not. Indeed, it is straightforward to show the existence of such price levels. This is what we do in the next proposition.

**Proposition 8** There exists a fixed cost level $\bar{c} > 0$ such that the $L$-type firm does not exert any manipulative effort for any price level while the $H$-type exerts positive manipulative effort $b_H > 0$ for a non-empty set of prices.

We omit the proof of Proposition 8 since the result directly follows from Proposition 1 and Theorem 1. The numerical calculations in Figure 8 illustrate this result for normally distributed signals and a quadratic cost function. In the left panel, we illustrate a situation where the fixed cost is binding only for sufficiently high price levels. When prices are low enough, both types exert effort to manipulate. Importantly, there is a price threshold beyond which the $L$-type no longer exerts effort although the $H$-type continues to do so. In the right panel of Figure 8 we illustrate a case where the fixed cost is so high that the $L$-type does not exert any effort regardless of the price level while the $H$-type chooses to manipulate for a wide range of prices.

![Figure 8: Manipulation levels when there is fixed cost](image)

Figure 9 illustrates the impact of a fixed cost policy on consumer surplus. It depicts the resulting change in consumer surplus as a function of the fixed cost level. In the
left panel, we plot the change in consumer surplus when the price is low while in the right panel we consider a higher price level. When the fixed cost is high enough, the bias levels are zero for both types, and so the net effect of the fixed cost on consumer surplus depends on the price level. Comparing the two panels, one can see that the magnitude of the change in consumer surplus is substantially higher under a low price than a high price. The reason is that manipulation raises demand for the $H$-type when the price is high enough. Therefore, any policy that drives down manipulative effort by both types will be less beneficial for consumers when prices are high. In conclusion, the optimal fixed cost level should be high enough to deter the low-quality firm but not too high to deter the high-quality firm from exerting manipulative effort.

![Figure 9: Effect of increasing fixed cost on consumer surplus](image)

We have discussed the impact of introducing a fixed cost when prices are uninformative. The implications will be similar when prices are endogenous and informative. One can show that there is still a threshold fixed cost level beyond which the $L$-type firm will stop exerting manipulative effort to bias product reviews. However, in the presence of a fixed cost, the $L$-type firm will continue using mixed pricing as a false advertising strategy.

### 6.2 Ex-ante choice of the bias in product reviews

We have assumed so far that the firm chooses the bias level in product reviews after it observes the quality of its product. This assumption is more relevant for the cases where firms invest in production before making decisions about marketing. However, analyzing the ex-ante choice of bias provides further insights about the strategic constraints a firm faces in our benchmark model without commitment. Moreover, the analysis of manipulation behavior of the firm when there is commitment enables us to compare our model to
Bayesian persuasion models in the literature, where the player who chooses the information structure is commonly assumed to have full commitment power (e.g., Kamenica and Gentzkow, 2011).

Suppose that the firm announces a state-contingent manipulation plan \((b_L, b_H) \in [0, \infty]^2\) given a price level \(\bar{p}\) and commits to it after observing the quality of the product. Consumers observe the plan but not the quality level. They receive a noisy private signal about quality and decide whether to purchase the product or not.

There is no change in consumers’ response to biased signals. Therefore, Corollary 1 extends to the ex-ante choice as well. For every state-contingent plan, there is a unique symmetric response by consumers, which can be characterized by the purchase threshold signal \(\bar{x}\). The equilibrium \(\bar{x}\) satisfies equation (9), which we replicate here for convenience:

$$
\frac{v_H(1-g)f(\bar{x} - v_H - b_H)}{(1-g)f(\bar{x} - v_H - b_H) + gf(\bar{x} - b_L - v_L)} = \bar{p}.
$$

The Implicit Function Theorem provides us with a continuously differentiable function \(\bar{x}(b_L, b_H)\) for consumers’ response. This in turn guarantees the existence of a symmetric pure-strategy equilibrium where \(\bar{x} \to \infty\) as \(\bar{p} \to v_H\) and \(\bar{x} \to -\infty\) as \(\bar{p} \to v_L\).

Firm’s profit maximization is equivalent to maximizing the following summation of the ex-post profits:

$$
(1-g)(\bar{p}(1 - F(\bar{x} - b_H - v_H)) - C(b_H)) + g(\bar{p}(1 - F(\bar{x} - b_L - v_L)) - C(b_L)).
$$

First-order conditions for an interior solution are exactly as the ones in the benchmark model. However, as will be established in the next proposition, the local maximum implied by these conditions is not a global maximum for the ex-ante profit of the firm. Instead, the profit-maximizing plan requires a corner solution, where at least one of the two effort levels must be zero.

**Proposition 9** Suppose that Assumptions 1, 2 and 4 hold. If the firm can ex ante commit to a particular manipulation plan that is contingent on product quality, the optimal plan will never feature strictly positive effort levels for both quality types.

Since the firm moves first and announces a plan that it can commit to, it enjoys the strategic advantage of a Stackelberg-leader. Its optimal plan takes into account how consumers would respond to different plans. This lies in contrast to the benchmark model where the information processing by consumers and the bias decision of the firm were
The firm knows that the consumers adjust their expectations for manipulation. Thus, if the firm reduces the bias levels uniformly without changing the difference between the two, consumers will respond by lowering their expectations for manipulation. This result follows from a calculation that is similar to the one presented in equation (10). Consequently, through a revealed preference argument, one can show that one of the bias levels will be equal to 0 in any equilibrium.

In contrast to the benchmark model with ex-post manipulation, here we cannot analytically rule out zero bias in both quality states as an equilibrium outcome. In some cases, the firm prefers simply not to intervene with the product reviews received by consumers. In particular, when \( \bar{p} = \mathbb{E}(v) \), consumer indifference condition (24) requires \( f(\bar{x} - b_L - v_L) = f(\bar{x} - b_H - v_H) \), which also implies that \( f'(\bar{x} - b_L - v_L) = -f'(\bar{x} - b_H - v_H) \). Hence, neither the plan \((b_L, 0)\) nor the plan \((0, b_H)\) leads to a positive marginal benefit. Therefore, the firm commits to zero manipulation for both states when \( \bar{p} = \mathbb{E}(v) \).

Implications of ex-ante commitment on manipulation are twofold when the price is endogenous. First, the firm would again choose zero manipulation for at least one of the quality types. Second, the firm would be able to exploit the interaction between price-signaling and biased product reviews more efficiently. In particular, the firm would choose to concentrate more on price-signaling when the quality of the product is low and on biased product reviews when the quality is high.

