MANIPULATION THROUGH BIASED PRODUCT REVIEWS*

KEMAL KIVANÇ AKÖZ†
CEMAL EREN ARBATLI‡
LEVENT CELIK §

January 15, 2020

Abstract

We study a signal-jamming model of product review manipulation in which rational consumers consult product reviews and price to better estimate a product's quality, and a firm, whose quality is either high or low, chooses its price and how much bias to insert into product reviews. We show that both firm types always exert positive effort to manipulate product reviews, and, depending on the equilibrium price level, one or both of them can increase its sales. When the high-type firm exerts more effort than the low-type, review manipulation benefits consumers by raising [lowering] their demand for the high-quality [low-quality] product.

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

*Acknowledgements: We would like to thank Jacopo Bizzotto, Olivier Bochet, Anett Erdmann, Sergei Izmalkov, Maarten Janssen, Matthew Nagler, Regis Renault, Sandro Shelegia, Alex Smolin, Chris Wilson and John Wooders as well as the Editor and an anonymous referee for many helpful comments. We also thank the participants of 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, 2017 Lisbon Meetings in Game Theory and Applications, 2018 RES Conference, and 2018 ASSET Conference for valuable discussions. Cemal Eren Arbatlı and Levent Celik acknowledge financial support from the Russian Science Foundation (grant project no. 18-48-05007). All errors are our own.

†Authors' affiliations: National Research University Higher School of Economics, Myasnitskaya 20, 101000, Moscow, Russia. *Email: kakoz@hse.ru*

‡National Research University Higher School of Economics, Myasnitskaya 20, 101000, Moscow, Russia. *Email: darbatli@hse.ru*

§City, University of London, Northampton Square, London EC1V 0HB, United Kingdom; National Research University Higher School of Economics, Myasnitskaya 20, 101000, Moscow, Russia; and CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, Prague, Czech Republic. *Email: Levent.Celik@city.ac.uk*

I. INTRODUCTION

Information provision in the Internet era has become increasingly decentralized due to expanding scale and scope of crowdsourcing and word-of-mouth communication. More and more consumers rely on product reviews on Amazon, hotel and restaurant reviews on TripAdvisor, or movie reviews on IMDb to make more informed choices.[1] As participation in online review platforms increases, firms become more tempted to boost their online ratings. 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.[2] Posting or funding fake reviews, incentivizing consumers to recommend a product,[3] or selectively funding and disseminating research results that provide favorable information about a product are some of these strategies.[4] Bing Liu, a data-mining expert at the University of Illinois, Chicago, estimated that already back in 2012 about a third of all online consumer reviews were fake.[5] Today, the fake review problem has reached such an extent that several auditing sites like ReviewMeta and FakeSpot have emerged so as to help consumers make better judgments about the products advertised on online retail platforms by filtering out inauthentic reviews. Not surprisingly, more and more consumers take online reviews with a grain of salt. Although many of them might rationally anticipate a positive bias in product reviews and recommendations, it is hard for most to assess the extent of this bias.[6]

In this paper, we formalize these ideas through a signal-jamming model of biased product reviews and investigate when and how review 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 the quality of its product. Consumers are uncertain about quality, but draw on two sources of information

[1] According to a report in 2010, Amazon was the largest single source of consumer reviews on the internet with 10s of millions of reviews. TripAdvisor, the largest travel site in the world, now has more than 400 million average monthly visitors and over 700 million reviews and opinions of travel-related businesses. IMDb, likewise, has 83 million registered users and over 5.3 million titles in its database.

[2] To demonstrate how fake online review business operates, NBC News ran a little experiment on Facebook. The results reveal that the supply of fake reviews responds very fast and at a large scale. Source: https://www.nbcnews.com/business/consumer/can-you-trust-online-reviews-here-s-how-find-fakes-n976756

[3] See, among others, Hu et al. [2011], Anderson and Simester [2014], and Mayzlin, Dover and Chevalier [2014].

[4] See, for example, Finucane and Boult [2004], and Sismondo [2008] for the case of medical research.

[5] Source: http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html.

[6] See Mayzlin et al. [2014] for a discussion on the difficulty of distinguishing biased online reviews from unbiased ones.

before making a purchase decision. First, by visiting a product review platform, they receive an individual-specific noisy signal about product quality.[7] Second, they observe the price set by the firm. The firm can shift the mean of quality signals on the product review platform by exerting costly effort. Consumers cannot observe the firm's efforts, but are aware of the fact that signals may be biased upwards, and rationally anticipate how much bias each type will insert into product reviews in equilibrium. However, since they cannot observe the firm type, and since the signals they receive contain random noise, they cannot pin down the underlying product quality based on these signals. We say that review manipulation is 'effective' for a given firm type and at a given price level if it leads to higher sales than what we would observe without manipulation.

We first characterize Perfect Bayesian equilibria (PBE) where both quality types pool on the same price, and hence the price is uninformative about product quality. These PBE can be supported for a range of pooling prices and differ from each other with respect to each type's manipulative effort and how effective manipulation is for each firm type. We show that when a product is underpriced relative to consumer priors about quality, the low-quality firm exerts more manipulative effort than the high-quality firm. The opposite is true when the price is above the *ex ante* expected quality. In equilibrium, the type that introduces more bias into consumer reviews makes higher sales than under the benchmark case where manipulation is absent and review signals are unbiased. Yet, at extreme prices, both types enjoy higher sales compared to the no-manipulation benchmark; i.e., manipulation becomes 'effective' in each quality state.

Next, we study PBE where the low-quality type plays a mixed pricing strategy such that the price is partly informative about quality. These partially-separating PBE differ from each other not only with respect to the degree of review manipulation by each type but also the degree to which a given price is informative about quality. To the best of our knowledge, this is the first paper in the literature to combine signal jamming with price signaling as complementary tools for persuasive advertising. Most papers in the existing literature focus either on the signaling role of price or direct claims about quality on a discrete and bounded space (e.g., Rhodes and Wilson [2018]; Piccolo, Tedeschi and Ursino [2015] and [2017]; Janssen and Roy, [2017]). Even in models where both price and quality claims are used simultaneously, firms are not permitted to exaggerate their own quality by assumption. In our model, the firm can utilize both price and manipulated quality signals –on an unbounded space– concurrently to persuade consumers about its quality, and this signal-jamming approach leads to positive manipulation

[7] The noise refers to the idiosyncratic component in signals and reflects the variation among consumers with respect to the type of product review platforms they use or other factors that affect their information sampling process in ways that are hard to predict. In other words, the reviews consumers read can be thought of as containing a random bias whose strength and direction cannot be discerned by consumers.

not only by the low quality type but also by the firm with a high quality product. This feature is absent in most previous models of false advertising –with the exception of Dellarocas [2006]– and is the result of an implicit arms race between the high and low-quality firm types. There is good amount of evidence suggesting that even highly regarded companies use fake promotional reviews to boost their ratings. Amazon's noncompliant review dataset, for instance, contains many reviews for well-established brands.[8] These suspicious reviews for high-quality products may be the outcome of a strategic response by well-established suppliers to overall review inflation.[9]

In our model, the implicit arms race is fueled by the fact that consumers are uncertain about the firm type they face. Hence, they cannot condition their response to the bias in reviews on the firm type. The anticipation of review manipulation by one firm type induces further manipulative effort by the other type since rational consumers will seek higher quality signals to purchase the same product. This arms race leads to wasteful spending on review manipulation, which may cause a reduction in the firm's profits relative to the no-manipulation benchmark. We show that a firm would not insert bias into product reviews in at least one of the states if it could credibly pre-commit to a state-contingent manipulation schedule.

There are only a few papers featuring manipulative advertising whereby consumers cannot tell apart noise from the firm-induced bias in quality signals (Mayzlin [2006]; Dellarocas [2006]; Drugov and Troya-Martinez [2019]). Yet, due to the unique features of our analysis, this paper offers several insights that are not present in related work. The most important feature of our model is the novel mechanism through which prices shape the incentives for manipulation and total sales by each type relative to the other type. At higher prices, consumers need higher signals to become convinced to purchase the product; and higher signals about quality are more likely when the product is of higher quality. Therefore, at the margin, the mass of indifferent consumers a firm can persuade via manipulation will be relatively larger for the high-type firm. The reverse is true when prices are low.

This mechanism is not present in previous models of false advertising because the price effects we uncover in our signal-jamming framework do not arise when manipulation of consumer beliefs occurs only via price-signaling or quality claims that arrive without noise. In our model, the degree of manipulation is unobservable, and thus operates as hidden advertisement. Therefore, differential cost of advertising among different firm types does not necessarily lead to separation in the usual sense and separation of types can occur only via prices. Instead, both

[8] This dataset contains a random subset of product reviews posted on Amazon and marked by the platform as violating Amazon's review guidelines.

[9] A New York Times article titled 'In a Race to Out-Rave, 5-Star Web Reviews Go for $5' gives the example of a high-end hotel offering 10% discount to guests who agree to write an 'honest but positive review.' The article notes that 'The boundless demand for positive reviews has made the review system an arms race of sorts. As more five-star reviews are handed out, even more five-star reviews are needed. Few want to risk being left behind.'

the extent of separation by the high-type and of mimicking by the low-type firm depend continuously on the distance between signal means under the two states. When equilibrium prices are relatively high, this distance increases *vis-a-vis* the no-manipulation case, separation becomes stronger and the high-type enjoys higher sales. In pooling PBE with high prices, expected consumer surplus is higher relative to the no-manipulation benchmark because consumers purchase the high-quality product relatively more often. The opposite is true when equilibrium prices are low.

While there are other papers where manipulation or false advertising can sometimes benefit consumers (e.g., Piccolo, Tedeschi and Ursino [2015]; Rhodes and Wilson [2018]; Janssen and Roy [2017]), the underlying mechanism in our paper is different: The presence of manipulation improves *ex ante* consumer surplus at high prices via its positive effect on total demand for the high-quality product and negative effect on demand for the low-quality product. In partially-separating PBE, review manipulation is more detrimental for consumers because it is always effective for the low-quality firm. The latter uses price signaling as a complementary tool to shape consumer beliefs. These equilibria involve mixing by the low-type firm between a given pooling price and a low separating price that is equal to consumers' valuation for the low-quality product. Even when consumers observe very high pooling prices, the mixing probability in equilibrium will be so low that consumers' baseline quality expectations remain sufficiently high to sustain review manipulation as an effective tool to increase sales. This novel interaction between signal-jamming and price-signaling has not been analyzed in earlier work on advertising.