### 6.3 Information precision

The dispersion of consumers’ private information depends negatively on the precision of the noise component \( \varepsilon \) in their signals. Signal precision has two effects on consumers’ inferences. The first effect concerns the informativeness of product reviews about quality. The more precise consumers’ signals are, the more informative they become about the underlying quality of the product. Therefore, in principle, it should get harder to deceive consumers when precision is high. We call this the ‘accuracy effect’ and, by itself, it should discourage biased reviews. The second effect concerns how the resulting change in the distribution of consumer signals affects the firm’s incentives to exert manipulative effort. When the precision is high, consumers concentrate more around the mean signal of each state and receive signals that are closer to each other. As a result, manipulative effort of the firm can sway a greater mass of consumers. This ‘concentration effect’ increases the
marginal revenue of manipulation and thus should raise the equilibrium levels of bias by each type. To quantify these counteracting effects of precision, we adopt the following modification to the private signal structure in the benchmark model. Given the quality level \( v_j \) and the bias level \( b_j \) for \( j \in \{H, L\} \), each consumer \( i \) receives the private signal

\[
x_i = v_j + b_j + \sigma \varepsilon_i,
\]

where \( \sigma > 0 \) is the dispersion parameter. The higher \( \sigma \) is, the higher is the dispersion among consumers’ private information on quality. We can describe the equilibria indexed by price levels \( \bar{p} \in (v_L, v_H) \) by following the same steps as before, only with slight changes in the notation. The sales quantity of the firm, given the signal threshold \( \bar{x} \), is the probability that a consumer receives a signal that is greater than \( \bar{x} \). That is,

\[
Pr(x_i \geq \bar{x}|v_j) = Pr\left(\varepsilon_i \geq \frac{\bar{x} - v_j - b_j}{\sigma}\right) = 1 - F\left(\frac{\bar{x} - v_j - b_j}{\sigma}\right).
\]

Similar to the benchmark model, the interior solution to the maximization problem of the firm always exists. The first-order condition for each type \( j \) is

\[
\bar{p} f\left(\frac{\bar{x} - v_j - b_j}{\sigma}\right) = \sigma C'(b_j),
\]

and the indifference condition for consumers that equates the expected quality to the price level is

\[
(v_H - \bar{p})(1 - g) f\left(\frac{\bar{x} - v_H - b_H}{\sigma}\right) = \bar{p} g f\left(\frac{\bar{x} - v_L - b_L}{\sigma}\right).
\]

Figure 10 illustrates the effects of signal precision on the equilibrium levels of manipulative efforts \( b_L \) and \( b_H \) under two different levels of \( \bar{p} \). The left panel depicts simulation results when the price is low enough so that \( b_L > b_H \). The right panel considers the opposite scenario where \( b_H > b_L \). In either case, as we raise the precision of the private signals (i.e., lower the signal dispersion \( \sigma \)), we first see a gradual increase in the equilibrium bias introduced by both types. In this range of precision, it appears that the concentration effect dominates the accuracy effect. However, as signal precision increases further, the accuracy effect starts to dominate. When \( \sigma \) is close to zero, consumers receive extremely informative signals about the quality for high precision values. Therefore, for every price, either almost all of the consumers decide to purchase the product or almost none of them purchases. Manipulation effort becomes irrelevant, and hence, is reduced to zero.
7 Conclusion

In this paper, we offer a model of information manipulation where firms can bias the information consumers collect through product review platforms. We study the implications of such manipulation when there is asymmetric information between the firm and consumers regarding the quality of the product and the manipulative effort of the firm is not observable to consumers. We show that the ability to increase demand via manipulative effort is not reserved only to one firm type. Depending on the price level, both firm types can be effective manipulators. Under relatively low prices, the low-quality firm has more incentives to bias product reviews, and the signal-jamming technology benefits the sales of the low-quality product more than the high-quality product. The opposite is true when prices are relatively high. The incentive of each firm type to exert manipulative effort is exacerbated due to an arms race between the two types. While manipulation might be harmful or beneficial for consumers depending on the context, the implicit arms race the firms engage in always leads to socially wasteful spending that would be avoided if firms could commit to a particular bias plan.

We further analyze the case where the firm can strategically use both price-signaling and product review manipulation. We show the existence of partially separating equilibria where the low-quality firm chooses a mixed strategy between setting a low price and mimicking the high-quality firm by setting a high price. This makes the price informative about the quality for the consumers. We find that both types of the firm still engage in product review manipulation, and the price continues to determine the terms of the implicit arms race between the two types. Moreover, product review manipulation is still effective in terms of raising the firm’s demand. In particular, holding the high-quality
firm’s equilibrium price constant, we show that the low-quality firm is always able to increase its demand by manipulating product reviews compared to the corresponding partially separating price equilibrium where there is no product review manipulation.

Our paper provides a novel rationale for regulating firms’ access to product review platforms. In particular, we show that there is an optimal fixed fee that platforms can charge firms for enabling reviews of their product. This fee helps the platform filter out the harmful manipulative effort from a beneficial one. A fixed fee for accessing the product review platforms is one of the simplest regulations. There are more sophisticated policy tools for regulation, such as quality control, customer protection and so on. However, these could complicate the information structure by increasing the number of public and private signals accessible by the consumers.

A Proofs for the results in the main text

Proof of Lemma 1 Suppose first that Assumption 3 holds. Then, each of the first-order conditions in (6) has a unique solution since Assumption 3 guarantees the strict concavity of the profit function globally.

Suppose for a moment that \( b_L + v_L \geq b_H + v_H \), which would imply that \( \bar{x} - b_L - v_L \leq \bar{x} - b_H - v_H \) and also \( b_L > b_H \). Then having a unique solution to equation (6) implies that if the \( L \)-type uses the lower bias level \( b_H \), its marginal revenue should be greater than its marginal cost. That is,

\[
\bar{p}f(\bar{x} - b_H - v_L) > C(b_H) = \bar{p}f(\bar{x} - b_H - v_H),
\]

which implies

\[
|\bar{x} - b_H - v_L| < |\bar{x} - b_H - v_H|,
\]

since the noise pdf \( f(\cdot) \) is unimodal. Then, \( v_L < v_H \) implies that \( \bar{x} > b_H + v_H \). On the other hand, \( b_L > b_H \) implies by the first-order conditions (6) that

\[
f(\bar{x} - b_L - v_L) > f(\bar{x} - b_H - v_H) \iff |\bar{x} - b_L - v_L| < |\bar{x} - b_H - v_H|.
\]

But this in turn implies that \( \bar{x} < b_H + v_H \). Hence, we reach a contradiction.

Now, suppose that Assumption 4 holds and \( b_L + v_L \geq b_H + v_H \iff b_L > b_H + v_H - v_L \). But at such a manipulative effort level by the \( L \)-type, the marginal cost would always be greater than the marginal revenue by Assumption 4.
Proof of Lemma 2 For notational simplicity, denote \( f_j \equiv f(x - v_j - b_j) \) and \( f'_j \equiv f'(x - v_j - b_j) \). Then,

\[
\frac{\partial E[v|x]}{\partial x} = \frac{\sum_{j \in J} v_j f'_j P(v = v_j) - \sum_{j \in J} f'_j P(v = v_j) \left( \sum_{j \in J} v_j f_j P(v = v_j) \right)}{\left( \sum_{j \in J} f_j P(v = v_j) \right)^2} > 0 \Leftrightarrow
\]

\[
\sum_{j \in J} v_j f'_j P(v = v_j) \left( \sum_{j \in J} f_j P(v = v_j) \right) - \sum_{j \in J} f'_j P(v = v_j) \left( \sum_{j \in J} v_j f_j P(v = v_j) \right) > 0 \Leftrightarrow
\]

\[
g(1 - g) f'_H f_L > g(1 - g) f_H f'_L \Leftrightarrow \frac{f'(\bar{x} - b_L - v_H)}{f(\bar{x} - b_H - v_H)} \geq \frac{f'(\bar{x} - b_L)}{f(\bar{x} - b_L)}, \quad (26)
\]

which holds since \( f \) is log-concave by Assumption 1 and \( b_H + v_H > b_L + v_L \) by Lemma 1.

Proof of Theorem 1 By Assumption 4 and the Inverse Function Theorem, for each bias pair \((b_L, b_H)\) there exists a unique \( \bar{x} \), and \( \bar{x}(b_L, b_H) \) is a continuously differentiable function. Therefore we can reduce the number of equations that define the equilibrium into the following two equations that are very similar to the first-order conditions in (6):

\[
\bar{p} f(\bar{x}(b_L, b_H) - v_j - b_j) = C'(b_j) \quad \text{for all } j \in J, \quad (27)
\]

which has a positive solution by Assumptions 1 and 2 and the Intermediate Value Theorem. Moreover, the solution is unique if Assumption 3 also holds.

Proof of Proposition 1 The proof follows from a simple revealed-preference argument. The profit level of the \( L \)-type is

\[
\bar{p}(1 - F(\bar{x} - b_L)) - C(b_L).
\]

If the \( H \)-type imitated the \( L \)-type, its profit would be

\[
\bar{p}(1 - F(\bar{x} - b_L - v_H)) - C(b_L) > \bar{p}(1 - F(\bar{x} - b_L)) - C(b_L),
\]
which implies the optimal profit level that the $H$-type can achieve is strictly higher than that of the $L$-type.

**Proof of Proposition 2** The proof follows from the inspection of the consumer indifference condition (8). When $\bar{p} = E(v)$, equation (8) implies that marginal revenues from manipulation for both types are equal to each other, which implies that equilibrium bias levels for both types of the firm equal to each other. By a similar argument, the marginal revenue of manipulation for $H$-type is higher than $L$-type if and only if $\bar{p} > E(v)$.

**Lemma 3** Suppose that Assumptions 1, 2 and 3 hold. When $\bar{p}$ converges to $v_L$, the signaling thresholds $\bar{x}$ and $\underline{x}$ diverge to $-\infty$, and when $\bar{p}$ converges to $v_H$, the purchase threshold signals $\bar{x}$ and $\underline{x}$ diverge to $\infty$.

**Proof of Lemma 3** When $\bar{p} < (1-g)v_H + gv_L$, $b_L > b_H$ and by the first-order conditions in (6)

$$f(\bar{x} - b_L - v_L) > f(\bar{x} - b_H - v_H) \Leftrightarrow |\bar{x} - b_L - v_L| < |\bar{x} - b_H - v_H|,$$

since $f(\cdot)$ is unimodal. Then by Lemma 1 $\bar{x} < b_H + v_H$.

On the other hand, when $\bar{p} > (v_H + v_L)/2$, $b_L < b_H$ and by the first-order conditions in (6)

$$f(\bar{x} - b_L - v_L) < f(\bar{x} - b_H - v_H) \Leftrightarrow |\bar{x} - b_L - v_L| > |\bar{x} - b_H - v_H|.$$

Then, by Lemma 1 $\bar{x} > b_L + v_L$.

The consumer indifference condition (11) can rewritten as follows:

$$\frac{v_H - \bar{p}}{\bar{p} - v_L} = \frac{f(\bar{x} - b_L - v_L)}{f(\bar{x} - b_H - v_H)}.$$

When $\bar{p} \to v_L$, LHS of the equation above converges to $\infty$ and therefore $f(\bar{x} - b_H - v_H) \to 0$, which implies $\bar{x} \to \{\infty, \infty\}$. But since $\bar{x} \leq b_H + v_H < \infty$, $\bar{x} \to -\infty$.

When $\bar{p} \to v_H$, LHS of the equation above converges to 0 and therefore $f(\bar{x} - b_L - v_L) \to 0$, which implies $\bar{x} \to \{-\infty, \infty\}$. But since $\bar{x} \geq b_L + v_L > -\infty$, $\bar{x} \to \infty$.

**Lemma 4** Suppose that Assumptions 1, 2 and 3 hold. The purchase threshold signal $\bar{x}$ strictly increases with price $\bar{p}$. That is, $\partial \bar{x}/\partial \bar{p} > 0$.  

41
Proof of Lemma 4. To simplify the notation let \( f(\bar{x} - b_L - v_L) = f_L, \) \( f(\bar{x} - b_H - v_H) = f_H, \) \( f'(\bar{x} - b_L - v_L) = f'_L, \) \( f'(\bar{x} - b_H - v_H) = f'_H. \)

Implicit differentiation of the first-order conditions in (6) enables us to calculate \( \partial b_L / \partial \bar{p} \) and \( \partial b_H / \partial \bar{p} \) as follows.

\[
\frac{\partial b_L}{\partial \bar{p}} = \frac{f_L}{\tilde{p} f'_L + C''(b_L)} + \frac{\partial f'_L}{\partial \bar{p}} \quad \text{and,}
\]
\[
\frac{\partial b_H}{\partial \bar{p}} = \frac{f_H}{\tilde{p} f'_H + C''(b_H)} + \frac{\partial f'_H}{\partial \bar{p}}
\]

Implicit differentiation of the consumer indifference condition (11) yields \( \partial \bar{x} / \partial \bar{p} \) as follows:

\[
-f_H(1 - g) + (1 - g)(v_H - \bar{p}) f'_H \left( \frac{\partial \bar{x}}{\partial \bar{p}} - \frac{\partial b_H}{\partial \bar{p}} \right) = g f_L + g(\bar{p} - v_L) f'_L \left( \frac{\partial \bar{x}}{\partial \bar{p}} - \frac{\partial b_L}{\partial \bar{p}} \right)
\]

\[
\frac{\partial \bar{x}}{\partial \bar{p}} = \frac{g f_L + (1 - g) f_H}{(1 - g)(v_H - \bar{p}) f'_H - g(\bar{p} - v_L) f'_L} + \frac{g(\bar{p} - v_L) f'_L}{(1 - g)(v_H - \bar{p}) f'_H - g(\bar{p} - v_L) f'_L} \frac{\partial b_L}{\partial \bar{p}}
\]

Now, suppose that \( \partial \bar{x} / \partial \bar{p} = 0. \) Then combining the calculations above, \( \partial \bar{x} / \partial \bar{p} = 0 \) implies that

\[
\frac{g f_L + (1 - g) f_H}{(1 - g)(v_H - \bar{p}) f'_H - g(\bar{p} - v_L) f'_L} + \frac{g(\bar{p} - v_L) f'_L}{(1 - g)(v_H - \bar{p}) f'_H - g(\bar{p} - v_L) f'_L} \frac{f_H}{f'_L} = \frac{g(\bar{p} - v_L)}{(1 - g)(v_H - \bar{p}) f'_H - g(\bar{p} - v_L) f'_L} f_L + \frac{C''(b_L)}{f'_L} = 0,
\]

which after some reorganization implies that

42
\[
0 = \frac{gf_L}{(1-g)(v_H-\bar{p})f_H' - g(\bar{p} - v_L)f_H'} \left( 1 - \frac{\bar{p} - v_L}{\bar{p}} \frac{\bar{p}f_L'}{\bar{p}f_H' + C''(b_L)} \right) + \frac{(1-g)(v_H-\bar{p})f_H' - g(\bar{p} - v_L)f_H'}{(1-g)(v_H-\bar{p})f_H' - g(\bar{p} - v_L)f_H'} \left( 1 + \frac{v_H - \bar{p}}{\bar{p}} \frac{\bar{p}f_H'}{\bar{p}f_H' + C''(b_H)} \right) \neq 0.
\]

But this is a contradiction. To see why, note that \(C''(b_L) > 0\) implies
\[
0 < \bar{p} - v_L \bar{p} \frac{\bar{p}f_L'}{\bar{p}f_H' + C''(b_L)} < 1,
\]
and
\[
\frac{v_H - \bar{p}}{\bar{p}} \frac{\bar{p}f_H'}{\bar{p}f_H' + C''(b_H)} < -1 \iff C''(b_H) < v_H f_H'.
\]

leading to a contradiction by Assumption 3.