We discuss the impact of various policies that might be desirable from the perspective of (i) a firm that has not observed its type yet, (ii) an e-commerce platform that charges the firm a commission on its sales and (iii) a consumer protection agency who only cares about consumer surplus. We show that *ex ante*, a firm always expects to receive a higher profit if it could credibly reveal its type than under a pooling or partially-separating PBE. Similarly, we show that review platforms also prefer policies that induce a firm to reveal its type over the *status quo* where the low-type fully or partially mimics the high-type. The platform can, for instance, use quality certification requirements to achieve type revelation as long as monitoring costs are sufficiently low. A consumer protection agency, on the other hand, prefers pooling and partially-separating PBE outcomes over complete type revelation even when the former involves manipulated reviews. Yet, it may want to reduce manipulative effort in a pooling PBE with low prices and encourage it when prices are high.

The rest of the paper is organized as follows. Section II offers a brief review of the related literature. Section III lays out our benchmark model and the equilibrium definition. Section IV characterizes the pooling price PBE and presents our main results regarding the effect of manipulation on total demand and consumer surplus. It also provides the intuition for why and when manipulation can be effective in influencing aggregate consumer behavior. In Section V, we characterize the partially-separating PBE and compare them with pooling PBE with respect to manipulative efforts and consumer welfare. Section VI discusses the consequences of various policy interventions. In Section VII, we analyze a modified model where the firm has commitment power. Finally, Section VIII concludes the paper. The proofs of the main results as well as some of the technical details are contained in Appendices A and B at the end. All further results and their proofs are relegated to an Online Appendix which can be accessed on the *Journal*'s editorial web site.

II. RELATED LITERATURE

Regarding firms' incentives to manipulate online product reviews, we are aware of two closely related papers that take the signal-jamming approach as we take in this paper.[10] 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 of 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 by construction about product quality. In contrast, we endogenize the price in a way that allows it to carry information about product quality. Differently from these papers though, we demonstrate a novel mechanism through which price governs the relative marginal benefits of manipulation for each type. Also unlike Mayzlin [2006] and Dellarocas [2006], under sufficiently extreme price levels, our model features equilibria where manipulation increases sales for both firm types.

Our work is also related to models where firms can use price to signal their quality (e.g. Milgrom and Roberts [1986]; Bagwell and Riordan [1991]; Riordan and Judd [1994]). In Riordan and Judd [1994], a firm invests in quality improvement after consumers observe the first-period quality with some noise. This investment helps the firm shape consumer beliefs, but is otherwise different from manipulation via biased product reviews both conceptually and

[10] Signal-jamming models have been analyzed in the literature in various other contexts. Some of the relevant studies are Matthews and Mirman [1983], Fudenberg and Tirole [1986], Holmström [1999], Edmond [2013], Caselli et al. [2014] and Aköz and Arbatlı [2016].

in terms of its welfare implications. In our model, manipulation affects demand without any quality improvement and can harm the consumer when exercised by the low-quality type.

There are a few recent papers that feature manipulative advertising (Gardete [2013]; Rhodes and Wilson [2018]; Piccolo, Tedeschi and Ursino [2015] and [2017]; Janssen and Roy [2017]). These papers are particularly related to our analysis of equilibria where prices are partially informative. Some of the equilibria in these papers allow deceptive advertising by the low-quality firm that affects consumer posterior beliefs about product quality. Although these models allow 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, and unlike the aforementioned studies, our paper features advertising in the form of hidden action.

More broadly, our paper is 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 in equilibrium, 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 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 [2019] 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.[11]

Another branch of the advertising literature studies dissipative advertising where spending on advertising indirectly signals quality (e.g., Nelson [1974]; 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]; Johnson and Myatt [2006][12]), we assume that the firm observes its product quality and conditions the bias level on its quality. [13]

We show that a firm would not manipulate product reviews in at least one of the quality states, if it could commit to do so. This relates our paper to Miklós-Thal and Zhang [2013] who

[11] See also Grunewald and Krakel [2017] for a related duopoly analysis.

[12] Manipulation 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 are quite different. See Figure 2 in subsection IV(i) and the discussion therein.

[13] See Renault [2016] for an overview of the recent literature on advertising.

argue that demarketing may be a desirable strategy for firms when consumers partly attribute product success to successful marketing. While Miklós-Thal and Zhang [2013] assume that marketing is persuasive, our paper explores the conditions under which it is so.

III. MODEL

Consider a firm releasing a new product whose quality is unknown to consumers. There is a continuum of consumers with unit measure each of whom 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.[14]

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, they 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. 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 firm first observes the quality v_j of its product. Then it sets a price p_j and exerts some effort to manipulate the reviews of its product. 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 in the mean of all signals that consumers receive. In particular, we assume that each private signal x_i has three components: (i) true quality of the product v_j, (ii) the bias b_j inserted by the firm, and (iii) a consumer-specific variation in the information collected, which we denote as $\varepsilon_i \in \mathbb{R}$ for consumer i. We assume that the noisy signal x_i is additively separable in these three components so that $x_i = v_j + b_j + \varepsilon_i$. Consumer-specific variation ε_i is distributed independently across consumers and generated by a known cumulative distribution function F and a corresponding density function f.[15] Assumption 1 below states the restrictions we place on the noise distribution.

[14] Our main results can be generalized –after an appropriate normalization– to the case where the high-quality firm has a higher marginal cost of production.

[15] Informational heterogeneity that we assume here is not critical for the results. We could assume that all consumers face a common noise ε in the information they receive so that all consumers receive the same signal. In this case, 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 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.

Assumption 1 *The density function f for the idiosyncratic noise ε_i is continuous, log-concave, symmetric around zero, has unbounded support and finite moments*[16]*, and satisfies the tail property* $\lim_{x\to-\infty}\frac{f'(x)}{f(x)}=\infty$[17].

Upon observing the unit price p_j and 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

$$u=v_j-p_j. \tag{1}$$

If the product is not purchased, then $u=0$.[18]

We assume, without loss of generality, that a consumer buys the product when indifferent. We denote the binary purchase decision of the consumer by a function $s(x_i,p_j)\in\{0,1\}$ such that $s(x_i,p_j)=1$ if i purchases the product. Therefore, from the perspective of the type-j firm, the total amount of sales can be written as

$$S(v_j,b_j,p_j)\equiv\int_0^1 s(x_i,p_j)di=\int_0^1 s(v_j+b_j+\varepsilon_i,p_j)di. \tag{2}$$

The profit to the firm is given by

$$\pi_j=p_jS(v_j,b_j,p_j)-C(b_j), \tag{3}$$

where $C(.)$ is the cost associated with the bias b_j. This cost function encompasses (i) the direct costs of hiring employees or fake reviewers as well as incentivizing consumers to promote the product, (ii) the opportunity cost of time spent on online review forums, (iii) the cost of strategic research expenditures (either through funding research projects or directly conducting research in an R&D department of the firm), and (iv) the expected fines and reputation costs if the firm is caught manipulating its product reviews. 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

[16]Note that log-concavity implies unimodality as well. See An [1998] for further discussion regarding the relation between log-concavity and unimodality.

[17]Note that normal distribution satisfies both strict log-concavity and the tail property.

[18]It is also possible to introduce preference heterogeneity 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, ψ_i is the individual match quality between the consumer and the product. See the discussion in Section VIII.

state the restrictions we place on the cost function.[19]

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 any 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.

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

Assumption 3 imposes a lower bound on the convexity of the cost of manipulative effort, which guarantees that the profit function is strictly concave in the level of bias. This assumption is useful in ruling out multiple equilibria which are qualitatively similar but feature different levels of manipulative effort.[20]

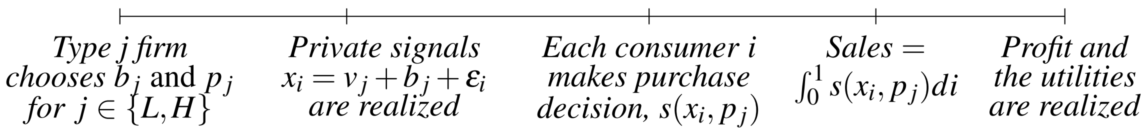

Figure 1: Timeline

The timeline of the game is illustrated in Figure 1. We employ the standard perfect Bayesian equilibrium (PBE) as the solution concept for our analysis. Intuitively, PBE requires sequential rationality and Bayesian updating for posterior beliefs whenever possible. We allow for mixed pricing strategies over 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' purchasing decisions are symmetric.[21] Each PBE in our

[19] Some manipulative activities could differ in terms of their costs across the firm types. For example, it could be easier for the high-type firm to compensate consumers in return for their promotional online reviews. Similarly, reputation costs, relative to the expected revenue, could be lower for the low-type firm. It is straightforward to extend our analysis to heterogeneous manipulation costs. Theorems 1 and 3 regarding the existence of equilibria with positive manipulation would not change. The existence and uniqueness of the pricing thresholds stated in Theorem 2 would also continue to hold, but their levels would now also depend on the difference between the costs.

[20] Suppose that the error distribution is standard normal, $C(b) = b^2$ and $v_L = 0$, $v_H = 1$. Then, both Assumptions 2 and 3 will be satisfied since $v_H \max_{x \in \mathbb{R}} f'(x) < 0.25$.

[21] 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. Moreover, we assume throughout the analysis the most pessimistic off-equilibrium beliefs such that any deviation is associated with the low-type firm.

model consists of (i) a profit-maximizing pricing decisions by each firm type –potentially involving mixed strategies– given the bias levels b_L and b_H, and the purchasing rule used by each consumer; (ii) a profit-maximizing bias level for each firm type at each price level that is chosen with positive probability; (iii) posterior beliefs about product quality conditional on the signal realization x and observed price p, formed via Bayesian updating whenever possible; and (iv) a purchasing decision by each consumer as a best response to pricing and bias strategies of each firm type and the realized quality signal x. The formal statement of the equilibrium definition is relegated to Appendix B.

IV. ANALYSIS OF POOLING EQUILIBRIA

The informativeness of the price depends on the equilibrium coordination of expectations. If consumers do not expect to learn anything from the price, they will consult only their private signals to make inferences. The firm then does not have any incentives to set a price different from what consumers have anticipated.