We have established that \(\partial \bar{x}/\partial \bar{p}\) cannot be 0. But this implies by Lemma 3 that \(\partial \bar{x}/\partial \bar{p}\) is globally positive.

**Proof of Theorem 2** We will start with the effective manipulation by the \(L\)-type. \(L\)-type does effective manipulation if and only if \(b_L > \bar{x} - \bar{x}\), which is equivalent to \(\bar{x} > \bar{x} - b_L\). By Corollary 1 this is equivalent to \(E(v|\bar{x} - b_L) < \bar{p}\) if manipulation is restricted to be 0. That is, when there is no manipulation, a consumer who receives a signal that equals to \(\bar{x} - b_L\), should expect that the quality should be lower than the price \(\bar{p}\). The posterior expectation of such a consumer is calculated as
\[
\frac{(1-g)v_H f(\bar{x} - b_L - v_H) + gv_L f(\bar{x} - b_L - v_L)}{(1-g)f(\bar{x} - b_L - v_H) + gf(\bar{x} - b_L - v_L)} < \bar{p}.
\]
\[
(1-g)(v_H - \bar{p})f(\bar{x} - b_L - v_H) - g(\bar{p} - v_L)f(\bar{x} - b_L - v_L) \iff f(\bar{x} - b_L - v_H) < f(\bar{x} - b_H - v_H) \iff |\bar{x} - b_L - v_H| > |\bar{x} - b_H - v_H|.
\]

In sum,
\[
b_L > \bar{x} - \bar{x} \iff |\bar{x} - b_L - v_H| > |\bar{x} - b_H - v_H|.
\]
We will show below that this equivalence condition for $L$-type’s effective manipulation holds when $\bar{p} < (1 - g)v_H + gv_L$ or when $\bar{p} > (1 - g)v_H + gv_L$ and $\bar{x} > (b_L + b_H)/2 + v_H$.

Now, when $\bar{p} < (1 - g)v_H + gv_L$, $b_L > b_H$ by Proposition 2. Then by first-order conditions (6)

$$f(\bar{x} - b_L - v_L) > f(\bar{x} - b_H - v_H) \Leftrightarrow |\bar{x} - b_L - v_L| < |\bar{x} - b_H - v_H|,$$

since $f(\cdot)$ is unimodal. Then by Lemma 1 the condition above is equivalent to

$$\bar{x} < b_H + v_H,$$

$$\bar{x} - b_H - v_H < \bar{x} - b_L - v_L < b_H + v_H - \bar{x} \Rightarrow$$

$$\bar{x} < \frac{b_L + b_H}{2} + \frac{v_L + v_H}{2} < \frac{b_L + b_H}{2} + v_H$$

On the other hand, when $b_L > b_H$, condition 28 is equivalent to

$$\bar{x} < b_L + v_H$$

$$\bar{x} - b_L - v_H < \bar{x} - b_H - v_H < b_L + v_H - \bar{x} \Leftrightarrow$$

$$\bar{x} < \frac{b_L + b_H}{2} + v_H,$$

which holds when $\bar{p} < (1 - g)v_H + gv_L$.

When $\bar{p} > (1 - g)v_H + gv_L$, $b_H > b_L$ by Proposition 2. Therefore, because of Lemma 1 condition 28 becomes equivalent to

$$\bar{x} > b_L + v_H$$

$$b_L + v_H - \bar{x} < \bar{x} - b_H - v_H < \bar{x} - b_L - v_H \Leftrightarrow$$

$$\bar{x} > \frac{b_L + b_H}{2} + v_H.$$

This completes the argument for effective manipulation by the $L$-type.

By a similar argument as above, $H$-type does effective manipulation if and only if
\[ b_H > \bar{x} - x \iff \bar{x} > x - b_H \]

\[ (v_H - \bar{p})(1 - g)f(\bar{x} - b_H - v_H) - (\bar{p} - v_L)g f(\bar{x} - b_H - v_L) < 0 \]

\[ = (v_H - \bar{p})(1 - g)f(\bar{x} - b_H - v_H) - (\bar{p} - v_L)g f(\bar{x} - b_L - v_L) \iff \]

\[ f(\bar{x} - b_L - v_L) < f(\bar{x} - b_H - v_L) \iff |\bar{x} - b_L - v_L| > |\bar{x} - b_H - v_L|. \]

In sum,

\[ b_H > \bar{x} - x \iff |\bar{x} - b_L - v_L| > |\bar{x} - b_H - v_L|. \]  \hspace{1cm} (29)

When \( \bar{p} < (1 - g)v_H + gv_L \), \( b_H < b_L \), therefore condition (29) is equivalent to

\[ \bar{x} < b_L + v_L, \]

\[ \bar{x} - b_L - v_L < \bar{x} - b_H - v_L < b_L + v_L - \bar{x} \iff \]

\[ \bar{x} < v_L + \frac{b_L + b_H}{2}. \]

When \( \bar{p} > (1 - g)v_H + gv_L \), \( b_H > b_L \), therefore condition (29) is equivalent to

\[ \bar{x} > b_L + v_L \]

\[ b_L + v_L - \bar{x} < \bar{x} - b_H - v_L < \bar{x} - b_L - v_L \iff \]

\[ \bar{x} > v_L + \frac{b_L + b_H}{2}, \]

which holds as long as \( b_H > b_L \).

Now, we will show that there exists a unique price \( p_L < (1 - g)v_H + gv_L \) such that \( \bar{x} = v_L + \frac{b_L + b_H}{2} \) and a unique price \( p_H > (1 - g)v_H + gv_L \) such that \( \bar{x} = v_H + \frac{b_L + b_H}{2} \).

First, note that existence of these price levels is a result of Lemmas 3 and 4. To show that \( p_L < (1 - g)v_H + gv_L \), note that
\[
\bar{x} = v_L + \frac{b_L + b_H}{2} \Rightarrow \\
\bar{x} - b_L - v_L = \frac{b_H - b_L}{2} \quad \& \quad \bar{x} - b_H - v_H = -(v_H - v_L) + \frac{b_L - b_H}{2} \\
\Rightarrow |\bar{x} - b_H - v_H| > |\bar{x} - b_L - v_L| \Rightarrow b_H < b_L,
\]
since the noise distribution is unimodal. Note moreover that \(b_H < b_L\) implies in this case that \(\bar{x} < b_L + v_L\). These imply by Proposition \(^2\) that any price level \(p_L < (1-g)v_H + gv_L\).

Now suppose that there are two price levels \(p_{L1} < p_{L2}\). Let \(\bar{x}_l, b_{Ll}, b_{Hl}\) be equilibrium variables associated with \(p_{Ll}\) for \(l \in \{1, 2\}\). Since \(\partial \bar{x}/\partial \bar{p} > 0\), \(\bar{x}_2 > \bar{x}_1\), and therefore

\[
b_{L1} + b_{H1} < b_{L2} + b_{H2}.
\]

Since \(\bar{x} < b_L + v_L\) and \(f(\cdot)\) is unimodal,

\[
p_{L2} f(\bar{x}_2 - b_{L1} - v_L) > p_{L1} f(\bar{x}_1 - b_{L1} - v_L),
\]

which implies \(b_{L2} > b_{L1}\) by Assumption \(^3\). Moreover, \(b_{L2} > b_{L1}\) implies that

\[
0 > \frac{b_{H2} - b_{L2}}{2} > \frac{b_{H1} - b_{L1}}{2} \Rightarrow b_{H2} > b_{H1} \Rightarrow \\
0 > -(v_H - v_L) + \frac{b_{L2} - b_{H2}}{2} > -(v_H - v_L) + \frac{b_{L1} - b_{H1}}{2},
\]

which is a contradiction. This shows that there exists a unique such \(p_L\). The proof that there exists a unique \(p_H > (1-g)v_H + gv_L\) is symmetric to the argument above.

**Proof of Proposition \(^3\)** For notational simplicity let \(f_j := f(\bar{x} - v_j - b_j)\) and \(F_j := F(\bar{x} - v_j - b_j)\) for \(j \in \{H, L\}\). With a noisy signal, the consumer purchases the product if and only if \(\bar{p} \leq gv_L + (1-g)v_H\), i.e., whenever expected quality under prior beliefs exceeds the price. Therefore, ex-ante (expected) consumer utility (surplus) without any signal is equal to
\[ E_{NS}(CS) = \begin{cases} g v_L + (1 - g) v_H - \bar{p}, & \text{if } \bar{p} \leq g v_L + (1 - g) v_H \\ 0, & \text{if otherwise.} \end{cases} \]

On the other hand, when the consumer acts upon the biased signal, expected consumer surplus will be

\[ E_S(CS) = g(1 - F_L)(v_L - \bar{p}) + (1 - g)(1 - F_H)(v_H - \bar{p}) \]

We need to show that \( E_S(CS) \geq E_{NS}(CS) \) always holds. When consumer decides based on signals, the purchase threshold signal is given by equation [9] which in turn implies that

\[ \frac{(1 - g)(v_H - \bar{p})}{g(\bar{p} - v_L)} = \frac{f_L}{f_H}. \] (30)

Consider the first case, i.e., \( \bar{p} \leq g v_L + (1 - g) v_H \). Then,

\[ E_S(CS) \geq E_{NS}(CS) \Leftrightarrow -g(v_L - \bar{p})F_L - (1 - g)(v_H - \bar{p})F_H \geq 0 \Leftrightarrow \]

\[ \frac{F_L}{F_H} \geq \frac{(1 - g)(v_H - \bar{p})}{g(\bar{p} - v_L)} = \frac{f_L}{f_H} \Leftrightarrow \]

\[ \frac{f_L}{F_L} \leq \frac{f_H}{F_H}, \] (32)

where the equality in [31] follows from equation [30]. By Lemma [1] \( b_L + v_L < b_H + v_H \). Therefore, \( \bar{x} - v_L - b_L > \bar{x} - v_H - b_H \). Also, since \( f \) is log-concave by Assumption [1], \( F \) must also be log-concave. Combining these two observations we obtain the inequality in [32] as desired.
Now consider the remaining case that $\bar{p} > gv_L + (1 - g)v_H$. Then,

$$E_S(CS) \geq E_{NS}(CS) \Leftrightarrow g(v_L - \bar{p})(1 - F_L) + (1 - g)(v_H - \bar{p})(1 - F_H) \geq 0 \Leftrightarrow$$

$$\frac{1 - F_L}{1 - F_H} \leq \frac{(1 - g)(v_H - \bar{p})}{g(\bar{p} - v_L)} = \frac{f_L}{f_H} \Leftrightarrow$$

$$\frac{f_L}{1 - F_L} \geq \frac{f_H}{1 - F_H}, \quad (33)$$

where, as before, the equality in $33$ follows from equation $30$. Since $f$ is log-concave, $1 - F$ is also log-concave. This in turn implies that $f(x)/(1 - F(x))$ is increasing in $x$. Since $\bar{x} - v_L - b_L > \bar{x} - v_H - b_H$ always holds, we obtain the inequality in $34$ as desired.

**Proof of Proposition 5** Let $((p_L, b_L), (p_H, b_H))$ be any pure-strategy separating PBE. By definition, $p_L \neq p_H$ and suppose for a moment both prices are strictly less than $v_H$. Then all consumers would buy from the high-type firm at price $p_H$, which gives incentive to the $L$-type to imitate $p_H > 0$. On the other hand, no PBE with $p_H \in \{0, v_H\}$ would be strict since either the firm or the consumers would be indifferent with other strategies. ■

**Proof of Proposition 4** When $v_L = 0$, the description of manipulation strategies and their existence is the same as in the exogenous price case, as described by Theorem 1. The prices can be supported by the off-equilibrium beliefs that any deviation to another price means that the firm is of $L$-type.

When $v_L > 0$, the firm has the outside option of choosing $v_L$ and making positive profits. Therefore, for extremely high prices where the expected sales are very close to zero, the firm may choose to deviate to $v_L$ because $\bar{x} \rightarrow \infty$. ■

**Proof of Theorem 3** The proof consists of two steps. In the first step, we will argue that equations (17) to (20) describe a partially separating PBE. Then in the second step, we will prove that equations (17) to (20) have a solution for high enough prices.

For the $L$-type to mix between two prices, it should be indifferent between the profit levels at the two prices. When $L$-type charges $v_L$, all consumers buy the good since their expected consumer surplus is at least 0. Therefore, the firm will sell the product to the whole market, providing a profit of $v_L$ to the firm. If the $L$-type chooses $\bar{p}$ with the corresponding bias level $b_L$, the expected profit would be given as in the RHS of equation (17), which ensures that $L$-type is indifferent between choosing $v_L$ and $\bar{p}$.

When consumers observe $\bar{p}$, they remain uncertain about the type of the firm because both types could have chosen this price level. Therefore, all consumers consult to the pri-
vate signals they receive. Suppose for a moment that consumers use a purchase threshold signal $\bar{x}$ when they observe the price $\bar{p}$. Expecting that consumers use this threshold, the optimal manipulation levels are given by the first-order conditions (18) and (19).