Before we prove the existence and characterize the properties of a pooling PBE in which both firm types charge the same price, it is instructive to lay out the conditions that govern the behavior of the firm and the consumers in such a common price equilibrium with $p_L = p_H = \bar{p}$. Since here the price does not convey any information about quality, consumer expectations about quality would only depend on private signals and prior beliefs. First, note that consumer i buys the product if and only if her posterior expectation of product quality is greater than the price, i.e., $\mathbb{E}(v|x_i) \geq \bar{p}$. Therefore, the market demand is determined by the distribution of signals and how consumers interpret them. Assumption 1 imposes some regularity conditions on posterior beliefs that consumers could have. In particular, posterior expectation of quality is strictly increasing in the level of quality signal x when the noise distribution is log-concave. As a result, the purchasing decision admits a simple monotonic threshold structure. To see how this behavior can be supported in equilibrium, first suppose that consumers indeed follow a monotonic strategy in which they purchase the product whenever their private signal x_i exceeds a common threshold $\bar{x}$. We show that the resulting profit-maximizing bias level for each type is strictly positive, and the mean quality signal in the high state always lies above the mean signal in the low state, i.e., $v_H + b_H > v_L + b_L$. If the opposite were true, consumers would purchase only when their signal was lower than a given threshold, which eliminates any incentive by the low-type firm to manipulate the product reviews, and therefore contradicts with the initial supposition that $v_H + b_H < v_L + b_L$. Thus, average signals must be monotonically increasing in quality, which in turn implies, under log-concavity of the noise distribution, that quality signals when the true product quality is high first-order stochastically dominate the signals when it is low. This means that a high signal will always generate a higher quality expectation than a low signal. This confirms the existence of a unique purchase threshold as postulated above. In what follows, we state these points more formally so as to prove the existence and uniqueness of a pooling price PBE.

Suppose 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., $x_i \geq \bar{x}$. Then the optimal bias level b_j solves

$$\max_{b_j \geq 0} \quad \bar{p}[1 - F(\bar{x} - v_j - b_j)] - C(b_j) \quad \text{for each } j \in \{L, H\}. \tag{4}$$

The first-order condition to this problem is given by

$$\bar{p} f(\bar{x} - v_j - b_j) = C'(b_j) \quad \text{for each } j \in \{L, H\}. \tag{5}$$

Assumptions 1 and 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 best-reply of the firm is unique.

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

$$\mathbb{E}(v|x) = \frac{\sum_j v_j f(x - v_j - b_j) P(v = v_j)}{\sum_j f(x - v_j - b_j) P(v = v_j)}. \tag{6}$$

Therefore, at a pooling 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

$$\begin{aligned} \mathbb{E}(v|\bar{x}) = \frac{\sum_j v_j f(\bar{x} - v_j - b_j) P(v = v_j)}{\sum_j f(\bar{x} - v_j - b_j) P(v = v_j)} &= \bar{p} \Leftrightarrow \\ \sum_j (v_j - \bar{p}) f(\bar{x} - v_j - b_j) P(v = v_j) &= 0. \end{aligned} \tag{7}$$

This condition pins down the purchase threshold as an implicit function $\bar{x} = \bar{x}(b_L, b_H, \bar{p})$ of the bias levels and the common price. The three equations given in (5) and (7) determine the equilibrium with biased product reviews.

In the absence of manipulation, consumers' posterior expectation of 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 $\underline{x}$ in the absence of any manipulation is then implicitly determined by the following equality:

$$\begin{aligned}
\mathbb{E}(v|\underline{x}) = \frac{\sum_j v_j f(\underline{x}-v_j)P(v=v_j)}{\sum_j f(\underline{x}-v_j)P(v=v_j)} &= \bar{p} \Leftrightarrow \\
\sum_j (v_j-\bar{p}) f(\underline{x}-v_j)P(v=v_j) &= 0.
\end{aligned} \tag{8}$$

Consumers' problem of estimating product quality is not trivial. A Bayesian consumer knows that 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 that the firm is of high type. 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 of the noise ε would lead to a higher signal x when the underlying quality is high, i.e., $v_H + b_H + \varepsilon > v_L + b_L + \varepsilon$. Lemma 1 below proves that posterior expectations are indeed increasing in signals if $v_H + b_H > v_L + b_L$.

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

We present the proof for Lemma 1 and all other omitted proofs in Appendix A. Lemma 1 states that posterior expectations of consumers are monotonic in the signal they observe. This is a consequence of the monotone-likelihood ratio property of the probability distribution function $f(\cdot)$ guaranteed by the log-concavity assumption we impose in Assumption 1. An immediate consequence of Lemma 1 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 3 hold and that $b_L + v_L < b_H + v_H$. Then, for each price level $\bar{p} \in (v_L, v_H)$, there is a unique threshold $\bar{x}$ with $\mathbb{E}(v|\bar{x}) = \bar{p}$ such that only those consumers with $x_i \geq \bar{x}$ purchase the product.*

The next question is whether the inequality $b_L + v_L < b_H + v_H$ holds in every PBE. Lemma 2 shows that this is indeed the case.

Lemma 2 *Suppose that Assumptions 1 and 2 hold. Then, in any PBE the mean signal about product quality is higher when the firm is of high type than when it is of low type, i.e., $b_H + v_H > b_L + v_L$.*

Since $v_H > v_L$, the reverse inequality $b_L + v_L \geq b_H + v_H$ would imply that the L-type firm would do excessive manipulation. In such a case, it is possible to show under the assumption of log-concave $f(\cdot)$ that consumers' evaluation of the signals would be reversed; i.e., higher signals would imply a higher likelihood that the firm is of low type. For such inferences, L-type would not have incentives to shift the mean signal upwards, and therefore, $b_L + v_L \geq b_H + v_H$ cannot be part of any PBE.

Now, we can establish the existence of a pooling PBE with strictly positive bias levels conditional on the price $\bar{p}$ by combining the first-order conditions in (5) and Corollary 1. Theorem 1 below lays out the conditions for existence and uniqueness of such an equilibrium.

Theorem 1 (Existence and uniqueness of pooling PBE) *Suppose that Assumptions 1, 2 and 3 hold. Then,*

i. For each admissible pooling price $\bar{p} \in (v_L, v_H)$, the PBE is unique and features 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 $x_i \geq \bar{x}$.

ii. There always exist pooling PBE for prices that are close enough to v_L. When $v_L = 0$, a pooling PBE exists for any price in the interval $(0, v_H)$. If $v_L > 0$, on the other hand, there is a threshold $\tilde{p}$ such that a pooling PBE exists only for $\bar{p} \in (v_L, \tilde{p})$.

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. As v_L gets larger, it becomes harder to sustain a pooling PBE. Indeed, Lemma 6 in Appendix A shows that when $v_L \geq v_H/2$, prices that are greater than $\mathbb{E}(v)$ cannot be supported by any pooling PBE. On the other hand, if we assume the tail property $\lim_{x\to-\infty} \frac{f'(x)}{f(x)} = \infty$, we can specify how the profit function behaves at extreme prices that are close to v_L and show the existence of a pooling PBE for sufficiently low prices even when $v_L > 0$.

One important implication of Theorem 1 is that equilibrium bias levels are always strictly positive. In other words, the firm chooses to spend some of its resources on manipulating product reviews regardless of its product quality. This result distinguishes our model from the advertising literature and most variants of false advertising models where only the low-quality types engage in advertising (see, for example, Rhodes and Wilson [2018]). The reason underlying our finding is the implicit arms race between the two firm types. Given consumers' purchasing threshold signal $\bar{x}$, a given firm type benefits from shifting the mean signal via manipulation. But since $\bar{x}$ is adjusted upwards to account for such manipulation, the other type is also compelled to respond with a strictly positive amount of manipulation. As will be discussed in Section VII, strictly positive manipulation in both states is suboptimal from an *ex ante* perspective, but is unavoidable unless the firm has commitment power.

The following proposition states that in every pooling PBE the high-quality firm earns a higher profit than the low-quality type even when the former exerts more effort to bias product reviews and thus bears 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 is driven by the quality advantage of the high-type firm. Given that consumers use a unique purchasing threshold $\bar{x}$ and face the same price under both quality realizations, it follows from Lemma 2 that the high-quality firm will always make higher 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 amount of bias in 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. However, neither Lemma 2 nor Proposition 1 implies a universal ranking of bias levels by quality. In fact, in the next section, we show that this ranking depends on the level of equilibrium price.

IV(i). *The effect of biased product reviews on sales*

In this section, we present and discuss our central results on the effects of biased product reviews in the pooling price PBE. 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 level of equilibrium 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 firm type from manipulation. As we demonstrate later in this section, without this price effect on competitive advantage, manipulative efforts would not differ across types and would have no effect on sales or consumer surplus. The following proposition describes how the price level governs the relative bias each firm type inserts in product reviews in a pooling PBE.

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 $\bar{p} >$ (<) $\mathbb{E}(v)$, $b_H >$ (<) 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 level. Keeping the firm's actions fixed, the purchasing 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 purchasing the high-quality product 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 consumer indifference condition in (7) rewritten in the following form:

$$(v_H - \bar{p})(1-g)f(\bar{x} - b_H - v_H) = (\bar{p} - v_L)gf(\bar{x} - b_L - v_L). \tag{9}$$

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 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 (5), that $b_L > b_H$. This result reflects the fact that at the margin, the L-type has more to gain from manipulation when $\bar{p} < \mathbb{E}(v)$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.

Theorem 1 has established that both firm types exert some manipulative effort in every pooling PBE, and Proposition 2 has shown how these efforts change with the price level. However, neither of these results necessarily implies that review manipulation affects total sales. We know by Proposition 2 that both firm types insert the same level of bias into product reviews when the price coincides with the prior expected quality $\mathbb{E}(v)$. Consumers' posterior beliefs will then be the same as when $b_L = b_H = 0$. That is, at this price, both types exert effort to bias the reviews, but these efforts have no effect on consumer beliefs nor the market demand.

Under which pooling PBE, if any, can a given firm type increase its sales via manipulation? To answer this question, we fix a price level and compare consumer demands under two scenarios: one where manipulative efforts are unrestricted (manipulation equilibrium) and one in which manipulative effort is restricted to be 0 for both firm types (no-manipulation benchmark).[22] Specifically, we will say that type j does effective manipulation at price $\bar{p}$ if

$$1 - F(\bar{x} - b_j - v_j) > 1 - F(\underline{x} - v_j) \Leftrightarrow b_j > \bar{x} - \underline{x}. \tag{10}$$

[22] When $v_L = 0$, this comparison applies to any price level between v_L and v_H as all of these prices can be supported as a pooling PBE with or without manipulation. When $v_L > 0$, on the other hand, the demand effect of manipulation is well defined only for those prices that are admissible (i.e., can be supported as a pooling PBE) both under manipulation and no-manipulation cases.