Given the bias levels $b_L$ and $b_H$, consumers infer that the price $\bar{p}$ could be chosen by the $L$-type with probability $\bar{\alpha}/(1+\bar{\alpha})$ and $H$-type with probability $1/(1+\bar{\alpha})$. Then, given the private signal $x_i$, the expected quality is

$$\frac{v_H f(\bar{x} - v_H - b_H) + v_L \bar{\alpha} f(\bar{x} - v_L - b_L)}{f(\bar{x} - v_H - b_H) + \bar{\alpha} f(\bar{x} - v_L - b_L)}.$$ 

Consumers are indifferent when they receive the threshold signal $\bar{x}$. After some rearrangement, the indifference condition can be written as equation (20).

To establish the existence, first note that for every value of $\bar{x}$ and $\bar{p}$, the bias levels $b_L$ and $b_H$ are bounded since the cost function is unbounded but the p.d.f. function $f(\cdot)$ is bounded. Therefore, by equation (17), $\bar{x}$ should be finite as well. Now, as $\bar{x} \to \infty$, RHS of equation (17) converges to $-C(b_L) < 0$. As $\bar{x} \to -\infty$, the RHS of equation (17) converges to $\bar{p} - C(b_L)$. When we substitute $b_L$ from equation (18), RHS of equation (17) becomes

$$\bar{p}(1 - F(\bar{x} - b_L - v_L)) - C((C''^{-1}(\bar{p} f(\bar{x} - b_L - v_L)))),$$

which converges to $\bar{p} > v_L$.

Now, we need to check two inequalities, $\bar{\alpha} < 1$ and the profit of the $H$-type being greater than $v_L$. The latter follows from the quality advantage of the $H$-type. Even if the $H$-type chooses the same manipulation level as the $L$-type, its profit will be

$$\bar{p}(1 - F(\bar{x} - b_L - v_H)) - C(b_L) > \bar{p}(1 - F(\bar{x} - b_L - v_L)) - C(b_L) = v_L.$$

For $\bar{\alpha}$, consider equations (20) and (17). As $\bar{p}$ converges to $v_H$, (17) implies that $\bar{x}$ converges to a finite value. Given this, equation (20) implies that $\bar{\alpha}$ converges to 0. Therefore, there exists $\bar{p} < v_H$ such that for all $\bar{p} > \bar{p}$, $\bar{\alpha} < 1$. This establishes the existence. \[\Box\]

**Proof of Proposition 6** Implicitly differentiating the incentive compatibility condition for the $L$-type, equation (17) implies

$$1 - F(\bar{x} - b_L - v_L) - \bar{p} f(\bar{x} - b_L - v_L) \left( \frac{\partial \bar{x}}{\partial \bar{p}} - \frac{\partial b_L}{\partial \bar{p}} \right) - C'(b_L) \frac{\partial b_L}{\partial \bar{p}} = 0.$$
Substituting the first-order condition for the L-type, equation (18) yields

\[
\frac{\partial \bar{\bar{x}}}{\partial \bar{\bar{p}}} = \frac{1 - F(\bar{x} - b_L - v_L)}{\bar{\bar{p}}f(\bar{x} - b_L - v_L)} > 0.
\] (35)

The behavior of \(b_L\) and \(\pi_H\) both depend on the monotonicity of the hazard rate of the noise distribution. To show the monotonicity, note that the c.d.f. \(F\) being log-concave implies that the ratio \(f/F\) is a decreasing function. Moreover, the derivative of \(f/F\) is negative, i.e., \(f'F - f^2 < 0 \iff f^2 > f'F\). Since the noise distribution is symmetric, this also implies that \(f^2 > (-f')(1 - F)\). Moreover, the ratio \((1 - F)/f\) is also decreasing.

Implicitly differentiating the first-order condition for the L-type, equation (18) yields

\[
\frac{\partial b_L}{\partial \bar{\bar{p}}} = \frac{f^2(\bar{x} - b_L - v_L) + f'(\bar{x} - b_L - v_L)(1 - F(\bar{x} - b_L - v_L))}{f(\bar{x} - b_L - v_L)(\bar{\bar{p}}f'(\bar{x} - b_L - v_L) + C''(b_L))} > 0,
\]

since \(F\) is log-concave, which implies that the nominator is positive, and the second-order condition for the manipulation decision of the L-type implies that the denominator is positive.

Finally, the implicit derivative of the profit function after substituting the first-order condition for the H-type, equation (19) becomes

\[
\frac{\partial \pi_H}{\partial \bar{\bar{p}}} = 1 - F(\bar{x} - b_H - v_H) - \bar{\bar{p}}f(\bar{x} - b_H - v_H)\frac{\partial \bar{x}}{\partial \bar{\bar{p}}} > 0 \iff \frac{1 - F(\bar{x} - b_L - v_L)}{f(\bar{x} - b_L - v_L)},
\]

which always holds by Assumption 4. \(\square\)

**Proof of Proposition 7** Equation (20) implies that when \(\bar{\bar{p}} = \hat{\bar{p}}\alpha\),

\[
f(\bar{x} - b_L - v_L) = f(\bar{x} - b_H - v_H),
\]

which implies that \(b_L = b_H\) by the first-order conditions (18) and (19). Moreover, by a comparison, \(b_L > b_H\) if and only if \(\bar{\bar{p}} < \hat{\bar{p}}\alpha\).

To show that the L-type always effectively manipulates, compare the indifference conditions (17) and (21). Since

\[
\bar{\bar{p}}(1 - F(\bar{x} - v_L)) = v_L = \hat{\bar{p}}(1 - F(\bar{x} - b_L - v_L)) - C(b_L),
\]

50
the sales of the $L$-type when there is manipulation is always greater than its sales when there is no manipulation, i.e., $b_L > \bar{x} - \bar{x}$. Then, since $b_H \geq b_L$ when $\bar{p} \geq \hat{p}_\alpha$, we have

$$b_H \geq b_L > \bar{x} - \bar{x},$$

which completes the proof. ■

**Proof of Proposition 9** Consider any strictly positive pair of bias levels $(b_L, b_H) > 0$. Recall that consumers’ purchasing decisions are determined by the following equation:

$$\frac{v_H f(\bar{x} - v_H - b_H)}{f(\bar{x} - v_H - b_H) + f(\bar{x} - b_L)} = \bar{p}.$$

The left-hand side does not change if we subtract $\min\{b_L, b_H\}$ from $b_L, b_H$ and $\bar{x}$ since

$$\frac{v_H f(\bar{x} - \min\{b_L, b_H\} - v_H - (b_H - \min\{b_L, b_H\}))}{\sum_j f(\bar{x} - \min\{b_L, b_H\} - v_j - (b_j - \min\{b_L, b_H\}))} = \frac{v_H f(\bar{x} - v_H - b_H)}{f(\bar{x} - v_H - b_H) + f(\bar{x} - b_L)} = \bar{p}.$$

That is, if the firm reduces both bias levels by $\min\{b_L, b_H\}$, the consumers’ unique response will be to reduce $\bar{x}$ by $\min\{b_L, b_H\}$ since they expect the firm to exert manipulative effort at a uniformly lower level.