Recall that threshold $\underline{x}$ is the signal realization that makes a consumer indifferent between purchasing and not purchasing when there is no manipulation. It is uniquely defined by equation (8). Therefore, $\bar{x} - \underline{x}$ is the average adjustment in consumers' purchasing 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 firm type depends on the equilibrium 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 benchmark. Our second result is that in equilibria where prices are either sufficiently high or sufficiently low, 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 (Effect of manipulation on firm sales in a pooling PBE) *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 $\bar{x} = v_H + (b_L + b_H)/2$ 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 the L-type firm exerts more manipulative effort than the H-type at prices lower than $\mathbb{E}(v)$. Since consumers cannot adjust their beliefs separately for each firm type, their adjustment in the purchasing threshold signal, $\bar{x} - \underline{x}$, falls short of the bias inserted by the L-type. As a result, L-type increases its demand in this region compared to the no-manipulation benchmark. What is more surprising is that the L-type firm is able to increase its sales also for sufficiently high prices. This is normally where the H-type firm is expected to increase its sales because, by Proposition 2, it exerts more manipulative effort than the L-type. But at sufficiently high prices, a consumer needs a very high signal to be persuaded to purchase the product, and the H-type's higher efforts to manipulate in this range increases the likelihood of receiving higher signals.

To see the mechanism behind effective manipulation more concretely, consider the indifferent consumer in the no-manipulation case, who observes a signal realization $x = \underline{x}$ and has a posterior expected quality equal to $\bar{p}$.[23] Now, allow for manipulation. Given the anticipated equilibrium bias levels b_L and b_H, if a signal realization $x = \underline{x} + b_L$ induces a consumer to purchase the product at the same price $\bar{p}$, then it means the L-type firm has effectively manipulated consumers' beliefs in a way that increases its sales. This is so because if $\mathbb{E}(v \mid \underline{x} + b_L) > \bar{p}$, the new indifferent consumer must have observed a signal realization $\bar{x}$ that is strictly less than $\underline{x} + b_L$. A signal realization $x = \underline{x} + b_L$ under manipulation and $x = \underline{x}$ under no-manipulation are equally likely when the true quality is low. Hence, manipulation can 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., $f(\underline{x} + b_L - b_H - v_H) > f(\underline{x} - v_H)$. When $\bar{p} < \mathbb{E}(v)$, this result follows easily from unimodality of $f(\cdot)$ because $b_L > b_H$ and $\underline{x} < v_H$. When $\bar{p} > \mathbb{E}(v)$, on the other hand, $b_H > b_L$ and so it holds only if $\underline{x}$ is sufficiently high. In the limit, for instance, $\underline{x}$ diverges to ∞ as $\bar{p}$ converges to v_H,[24] and thus $f(\underline{x} + b_L - b_H - v_H) > f(\underline{x} - v_H)$ for any finite $b_L < b_H$ and v_H since we are in the right tail of $f(\cdot)$. Finally, to see why manipulation cannot reduce the sales for both firm types, we can again look at the indifferent consumer in the no-manipulation case. Suppose for a moment that there is a price for which manipulation leads to lower demand for both types. This necessarily implies that one of the firm types must be doing more manipulation than the other because when manipulation levels are the same, manipulation has no effect on demand. If this is the high-type firm, a lower demand with manipulation means that $f(\underline{x} + b_H - b_L - v_H) > f(\underline{x} - v_H)$. But, by unimodality of f, the high-type firm can reduce its manipulation level below b_L to reverse this inequality. The argument for the low-type firm is symmetric. Intuitively, since there are no demand effects when both firm types choose the same manipulation level, a firm type will be tempted to do more manipulation than the other type only if this is expected to generate a higher demand than the no-manipulation case. So if there is any manipulation in equilibrium, then demand must be higher for at least one of the firm types.

It is important to note the differences between our analysis of effective manipulation and Bayesian persuasion (as in Kamenica and Gentzkow [2011]) of a fixed consumer. Intuitively, when a firm wants to persuade a consumer into purchasing its product, it should provide some information so that the consumer can use the signal that the firm generates. Bayesian plausi-

[23] 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.

[24] See Lemma 4 in Appendix .

bility (p.2594 in Kamenica and Gentzkow [2011]) would then imply that only one of the two types could 'gain' from persuasion. A similar mechanism works in our model when we fix a consumer and examine her posterior expectation with and without manipulation. However, in our model the indifferent consumer who determines the market demand also depends on price. For example, if we fix the 'median' consumer whose noise ε_i is drawn as 0, then it is possible to show that $b_L > b_H$ implies that the posterior expected quality for this consumer is higher than that when there is no manipulation if and only if the firm is indeed of low type. Therefore, when $b_L > b_H$, it is only the low-type firm which can 'persuade' the median consumer. However, the indifferent consumer, who determines the size of the demand, is not necessarily the median consumer. Therefore, for extreme prices, the low-type firm can increase its demand even if it can persuade only a minority of consumers. Note that such a price effect is absent in the canonical Bayesian persuasion model where there is a fixed receiver.

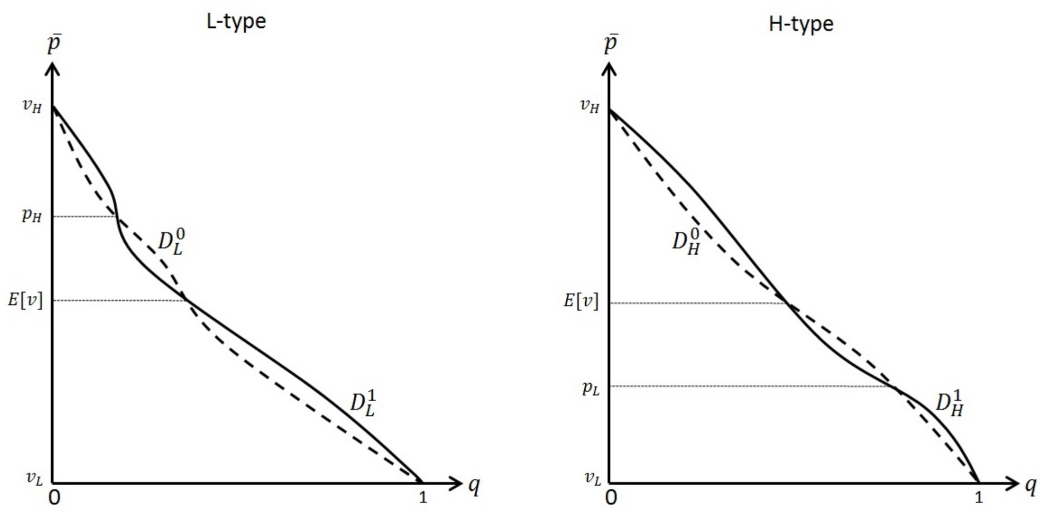

Figure 2: Demand curve with and without manipulation

An alternative way to illustrate Theorem 2 is to show how the demand curve changes after manipulation. In Figure 2 we plot the demand curves before and after manipulation for each firm type. In the figure, D^0 represents the demand curve before manipulation, so $D_j^0 = 1 - F(\underline{x} - v_j)$. Similarly, D^1 represents the demand curve after manipulation, so $D_j^1 = 1 - F(\bar{x} - v_j - a_j)$. As we can see from the figure, manipulation causes a change in the demand curve for both firm types. This role of advertising was the main focus of Johnson and Myatt [2006], where they argue that many economic activities, including advertising, change the dispersion of consumer valuations and, in turn, cause a rotation in the demand curve. In a fairly general model, they show that monopoly profits are a U-shaped function of the dispersion of consumer valuations. This means that the seller pursues either maximal dispersion (niche-market strategy), serving high-value consumers at a high price, or minimal dispersion (mass-market strategy), serving a large fraction of consumers at a lower price. However, the particular way rotation happens in our analysis and the reasons behind it are quite different than Johnson and Myatt [2006]. 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 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. Since consumers anticipate the low-quality [high-quality] monopolist to engage in relatively more manipulative advertising at low [high] prices, they downgrade [upgrade] their quality expectations at these prices. This, in turn, causes the quantity demanded by the low-quality [high-quality] type to go down [up] at low prices and go up [down] at high prices.

IV(ii). *The effect of biased product reviews on consumer surplus*

To understand the effect of review manipulation on consumer surplus, we first identify the net effect of manipulation on consumer surplus in a pooling PBE. To do that we compare, for a given price, the aggregate *ex ante* consumer surplus under a pooling PBE with manipulation and the consumer surplus under the no-manipulation benchmark. That is we compare

$$CS_{ea}^{M} = (1-g)(v_H - \bar{p})[1 - F(\bar{x} - b_H - v_H)] - g(\bar{p} - v_L)[1 - F(\bar{x} - b_L - v_L)], \tag{11}$$

and

$$CS_{ea}^{N} = (1-g)(v_H - \bar{p})[1 - F(\underline{x} - v_H)] - g(\bar{p} - v_L)[1 - F(\underline{x} - v_L)]. \tag{12}$$

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 *In a PBE with a pooling price $\bar{p}$, the net effect of manipulation on ex ante consumer surplus is negative when $p_L < \bar{p} < (1-g)v_H + gv_L = \mathbb{E}(v)$ and positive when $\mathbb{E}(v) < \bar{p} < p_H$.*

If manipulation is effective (i.e., demand-increasing) for the L-type but not for the H-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.

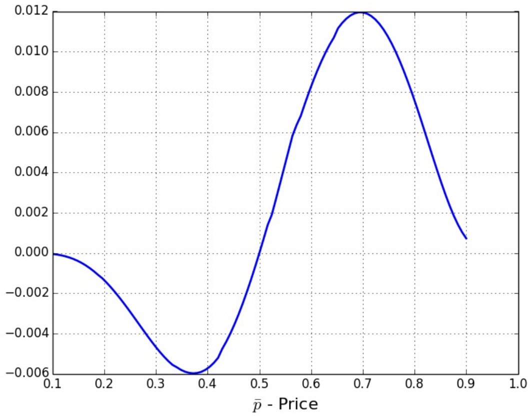

Figure 3: The net effect of manipulation on *ex-ante* consumer surplus

When $v_L > 0$, highest prices cannot be supported by pooling PBE. Does this help narrow down our predictions about the welfare implications of manipulation for pooling PBE? The answer is affirmative. Lemma 6 in Appendix A shows that when v_L is large enough (i.e., $v_L \geq v_H/2$), no price that is greater than the *ex ante* expected quality level can be supported as a pooling PBE. Therefore, all pooling price PBE, where the net effect of manipulation is positive cease to exist for large enough v_L.

When manipulation is effective for both types, i.e., when $\bar{p} < p_L$ or $\bar{p} > p_H$ as stated in Theorem 2, the resulting welfare effects are ambiguous. 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 how exactly *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 3 confirms that the net effect of manipulation on *ex ante* consumer surplus is negative when the price is below the *ex ante* expected quality and positive when the price is above it. However, the welfare loss in the former case ($\bar{p} < \mathbb{E}(v)$) as well as the welfare gain in the latter case ($\bar{p} > \mathbb{E}(v)$) diminishes as the price approaches its respective lower and upper bounds of v_L and v_H, respectively.

V. ANALYSIS OF PARTIALLY-SEPARATING EQUILIBRIA

So far we have focused on pooling price equilibria and analyzed how the nature of the manipulation race between the two firm types and its implications for firm sales and consumer surplus change as we move along the range of admissible prices. This was a natural starting point to illustrate the critical role prices play for equilibrium determination even when they are uninformative about quality.

In principle, the firm may want to use its price as an additional tool for managing the expectations of consumers. For example, the high-type firm might consider using 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. Therefore, there is no separating PBE where the firm can earn a positive profit (see Proposition 5 in Appendix B).[25]

Even if prices cannot be fully informative, they can be partially informative and thus allow some degree of separation between the types. The idea is that when we allow the firm to choose a probabilistic pricing strategy, the low-type firm may use a mixed pricing rule that mimics the hypothetical behavior of the high-type with some probability. Theorem 3 below describes such PBE and in particular shows that any price that cannot be supported by a pooling price PBE can be supported by partially-separating mixed PBE.

Theorem 3 (Partially-separating PBE) *Suppose that Assumptions 1, 2 and 3 hold and* $v_L > 0$*. For any* $\bar{p} \in (v_L, v_H)$*, there exists a partially-separating PBE whenever a pooling PBE does not exist. In this PBE, the high-quality firm type plays a pure strategy given by 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* $\bar{\alpha}$ *and the signal threshold* $\bar{x}$ *satisfy the following system of equations:*

$$v_L = \bar{p}(1 - F(\bar{x} - b_L - v_L)) - C(b_L) \tag{13}$$

$$(v_H - \bar{p})(1-g) f(\bar{x} - b_H - v_H) = \bar{\alpha} g (\bar{p} - v_L) f(\bar{x} - b_L - v_L), \tag{14}$$

while the bias levels b_L *and* b_H *are determined by the first-order conditions that are same as the equations in (5). If* $v_L = 0$*, on the other hand, there does not exist any partially-separating PBE.*

When $v_L > 0$, pooling PBE cannot be supported at high prices, while partially-separating

[25] Allowing for preference heterogeneity is one of the ways of generating separating PBE. See Gardete [2013] for a related model with vertical differentiation, where it is costly for the low-quality type to mimic the high-type. Another way to induce revealing prices is to tie the flow of information to fixed incentives. We discuss relevant extensions in Section VI.

PBE exist only at high prices.[26] Hence, it appears that partially-separating PBE and pooling PBE can be supported for two distinct sets of prices on the interval $[v_L, v_H]$, and at a given price these two types of equilibria are unlikely to co-exist. To see why this might be true, note that in any pooling PBE, by definition, the firm must be making more profits than v_L because otherwise it would deviate to a price of $p_L = v_L$ and earn a profit of v_L. In contrast, in all (strictly) partially-separating PBE, the firm must earn a profit exactly equal to v_L since otherwise it would not mix between v_L and a higher price. So, partially-separating PBE arise as a way to ensure a profit of v_L whenever the pooling profit falls below v_L.[27]

In a partially-separating PBE, 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 normally the H-type firm would 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 [2018] consider a similar false-advertising framework where the low-type firm uses a misleading mixed strategy. However, in our model, consumers do not constrain themselves merely to the information they obtain from the price since they also have access to biased information through product review platforms. Therefore relative incentives to bias quality signals and hence the ranking of manipulative efforts b_H and b_L under partially-separating PBE depends on the level of $\bar{p}$ in the same way as in pooling PBE.

Moreover, in partially-separating PBE, the quality of information that the price conveys to consumers interacts with manipulation incentives. In particular, the informativeness of the observed price increases with $\bar{p}$. By inspecting equation (14), one can notice that as $\bar{p}$ increases, all else equal, the L-type firm chooses $\bar{p}$ with a lower probability. As a result, when consumers observe a higher $\bar{p}$, they become more certain that the product is of high quality and demand more of it. The combination of a higher price and higher demand means greater revenue for both types of the firm. This suggests that the high-type should be earning higher profits at higher prices and moreover its profit should always be greater than v_L. On the other hand, since, as implied by equation (13), the L-type firm's profit must stay constant across all partially-separating PBE, the high-type firm always earns a higher profit than the low-type, thus extending Proposition 1 to partially-separating PBE. A partially-separating PBE with a higher

[26] In Lemma 6 in Appendix , we provide a sufficient condition for finding an upper-bound on prices for which a pooling PBE can be supported.

[27] The mechanism through which this works is simple. Suppose that fully pooling profit is below v_L. In such a case, the firm can increase its profit by lowering the probability with which it chooses this price below 1, thereby increasing consumers' posterior beliefs about quality, all else equal. The equilibrium value of $\alpha \in (0,1)$ solves exactly this trade-off. Unfortunately, the nonlinearity of equations (13) and (14) prevents us from solving the equilibrium level of α to analytically compare the existence of pooling and partially-separating PBE. Therefore, although we can argue, in the absence of manipulation, that pooling and partially-separating PBE cannot co-exist at a given price, we cannot rule out this possibility when there is manipulation.

revenue for the L-type must also feature higher spending on manipulation to counter-balance the higher revenues. As a result, the equilibrium bias level by the L-type firm as well as the profit of the H-type firm increases with the price level (see Proposition 6 in Appendix B).

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 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 consumer beliefs via manipulative effort. Like in the benchmark model, the H-type firm exerts more manipulative effort than the L-type at sufficiently high prices.

When analyzing pooling PBE, we have measured effectiveness of manipulation at each admissible price by comparing the equilibrium demand under manipulated reviews to the demand when manipulation is absent. Since jamming the product review signals is the only manipulation channel when the price is uninformative, this was the most natural comparison to make. However, in the partially-separating PBE that we consider in this section, a firm can manipulate consumer beliefs both via 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 (14), 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 unchanged, not only will the consumers use a different purchasing threshold signal $\underline{x} \neq \bar{x}$ 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 $\underline{\alpha}$. Therefore, we say that j-type firm's product review manipulation is effective if the marginal contribution of j-type firm's manipulative effort to its *ex post* demand is positive compared to the corresponding partially-separating PBE at the same price but without any review manipulation.

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. It is straightforward to show that, for a relatively high price $\bar{p}$ that is close enough to v_H, a partially-separating PBE without manipulation exists. Thus, when we make statements about the effectiveness of manipulation in this section, we implicitly assume that $\bar{p}$ admits a partially-separating PBE with manipulation as well as without. 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 $1 - F(\bar{x} - b_j - v_j) > 1 - F(\underline{x} - v_j) \Leftrightarrow b_j > \bar{x} - \underline{x}$.

The following proposition characterizes when the H-type firm exerts more manipulative effort than the L-type firm in a partially-separating PBE, and when manipulation is effective for each type.

Proposition 3 (Effective manipulation in a partially-separating PBE) *The bias levels $b_H = b_L$ if and only if*

$$\bar{p} = \frac{v_H + \bar{\alpha} v_L}{1+\bar{\alpha}} \equiv \hat{p}_{\bar{\alpha}}. \tag{15}$$

For every price level $\bar{p} > (<)\ \hat{p}_{\bar{\alpha}}$, $b_H > (<)\ b_L$. Manipulation is always effective for the L-type firm and is effective for the H-type only when $\bar{p} \geq \hat{p}_{\bar{\alpha}}$.

This result highlights an important common feature between pooling and partially-separating PBE and a distinction. The common feature is that for relatively higher prices, high-type firm inserts more bias into reviews than the low-type, and vice-versa for relatively lower prices. This has implied in pooling PBE that for some of the higher prices only the high-type can achieve effective manipulation. On the other hand, Proposition 3 shows that low-type can always do effective manipulation in any partially-separating PBE. The intuition behind this result is that the L-type firm can influence consumer beliefs not only through its manipulative effort but also the probability with which it mimics the H-type. Therefore, even for extremely high prices, the L-type firm is able to use a combination of these two channels to tilt the market demand.

VI. WELFARE IMPLICATIONS AND POLICY

In this section, we discuss the welfare implications of manipulated product reviews as described by our model and various policy tools that different policy-makers may want to employ in this context. We group policy tools in terms of their consequences for equilibrium outcomes. The types of policies we consider are: (1) revelation of firm type, (2) reduction in the level of review manipulation by both firm types or only the low-type, and (3) increasing the precision of consumer signals. We discuss how much each of these policy tools serves the objectives of policy-makers. We analyze the intervention of two types of policy-makers which differ by their objectives. The first policy-maker is what we term as the 'consumer protection agency,' who only cares about the *ex ante* consumer surplus. The second policy-maker is a 'platform,' and represents e-commerce platforms which typically charge a commission on firms' sales (e.g. online travel agencies such as Booking.com). Therefore, we assume that the platform-owner's aim is to increase the *ex ante* revenue of the firm.[28] In what follows, we assume that both types of agents are able to fully commit to the policies they announce.

[28]For now, we ignore any small fixed fees that a platform may wish to charge firms. If a fixed fee changes the behavior of the firm, it can do so only by discouraging the firm from using the platform. While a policy that would discourage both firm types would never be optimal, it is possible for the platform to set a fixed fee so as to screen out the low-type firm. We revisit this scenario later in the section when we discuss targeted policies.

VI(i). *Type-revealing policies*

Other things equal, do consumers, firms and platforms prefer PBE where firm types are revealed over PBE in which there is no or only partial separation? The following proposition answers this question by comparing the *ex ante* profits (for the firm), revenues (for the online platform) and consumer surplus under full separation versus under pooling or partially-separating equilibria.