Since manipulation is costly, the firm can always achieve a higher profit by reducing the minimum manipulative effort level to 0. ■
B Equilibrium Selection

Proposition 4 and Theorem 3 show the existence of two sets of PBE. One of the implications of these two results is that the pricing outcome of the game is resolved through the coordination between the expectations of the firm and consumers. In particular, the off-equilibrium path beliefs of the consumers for each PBE prevent the firm to deviate to a price other than the one prescribed by that PBE. The most conservative off-equilibrium beliefs are such that whenever consumers observe a price other than what is prescribed by a particular PBE, they believe that it is only the low-quality type that could make such a deviation. Therefore, they would not purchase the good for any price $\bar{p} > v_L$. These extreme beliefs support any PBE described by Proposition 4 and Theorem 3. However, some of the PBE can be supported by less extreme off-equilibrium beliefs as well. In this section, we investigate which PBE survive when we put some discipline on the off-equilibrium beliefs.

Firstly, observe that the assumption that production costs are the same for both types of the firm makes it impossible for the high-quality type to credibly communicate to the consumers that it is not a low-quality type. This is because if the high-quality type expects to get a higher profit by deviating to an off-equilibrium price, it expects that consumers would have favorable enough expectations. However, the low-quality type can simply imitate the same behavior as well. Therefore, from the consumers’ perspective, if an off-equilibrium price offers a higher profit to one type, it should offer the same for all types. This immediately implies that the Intuitive Criterion cannot eliminate any equilibria, even if it is a weak one. Indeed a straightforward check of the definition of the Intuitive Criterion proves the following result.

**Proposition 10** All PBE satisfy the requirements of the Intuitive Criterion.

The reason that the Intuitive Criterion does not help us refining the set of PBE is that it is impossible to devise credible deviations that benefit only one type of the firm. Therefore, the only way we can eliminate some PBE is by arguing that both types can benefit from deviating to an off-equilibrium price. For such an argument we employ the concept of “Undefeated Equilibrium”. We first make some observations about off-equilibrium beliefs that consumers can hold.

To simplify the analysis for equilibrium selection among the pooling PBE, we assume in what follows that the ex-post profit functions are strictly concave in price, and, therefore, there is an “optimal” price for each firm type that maximizes the ex-post profit among
the pooling equilibrium profit levels. This property enables us to rank the pooling PBE for each type.

**Assumption 5** Let $\Phi \subseteq (v_L, v_H)$ be the set of prices that could be supported by pooling PBE. The equilibrium profit functions are strictly concave in price and for each type $j \in \{L, H\}$ there exists a unique price $\bar{p}_j \in \Phi$ such that for any other price $p \in \Phi$

$$\pi_j(b^*_L(p), b^*_H(p), \bar{x}(p)) \geq \pi_j(b^*_L(\bar{p}_j), b^*_H(\bar{p}_j), \bar{x}(\bar{p}_j)).$$

The following Lemma 5 states that the best price $p_H$ for the high-type firm is higher than the best price $p_L$ for the low-type. The proof is a direct result of Proposition 1.

**Lemma 5** Suppose that Assumption 5 holds. The price $p_H$ that gives the highest profit to the $H$-type firm among all pooling PBE is strictly greater than the corresponding price $p_L$ for the $L$-type.

One implication of Lemma 5 is that there is an interval $[p_L, p_H]$ such that for any $p \in [p_L, p_H]$ and $p' \notin [p_L, p_H]$,

$$\pi_j(b^*_L(p), b^*_H(p), \bar{x}(p)) > \pi_j(b^*_L(p', p_H), \bar{x}(p'))$$

for each type $j \in \{L, H\}$.

Second implication of Lemma 5 is that there is no unique price that makes a pooling PBE the best one for both types. Therefore, for any pooling PBE, each type should face a less favorable out-of-equilibrium belief by consumers since, otherwise, at least one of the types could deviate to the price supported by the best PBE. Based on this argument, Proposition 11 below provides two “bounds” for off-equilibrium beliefs that support pooling PBE.

**Proposition 11** Let $\bar{p} \in (v_L, v_H)$ be any pooling price. Then the following belief about the informativeness of prices is sufficient to support a pooling PBE:

$$P(v = v_H|p) = \begin{cases} P(v = v_H) & \text{if } p = \bar{p} \\ 0 & \text{o.w.} \end{cases}$$

(36)

On the other hand, the following belief is necessary for any pooling PBE that supports $\bar{p}$:

There exists a range of prices $R(\bar{p}) \subset (v_L, v_H)$ such that for any price $p \in R(\bar{p})$
\[ P(v = v_H | p) = \begin{cases} P(v = v_H) & \text{if } p = \bar{p} \\ < P(v = v_H) & \text{o.w.} \end{cases} \quad (37) \]

**Proof** The sufficiency of expectations specified in equation (36) is straightforward. When the firm of either type deviates to any other price \( p \in (v_L, v_H) \) from the price \( \bar{p} \) prescribed by the equilibrium, it faces zero demand because consumers believe with certainty that the deviating type is the low-type firm.

The necessity of expectations specified in equation (37) comes from Lemma 5. For any price \( \bar{p} \in (v_L, v_H) \), if the deviation induces a demand associated with a belief at least as favorable as the prior beliefs, at least one of the types would deviate to \( p_L \) or \( p_H \) to achieve a higher profit. Since \( p_H \neq p_L \), there is no price that prior beliefs as “pre-signal” off-equilibrium beliefs could support a pooling PBE at that price. ■

One of the implications of Proposition 11 is that whenever consumers observe an off-equilibrium price, they should reduce the likelihood of a high-quality firm from the prior belief. If this is not the case for a particular price, that price cannot be supported by a pooling PBE.

Next, we analyze how much “Undefeated Equilibrium” might help in refining the set of equilibria. The original definition proposed by Mailath, Okuno-Fujiwara and Postlewaite (1993) was designed for a single sender and a single receiver. We extend the definition for our information manipulation model with many receivers, additional private information by receivers and mixed-pricing strategies as below.

**Definition 1** An equilibrium \( (\hat{b}_L, \hat{b}_H, \hat{\beta}_L, \hat{\beta}_H, \hat{s}(x,p), \hat{\mu}(x,p)) \) defeats \( (\bar{b}_L, \bar{b}_H, \bar{\beta}_L, \bar{\beta}_H, \bar{s}(x,p), \bar{\mu}(x,p)) \) if there exists a price \( p' \in [v_L, v_H] \) such that

1. for any type \( j \in \{L, H\} \) of firm, \( p' \notin \text{supp}(\beta_j) \) and \( K = \{j \in \{L, H\} | p' \in \text{supp}(\beta_j)\} \neq \emptyset \);

2. for any type \( j \in K \), the profit of \( j \)-type firm \( \pi(\bar{\beta}_j, \bar{b}_j, \bar{s}(\cdot, \cdot)) \geq \pi(\hat{\beta}_j, \hat{b}_j, \hat{s}(\cdot, \cdot)) \), and for at least one of the types \( j' \in K \), \( \pi(\bar{\beta}_{j'}, \bar{b}_{j'}, \bar{s}(\cdot, \cdot)) > \pi(\hat{\beta}_{j'}, \hat{b}_{j'}, \hat{s}(\cdot, \cdot)) \);

3. there exists a type \( j \in K \) such that

\[ \bar{\mu}(x, p')(j) \neq \frac{P(x|j)P(j)\phi(j)}{P(x|L)P(L)\phi(L) + P(x|H)P(H)\phi(H)}, \]

for any function \( \phi : \{L, H\} \to [0, 1] \) satisfying

54
\[ j' \in K \text{ and } \pi(\hat{\beta}_{j'}, \hat{b}_{j'}, \hat{s}(\cdot, \cdot)) > \pi(\hat{\beta}_{j}, \hat{b}_{j}, \hat{s}(\cdot, \cdot)) \Rightarrow \phi(j') = 1, \]
\[ j' \notin K \Rightarrow \phi(j') = 0. \]

An equilibrium is called undefeated if and only if it is not defeated by another equilibrium.