Proposition 4 *Ex ante consumer surplus is always higher, whereas ex ante profit and revenue of the firm are always lower in a pooling or partially-separating PBE than they are when the firm can pre-commit to a revealing pricing strategy.*

In a PBE with revealing prices, all uncertainty about product quality vanishes, and this enables type-j firm to capture all consumer surplus, without any need for manipulation, by charging a price of v_j. In contrast, in any pooling or partially-separating PBE, the firm has to leave some positive surplus to the consumers because of ensuing uncertainty. Thus, whether a decision-maker would ultimately choose to implement a type-revealing policy or not depends inherently on its objective and the associated costs of its implementation.

Consumer protection agency: A consumer protection agency that aims to maximize consumer surplus would never want to induce perfect type revelation. Instead, it would want to implement a PBE with manipulation as in our benchmark model –even when manipulative efforts are positive. While price pooling opens the door to review manipulation, at the same time it protects consumers by hindering the high-type firm from extracting their surplus fully. Although in equilibrium consumers will incur a utility loss when they purchase the low-quality product, they will enjoy a positive net utility when the product quality turns high. And, by Proposition 4, the latter gain dominates the former loss from an *ex ante* perspective.

Online platform: Unlike a consumer protection agency, online platforms do have an incentive to induce type revelation. This can be achieved, for instance, by certification and information disclosure requirements.[29] To be more concrete, consider a monopolist e-commerce platform. Suppose that the platform eliminates any search frictions that otherwise cause demand distortions. If the firm does not use the platform as a mediator, it faces a market with additional search frictions, hence a lower demand. Now, suppose that the platform requires the firm to provide a verifiable public report about its quality in case it wishes to use the platform. If the firm agrees, it provides the report and the platform then incurs some cost of verification. Since

[29] Amazon, for example, takes serious measures when sellers ship products that do not match the information provided on the product detail page. See https://sellercentral.amazon.com/gp/help/external/G202010130 for details.

the firm fully reveals its type, the *ex ante* expected revenue it generates is $\mathbb{E}(v)$ (see the discussion in Section 7 for further details). In the Online Appendix, we provide an upper-bound on the cost of verification that is sufficient to make each firm type prefer to use the platform and disclose its type.

An alternative tool the platform can use to achieve type revelation is to charge a fixed fine $\bar{c} > 0$ to the firm in case of a consumer complaint. If the product has low quality, consumers who purchase the product will end up with a net utility loss in any pooling PBE. Provided that the platform can incentivize the consumers to submit a complaint or enter a negative review revealing their negative experiences with the product, it can ensure the existence of a separating PBE by setting $\bar{c} > v_H - v_L$ so that only the high-type firm sells through the platform and earns v_H whereas the low-type firm stays out and earns v_L. Although if the low-type firm pretends to be a high type by using the platform, consumers will think it offers high quality and be willing to pay v_H, the fine $\bar{c}$ is high enough to deter the low-type firm from doing so.

VI(ii). *Partially-revealing policies*

Online platforms often offer additional services to firms and charge some fixed subscription fees. Amazon Vine Program, and to some extent the Early Reviewer Program of Amazon,[30] provides firms with the option of independent and unbiased product reviews (i.e., $b = 0$). Existence of such a service itself could potentially reveal information as consumers may interpret subscription as a signal of high quality. We argue below that these additional services and fixed fees for these services may reveal information through a partially-separating PBE.

As an example, suppose that the platform charges a fixed fee $\bar{c} > 0$ to the firm if the latter wants to subscribe and open its product to reviews. Consumers can observe if the firm has opted in to use the service or not. This would open up an additional signaling dimension. Now, fix the equilibrium expectations so that if both firm types choose to pay the fee, the expected play is a pooling PBE with a price $\bar{p}$. If $\bar{c}$ is small enough, a pooling PBE exists in which both firm types would pay the fee. However, if the platform does not prefer the firm to play a pooling PBE, it can incentivize the low-type firm to deviate to opting out by setting $\bar{c} > \pi_L(\bar{p}) - v_L$. On the other hand, such a fixed fee would not induce a fully revealing equilibrium. If there was a revealing PBE, the high-type firm would pay the fee and open its product to reviews, while the low-type would opt out, and the consumers would interpret the opting-in behavior as a signal for the high-type. In such a case, however, the low-type firm would deviate from the revealing strategy profile by pretending to be a high-type and thereby earning $v_H - \bar{c}$.

Nevertheless, there might exist a partially-separating PBE in the following sense: The high-

[30] Sources: https://www.amazon.com/gp/vine/help and https://www.amazon.com/gp/help/customer/display.html?nodeId=202094910&tag=bisafetynet2-20, respectively.

type firm always pays the fee and subscribes, whereas the low-type is indifferent between subscribing and not, and so randomizes between these two. If the low-type firm opts out, it reveals itself and charges v_L. If it subscribes, it mimics the high-type. Whenever consumers observe a firm in the platform, they update their prior beliefs according to the mixed strategy of the low-type. Even if the price is fully pooling, the opting-in decision in this case provides partial information about quality, hence potentially increasing the final demand. In the Online Appendix, we provide the equilibrium conditions for such a partially-revealing PBE.

VI(iii). *Raising costs of manipulation*

Another common way of intervening in markets with product review manipulation is running algorithms to detect and punish fake reviews.[31] Such a policy makes it harder for firms to produce fake or promotional reviews. To generate more articulate reviews that are not filtered by algorithms, firms may work with online bloggers or offer higher compensation to consumers who would write promotional reviews, for which the firm has to incur a higher cost for each biased review. For our model, such policies could be represented by an increase in a parameter that shifts the marginal cost of manipulation for both firm types. How would an increase in marginal cost of manipulation affect the *ex ante* expected revenue of the firm. On the one hand, reductions in manipulation levels would make consumers more trusting of their private signals. However, since for almost all prices, manipulation by at least one type is effective in increasing the demand (see Theorem 2), lower manipulation may result in reduced equilibrium demand and therefore reduced revenues. In fact, our numerical solutions presented in the Online Appendix suggest that the net demand effect is negative for both pooling and partially-separating PBE. These results suggest that a revenue-maximizing platform may want to avoid a policy that increases marginal costs of manipulation for both types.

Would a consumer protection agency choose to raise marginal costs of manipulation –using similar tools that the platform has at its disposal– when the types are playing a pooling PBE? The answer depends on which PBE the agency anticipates since, as stated in Corollary 2 and illustrated in Figure 3, the welfare effect of reducing the level of manipulation depends on the price level. For pooling PBE, if $\bar{p} < \mathbb{E}(v)$, the regulator can raise consumer surplus by limiting manipulative effort, while the opposite is true if $\bar{p} > \mathbb{E}(v)$. In contrast to pooling PBE, the net effect of manipulation on consumer surplus in a partially-separating PBE is more complicated. Recall that the L-type firm can always do effective manipulation in a partially-separating PBE (see Proposition 3). This is related to the ability of the L-type to use price-signaling as an

[31] Amazon claims that it constantly analyzes all incoming and existing reviews in its platform via machine learning algorithms to filter out inauthentic reviews. Yet, independent analyses by ReviewMeta indicate that a significant amount of suspicious reviews remain on Amazon's platform. This suggests that there is an ongoing arms race between groups that police fake reviews and those who produce them.

additional manipulation tool. Therefore, biased reviews are potentially more harmful under partially-separating PBE than under pooling PBE. Indeed, our numerical calculations show that, for each fixed price level that is admissible for a partially-separating PBE, the net effect of manipulation on consumer surplus is negative. Moreover, the net harm that manipulation causes increases with the price level.

VI(iv). *Increasing the signal precision*

The dispersion of consumers' private information depends negatively on the precision of the noise component ε in their signals. In principle, a platform or regulator can supplement existing consumer reviews with more precise information about quality. Or it can devise various strategies, such as deleting extreme review signals, to increase the transparency of consumer reviews. Whether policy-makers would implement these policies depend on the net effect of precision on equilibrium outcomes.

Signal precision has two effects on consumer 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. We call this the 'accuracy effect' and, by itself, it should discourage biased reviews. On the other hand, 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 could raise the equilibrium levels of bias by each type.

When the signals are extremely precise, almost all of the consumers receive signals that are very close to the actual value of the product. This makes it extremely hard to shift the demand through manipulation and significantly reduces the incentives for manipulation. We show in the Online Appendix that the equilibrium manipulation levels converge to 0 as precision increases indefinitely. On the other hand, our numerical calculations presented in the Online Appendix illustrate the effects of signal precision on the equilibrium levels of manipulative efforts b_L and b_H, and welfare under two different levels of $\bar{p}$, respectively. An increase in signal precision leads to an increase in the equilibrium bias levels introduced by both types within a particular range of precision. This suggests that the concentration effect of precision can sometimes dominate the accuracy effect. However, as signal precision increases further, the accuracy effect overtakes the concentration effect.[32]

The effect of precision on welfare seems to be monotonic. Higher precision increases *ex*

[32] There is a strategic discontinuity with respect to σ. As long as $\sigma > 0$, it is impossible for consumers to infer the bias levels from the signals even when $v_L + b_L < v_H + b_H$. However, when $\sigma = 0$, the interaction between the firm and the consumers turns into a signaling game. Therefore, it is possible to find a pooling equilibrium in which $v_L + b_L = v_H$ if the pooling price $\bar{p} \geq \max\{\mathbb{E}(v), C(v_H - v_L)\}$.

ante consumer surplus for a wide range of parameters. This implies that a regulator might employ tools to increase signal precision. On the other hand, the effect of precision on *ex ante* revenues seems to depend on the price level. For higher prices, increasing signal precision increases the *ex ante* revenue, while our numerical results suggest an effect in the opposite direction for lower prices. Therefore, the policy choice of a platform may critically depend on the expected equilibrium price level.

VII. EQUILIBRIUM BEHAVIOR WITH COMMITMENT

In this section, we discuss the preferences of a firm from an *ex ante* perspective. First, we establish that if there was a way for the firm to credibly reveal its type, full revelation would be *ex ante* firm optimal. Commitment to a state-contingent pricing scheme would achieve this outcome. That is, if the firm could commit to charging a different price in each state, the price level would reveal its type.