The relation of defeating constitutes a strict partial order among the PBE. It is possible to rule out some of the outcomes using this order. However, due to the multiplicity of off-equilibrium beliefs that may support the same outcome, it is not possible to select a single equilibrium outcome or even one type of outcome. We show that a strict subset of pure-strategy pooling PBE (described in Proposition 4) are undefeated and that partially separating PBE (described in Theorem 3) that have relatively conservative off-equilibrium beliefs can all be defeated by other partial-separating PBE. Moreover, pooling PBE and partially-separating PBE cannot be compared by the defeating relationship.

Proposition 12 below shows that the interval of prices \([\bar{p}_L, \bar{p}_H]\) specified by Assumption 5 are undefeated.

**Proposition 12** Suppose that Assumptions 1, 2, 3 and 5 hold. A pooling PBE is undefeated only if it supports a price \(\bar{p} \in [\bar{p}_L, \bar{p}_H]\). Moreover, pooling PBE that support a price in the interval \([\bar{p}_L, \bar{p}_H]\) are not defeated by other pooling PBE.

**Proof** We first show that any pooling PBE that supports a price that is outside of the interval \([\bar{p}_L, \bar{p}_H]\) is defeated by another pooling PBE. This argument has two parts and the second part is very similar to the first part, so we will present only the first part.

Let \(\bar{p} < \bar{p}_L\) and consider any two pooling PBE that support \(\bar{p}\) and \(\bar{p}_L\), respectively. The set of types \(K\) that might want to deviate from \(\bar{p}\) to \(\bar{p}_L\) consists of both firm types. That is, \(K = \{L, H\}\), and for each type, the equilibrium payoff at \(\bar{p}_L\) is strictly higher than \(\bar{p}\). To see that condition 3 in Definition 1 also holds, note that for \(\bar{p}\) to be supported by a pooling PBE, consumers’ off-equilibrium pre-signal belief \(P_{\bar{p}}(v_H|\bar{p}_L) < P(v_H)\). Therefore, consumers’ off equilibrium belief for any deviation from \(\bar{p}\) to \(\bar{p}_L\) cannot be the same as the Bayesian updating prescribed in condition 3 in Definition 1.

A similar argument shows that any pooling PBE that supports \(\bar{p} > \bar{p}_H\) is defeated by a pooling PBE that supports \(\bar{p}_H\).

It is clear that two PBE that differ only with respect to their off-equilibrium beliefs cannot defeat each other. Moreover, no two pooling PBE that support two different prices...
in $[\bar{p}_L, \bar{p}_H]$ can defeat each other since by Assumption 5 one of them offers a higher profit to one of the types. □

Proposition 13 shows that we can rank partially separating PBE based on the price they support if they have rather conservative off-equilibrium beliefs.

**Proposition 13** Consider any partially separating PBE that supports a price $\bar{p} \in (v_L, v_H)$. If the off-equilibrium probability $P(v_H|p') < 0.5$ for any price $p' \neq \bar{p}$, there is another partially separating PBE that supports $\hat{p} > \bar{p}$ and defeats the PBE that supports $\bar{p}$.

**Proof** Consider any two partially separating PBE indexed by the prices $p_1 < p_2$. Suppose that the first PBE assigns $P_1(v_H|p') < 0.5$ to any off-equilibrium price $p' \neq p_1$. Our claim is that the partially separating PBE that supports $p_2$ defeats the first PBE.

First, note that $p_2$ is not in the set of prices that are prescribed by the first PBE. Therefore, $K = \{L, H\}$. The profit of the $L$-type remains the same across the two PBE as it is fixed at $v_L$. On the other hand, the profit of the $H$-type is strictly higher in the second PBE by Proposition 6. Examining condition 3 in Definition 1 reveals that consumers’ possible off-equilibrium beliefs after observing a deviation from $p_1$ to $p_2$ assign a probability to the $H$-type that ranges from 0.5 to 1. This probability is higher than the prescribed off-equilibrium belief for the first PBE by hypothesis. □
C Firm’s Manipulation Plan under Commitment

The firm has three choices for manipulative effort plan: \((b_L, 0)\), \((0, b_H)\) or \((0, 0)\). The choice of the firm depends on the sign of the net marginal revenue of manipulative effort. The net marginal revenue for \(b_H\) given that \(b_L = 0\) is

\[
(1 - g)\bar{p}f(\bar{x} - b_H - v_H) \left( \frac{\partial \bar{x}}{\partial b_H} - 1 \right) + g\bar{p}f(\bar{x} - v_L) \frac{\partial \bar{x}}{\partial b_H} - C'(b_H),
\]

(38)

where \(\frac{\partial \bar{x}}{\partial b_H}\) can be implicitly defined using equation (24) as follows:

\[
\frac{\partial \bar{x}}{\partial b_H} = \frac{(1 - g)(v_H - \bar{p})f'(\bar{x} - v_H - b_H)}{(1 - g)(v_H - \bar{p})f'(\bar{x} - v_H - b_H) - g\bar{p}f'(\bar{x})}. \tag{39}
\]

The net marginal revenue for \(b_L\) given that \(b_H = 0\) is

\[
(1 - g)\bar{p}f(\bar{x} - v_H) \frac{\partial \bar{x}}{\partial b_L} + g\bar{p}f(\bar{x} - b_L) \left( \frac{\partial \bar{x}}{\partial b_L} - 1 \right) - C'(b_L),
\]

(40)

where \(\frac{\partial \bar{x}}{\partial b_L}\) can be implicitly defined using equation (24) as follows:

\[
\frac{\partial \bar{x}}{\partial b_L} = \frac{g\bar{p}f'(\bar{x} - b_L - v_L)}{g\bar{p}f'(\bar{x} - b_L - v_L) - (1 - g)(v_H - \bar{p})f'(\bar{x} - v_H)}.
\tag{41}
\]

The firm first calculates the optimal bias level \(b_H\) and the corresponding signal threshold \(\bar{x}\) using the first-order conditions (38), (39) and (24), then calculates optimal \(b_L\) and the corresponding \(\bar{x}\) using the first-order conditions (40), (41) and (24), and finally calculates the profit levels at \((0, b_H, \bar{x}(0, b_H))\), \((b_L, 0, \bar{x}(b_L, 0))\) and \((0, 0, \bar{x})\). The manipulation plan of the firm is the triple among the three alternatives that maximizes its profit.
References


JANSSSEN, M. C., and ROY, S. Regulating false disclosure, May 2017. mimeo.