In section VII(ii) below, we assume that such type-revelation is not feasible and instead focus on commitment to a state-contingent manipulation plan. We first illustrate how this type of commitment can in principle eliminate the implicit competition between the types and *ex ante* benefit the firm. Then, we proceed to describe what the equilibrium behavior of the firm looks like when it could commit to a previously announced manipulation schedule within the context of a pooling PBE. We find that, at any arbitrary pooling price, the firm manipulates in at most one of the two quality states, and the *ex ante* profit increases by removing the implicit arms race between the types.

VII(i). *Ex ante commitment to type-revealing prices*

Type revelation is optimal for the firm from an *ex ante* point of view. We have already stated this result in Proposition 4. In the proof of Proposition 4, we also show that the *ex ante* consumer surplus in a pooling (or partially-separating) PBE is always positive. Since the total surplus in such a PBE is smaller than under type-revelation, the *ex ante* profit in any pooling (or partially-separating) PBE must be lower than the *ex ante* profit under type revelation.[33]

VII(ii). *Ex ante commitment to state-contingent manipulation*

Now, suppose that the firm cannot commit to a revealing price strategy but can commit to a state-contingent manipulation plan.[34] We have previously shown in Theorem 1 that both firm

[33]This implies that, in our context, full revelation is always socially beneficial even though it harms consumers. Corts [2014] considers a related but different setup, where he finds that information revelation is not always socially optimal.

[34]It is possible to observe such an asymmetry in commitment abilities within a firm if the marketing department of the firm is separate from other departments.

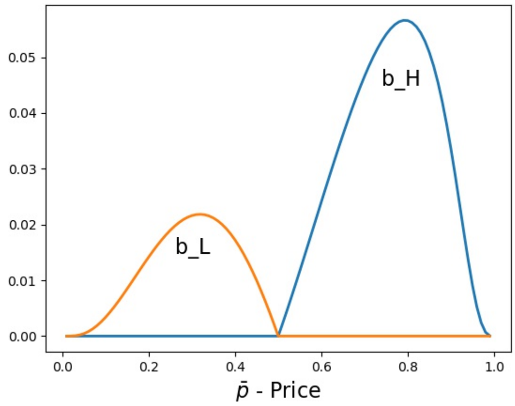

Figure 4: Bias levels with commitment

types exert strictly positive manipulative effort in a pooling PBE. However, from an *ex ante* perspective, this behavior is not optimal for the firm.[35] This is because if the firm reduces both manipulation levels by the same increment, the consumers' signaling threshold would not change. Therefore, the firm would benefit to commit to setting the minimum manipulation to zero, and keeping the other positive. The firm will choose $b_L > 0$ or $b_H > 0$ based on the state in which manipulation is relatively more effective. Given the results in Theorem 2, one would expect the firm to choose $b_L > 0$ only when the expected pooling price is low, and $b_H > 0$ when the price is relatively high. Moreover, when manipulation is ineffective in both states (e.g., when $\bar{p} = \mathbb{E}(v)$), the firm will choose zero manipulation for each state.[36] The numerical solutions for optimal levels of manipulation (for $g = 0.5$) at different pooling price levels are depicted in Figure 4. They suggest that the optimal plan is $(b_L, 0)$ if $\bar{p} < \mathbb{E}(v)$ and $(0, b_H)$ otherwise. These solutions confirm our intuition that the firm would commit to positive manipulation only in the state where marginal benefit of manipulation is relatively higher.

VIII. CONCLUSION

Manipulation in product reviews is commonplace especially in markets where incentives for promotional or fake reviews are high. In this paper, we offer a model of information manipulation where firms can strategically influence the quality expectations of consumers both by inducing an unobservable bias in the information that consumers collect through product review platforms and also by strategically using price-signaling. We show that the ability to increase demand via manipulative effort is not reserved only to one firm type. Depending on the price level, each of the firm types can be effective manipulators.

[35]We prove this result in Proposition 2 in the Online Appendix.
[36]The optimality conditions to this problem are stated in the Online Appendix.

Our paper identifies a novel interaction between price-signaling and information manipulation via signal-jamming. This enables us to provide a more nuanced welfare and policy analysis for regulating firms' behavior on product review platforms. We find that policies that monitor and reveal the type of the firm may not benefit the consumers especially when the firm has monopoly power. The welfare consequences of manipulation in oligopolies remains to be an interesting open question. On the other hand, commonly employed policy interventions such as using algorithms to detect and delete fake reviews will benefit consumers only when prices are low enough such that the low-quality firm type has more incentive to manipulate consumer reviews. Moreover, policies that specifically target the low-type firm (e.g., *ex post* punishment of fake reviews when the quality turns out to be low) would always increase the consumer surplus.

Our model is relevant for understanding promotional reviews for experience goods that are offered by monopolies or sold in markets characterized by monopolistic competition. Best examples of the latter type of goods are books, hotels and restaurants. In these markets, consumers heavily rely on online review platforms to collect information about product attributes and experience of other customers. We show that equilibria exist where the low-quality sellers choose to mimic –fully or partially– the pricing behavior of high-quality sellers. A range of prices can be supported in these equilibria giving way to multiplicity. One can think of this multiplicity as driven by factors other than product qualities offered in a market, and this multiplicity allows us to generate a novel prediction that one can in principle take to data: For products that are viewed as 'overpriced' –given the information consumers possess about the possible quality levels and their prior likelihoods– promotional (i.e., manipulated) reviews should be more common for high-quality products *vis-a-vis* the low-quality products. On the other hand, when the product is 'underpriced', i.e., $\bar{p} < \mathbb{E}(v)$ in the language of our model, the low-quality firms have more incentive to manipulate product reviews, resulting in higher share of fake reviews for low-quality products. A related implication of our model is that the gap between the prevalence of manipulated reviews for high and low-quality products should be smaller in review platforms where it is more costly to post fake or promotional reviews. While our model is too stylized to directly guide empirical tests of these hypotheses, one can in principle measure prior quality expectations of consumers ($\mathbb{E}(v)$) via consumer surveys and compare these to prevailing prices in order to distinguish products that are viewed as over- and under-priced. Furthermore, one can use algorithms –already used by some e-platforms– to detect and quantify fake product reviews posted on these platforms. Finally, average scores or more objective assessment of product attributes can help us classify each product into a quality group. Formulating a model that is tailored to fit the characteristics of a specific market and thereby generating sharper predictions to test would certainly be a valuable contribution one can consider for future research.

In the analysis, we have assumed that all consumers are *ex ante* identical in the sense that their valuations for the seller's product are the same. The ensuing heterogeneity was introduced *ex post* by means of an idiosyncratic quality signal received by each consumer. We could alternatively consider preference heterogeneity at the outset and assume that consumers received a common quality signal. Specifically, suppose that consumer i's payoff is $v_j + \psi_i - p$ when she purchases the product at price p. Here, ψ_i represents the idiosyncratic preference of consumer i for the product. If ψ_i follows a well-behaved distribution over some bounded interval, each consumer would then follow a different purchasing threshold signal defined by their individual ψ_i, and as such, each would have a different likelihood of buying the product at a given price p. As a result, the seller would face a downward-sloping demand curve that depends on the *average* likelihood of purchase in the population. The analysis becomes analytically less tractable as each consumer would follow a different equilibrium strategy. However, the main results regarding the existence and effectiveness of manipulation we have reached in our main analysis would carry over to this alternative setting. Moreover, both low and high-quality sellers can still achieve effective manipulation for some prices, and consumers become better off with manipulation than without whenever $b_H > b_L$. On the other hand, some of the welfare implications could change as fully-revealing prices also create consumption distortions now.

APPENDIX A. PROOFS FOR THE ANALYSIS OF POOLING EQUILIBRIA

Proof of Lemma 1 For notational simplicity, denote $f_j \equiv f(x-v_j-b_j)$ and $f'_j \equiv f'(x-v_j-b_j)$. Then,

$$\frac{\partial \mathbb{E}(v|x)}{\partial x} =$$

$$\frac{\sum_{j\in J} v_j f'_j P(v=v_j)\left(\sum_{j\in J} f_j P(v=v_j)\right) - \sum_{j\in J} f'_j P(v=v_j)\left(\sum_{j\in J} v_j f_j P(v=v_j)\right)}{\left(\sum_{j\in J} f_j P(v=v_j)\right)^2} > 0 \Leftrightarrow$$

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

$$(v_H-v_L)g(1-g)f'_H f_L > (v_H-v_L)g(1-g)f_H f'_L \Leftrightarrow$$

$$\frac{f'(x-b_H-v_H)}{f(x-b_H-v_H)} > \frac{f'(x-v_L-b_L)}{f(x-v_L-b_L)}, \tag{16}$$

which holds since f is log-concave by Assumption 1 and $b_H+v_H > b_L+v_L$ by supposition. ■

Proof of Lemma 2 If $b_L = 0$ in a PBE, then the hypothesis holds trivially because $v_H + b_H > v_L$. Therefore, suppose that $b_L > 0$. Now, suppose on the contrary that $b_L + v_L > b_H + v_H$. Then, it is possible to show that consumers use a single threshold $\bar{x}$; however, they purchase the product if and only if $x_i \leq \bar{x}$. The proof of this follows the same steps as in the proof of Lemma 1. But then the profit for the L-type becomes $\bar{p}F(\bar{x}-b_L-v_L) - C(b_L)$, which is strictly decreasing in b_L for any $\bar{x}$. This implies that the L-type does not have any incentive to do manipulation, which contradicts with $b_L > 0$.

If $b_L + v_L = b_H + v_H$, the posterior of each consumer would be equal to her prior. In this case, the marginal contribution of the L-type's manipulation to the revenue would be zero, which is a contradiction with $b_L > 0$. ■

Lemma 3 *Suppose that Assumptions 1, 2 and 3 hold. The three pooling PBE conditions given in (5) and (7) have an interior solution for each price.*

Proof of Lemma 3 By Corollary 1 and the Implicit Function Theorem, for each bias pair (b_L, b_H), there exists a unique $\bar{x} = \bar{x}(b_L, b_H)$, a continuously differentiable function. Therefore we can reduce the number of equations that define PBE into the following two equations that are very similar to the first-order conditions in (5):

$$\bar{p}f(\bar{x}(b_L,b_H)-v_j-b_j)=C'(b_j) \quad \text{for } j=L,H. \tag{17}$$

It is possible to show, by Brouwer's fixed point theorem that these two equations have a positive solution by defining a function from bias levels to bias levels and using the Assumptions 1, 2 to show continuity of this map in a compact set defined as the Cartesian product of two compact intervals defined between 0 and and a bias level high enough that the firm would never choose a bias above this bound. ■

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

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

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

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

which implies the optimal profit level that the H-type can achieve must be strictly higher than that of the L-type. ■

Proof of Proposition 2 The proof follows from the inspection of the consumer indifference condition (7). When $\bar{p}=\mathbb{E}(v)$, equation (7) implies that marginal revenues from manipulation for both types are equal to each other. Thus, given identical cost functions for manipulation, equilibrium bias levels for both types will be equal to each other. By a similar argument, the marginal revenue of manipulation for H-type is higher than for L-type if and only if $\bar{p}>\mathbb{E}(v)$. ■

Lemma 4 *Suppose that Assumptions 1, 2 and 3 hold. When $\bar{p}$ converges to v_L, the purchasing signal thresholds under manipulation $\bar{x}$ and without manipulation $\underline{x}$ both diverge to $-\infty$. When $\bar{p}$ converges to v_H, $\bar{x}$ and $\underline{x}$ both diverge to ∞.*

Proof of Lemma 4 When $\bar{p}<(1-g)v_H+gv_L$, $b_L>b_H$ and by the first-order conditions in (5)

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

since $f(\cdot)$ is log-concave and log-concavity implies unimodality. Then, by Lemma 2, $\bar{x} < b_H + v_H$.

On the other hand, $b_L < b_H$ when $\bar{p} > (1-g)v_H + gv_L$, and by the first-order conditions in (5)

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

Then, by Lemma 2, $\bar{x} > b_L + v_L$.

The consumer indifference condition (9) can rewritten as follows:

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

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

Lemma 5 *Suppose that Assumptions 1, 2 and 3 hold. The purchasing signal threshold $\bar{x}$ strictly increases with price $\bar{p}$. That is, $d\bar{x}/d\bar{p} > 0$.*

Proof of Lemma 5 To simplify notation, let $f(\bar{x}-b_j-v_j) = f_j$ and $f'(\bar{x}-b_j-v_j) = f'_L$, $j = L, H$. By using the consumer indifference condition (9), define $\Omega(\bar{x},\bar{p})$ as

$$\Omega(\bar{x},\bar{p}) = (1-g)(v_H-\bar{p})f(\bar{x}-v_H-b_H(\bar{x},\bar{p})) - g(\bar{p}-v_L)f(\bar{x}-v_L-b_L(\bar{x},\bar{p})), \tag{18}$$

where $b_H(\bar{x},\bar{p})$ and $b_L(\bar{x},\bar{p})$ are uniquely defined by the first-order conditions in (5). By Assumption 3, $b_H(\bar{x},\bar{p})$ and $b_L(\bar{x},\bar{p})$ are unique for each given $\bar{x}$ and $\bar{p}$. Totally differentiating $\Omega(\bar{x},\bar{p})$ with respect to $\bar{p}$ yields

$$\frac{d\Omega(\bar{x},\bar{p})}{d\bar{p}} = \frac{\partial\Omega(\bar{x},\bar{p})}{\partial\bar{p}} + \frac{\partial\Omega(\bar{x},\bar{p})}{\partial\bar{x}}\frac{d\bar{x}}{d\bar{p}}.$$

We know that as the price changes, $\bar{x}$ adjusts so that the consumer indifference condition continues to hold. This means that $\frac{d\Omega(\bar{x},\bar{p})}{d\bar{p}} = 0$ is always true. But this also implies that if $\frac{d\bar{x}}{d\bar{p}} = 0$, then $\frac{\partial\Omega(\bar{x},\bar{p})}{\partial\bar{x}}$ has to be equal to zero as well. We show below that this is impossible.

Suppose to the contrary that $\frac{d\bar{x}}{d\bar{p}} = 0$. Then, $\frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{x}}$ can be calculated as

$$
\begin{aligned}
\frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{x}} &= g\left((\bar{p}-v_L)f_L'\frac{\partial b_L}{\partial \bar{p}} - f_L\right) - (1-g)\left((v_H-\bar{p})f_H'\frac{\partial b_H \partial \bar{p}}{+} f_H\right) \\
&= g\left((\bar{p}-v_L)\frac{f_L' f_L}{\bar{p}f_L' + C''(b_L)} - f_L\right) - (1-g)\left((v_H-\bar{p})\frac{f_H' f_H}{\bar{p}f_H' + C''(b_H)} + f_H\right) \\
&= g\left(\frac{\bar{p}f_L' f_L - v_L f_L' f_L - \bar{p}f_L' f_L - C''(b_L) f_L}{\bar{p}f_L' + C''(b_L)}\right) \\
&\quad -(1-g)\left(\frac{v_H f_H' f_H - \bar{p}f_H' f_H + \bar{p}f_H' f_H + C''(b_H) f_H}{\bar{p}f_H' + C''(b_H)}\right) \\
&= -g f_L \frac{v_L f_L' + C''(b_L)}{\bar{p}f_L' + C''(b_L)} - (1-g)\frac{v_H f_H' + C''(b_H)}{\bar{p}f_H' + C''(b_H)},
\end{aligned}
$$

where $\frac{\partial b_j}{\partial \bar{p}} = \frac{f_j}{\bar{p}f_j' + C''(b_j)}$ follows from implicit differentiation of the first-order conditions in (5). Note that, in the calculation of $\frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{x}}$, we ignore the indirect effect of $\bar{p}$ on $\bar{x}$.

Now, by Assumption 3, we know that $\bar{p}f_j' + C''(b_j) > 0$ for each j and for all $\bar{p}$ and $\bar{x}$. This implies that $\frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{x}} < 0$ for all $\bar{p}$ and $\bar{x}$, which is a contradiction.

We have established that $d\bar{x}/d\bar{p}$ cannot be 0. But this implies by Lemma 4 that $d\bar{x}/d\bar{p}$ must be globally positive. ■

Proof of Theorem 1 Lemma 3 shows the existence of a triple $(b_L, b_H, \bar{x})$ that solves the equilibrium conditions at each price level $\bar{p} \in (v_L, v_H)$. To show that this triple is unique for each price $\bar{p}$, note that

$$
\frac{d\Omega(\bar{x},\bar{p})}{d\bar{p}} = \frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{p}} + \frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{x}}\frac{d\bar{x}}{d\bar{p}} = 0,
$$

where $\Omega(\bar{x},\bar{p})$ is as defined in equation (18) in the proof of Lemma 5. By Lemma 5 and from the calculations we do in the proof of Lemma 5, we know that $\frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{p}} < 0$ and $\frac{d\bar{x}}{d\bar{p}} > 0$, which implies that $\frac{\partial \Omega(\bar{x},\bar{p})}{\partial \bar{x}} > 0$. But, this means that there exists a unique $\bar{x}$ that solves the consumer indifference condition $\Omega(\bar{x},\bar{p}) = 0$ given in (9).

This does not guarantee that any triple $(b_L, b_H, \bar{x})$ indeed forms a pooling PBE for any price $\bar{p}$. To verify the existence of a pooling PBE at a price $\bar{p}$, we need to check if the deviation profit of a given firm type is indeed lower than the profit implied by the candidate PBE. The deviation payoff for both types is v_L since the off-equilibrium beliefs of consumers imply that a deviator must be of low type. Thus the only profitable deviation from equilibrium is to charge v_L and achieve full demand. Since in any equilibrium $\pi_H > \pi_L$, any deviation that is not profitable for the L-type will also be unprofitable for the H-type. To see that the profit for the L-type is greater than v_L for sufficiently low prices, consider the derivative of the profit function of the

L-type with respect to price:

$$\frac{d\pi_L}{d\bar{p}} = 1 - F(\bar{x} - b_L - v_L) - C'(b_L)\frac{db_L}{d\bar{p}} - \bar{p}f(\bar{x} - b_L - v_L)\left(\frac{d\bar{x}}{d\bar{p}} - \frac{db_L}{d\bar{p}}\right). \tag{19}$$

We can substitute the last term by implicitly differentiating the first-order condition for the L-type as follows:

$$f(\bar{x} - b_L - v_L) + \bar{p}f'(\bar{x} - b_L - v_L)\left(\frac{d\bar{x}}{d\bar{p}} - \frac{db_L}{d\bar{p}}\right) = C''(b_L)\frac{db_L}{d\bar{p}} \Rightarrow$$
$$\bar{p}\left(\frac{d\bar{x}}{d\bar{p}} - \frac{db_L}{d\bar{p}}\right) = \frac{1}{f'(\bar{x} - b_L - v_L)}C''(b_L)\frac{db_L}{d\bar{p}} - \frac{f(\bar{x} - b_L - v_L)}{f'(\bar{x} - b_L - v_L)}.$$

Then, we can rewrite the derivative in (19) as

$$\begin{aligned}\frac{d\pi_L}{d\bar{p}} =& 1 - F(\bar{x} - b_L - v_L) - C'(b_L)\frac{db_L}{d\bar{p}} \\ & - \frac{f(\bar{x} - b_L - v_L)}{f'(\bar{x} - b_L - v_L)}C''(b_L)\frac{db_L}{d\bar{p}} + \frac{[f(\bar{x} - b_L - v_L)]^2}{f'(\bar{x} - b_L - v_L)}.\end{aligned} \tag{20}$$

As we take the limit of $\frac{d\pi_L}{d\bar{p}}$ as $\bar{p} \to v_L$, we have $\bar{x} \to -\infty$ and $b_L \to 0$. As a result, $1 - F(\bar{x} - b_L - v_L) - C'(b_L)\frac{db_L}{d\bar{p}}$ converges to 1. The remaining two terms in (20) converge to 0 since $\lim_{x \to -\infty} \frac{f'(x)}{f(x)} = \infty$ while $C''(b_L)$ and $\frac{db_L}{d\bar{p}}$ remain finite in the limit. To see why $\frac{db_L}{d\bar{p}}$ is always finite, note that for any p and $\delta > 0$,

$$\frac{|C(b_L(p+\delta)) - C(b_L(p))|}{\delta} \leq$$
$$|(f(\bar{x}(p+\delta) - b_L(p+\delta) - v_L) - f(\bar{x}(p) - b_L(p) - v_L)| \leq f(0).$$

This is because even for prices very close to v_L, that is when $p \to v_L$, $\bar{x}$ diverges to $-\infty$ and therefore $f(\bar{x}(p) - b_L(p) - v_L)$ converges to 0. Moreover, since f is unimodal and symmetric around 0, $f(0)$ is the (finite) upper-bound for $f(\bar{x}(p) - b_L(p) - v_L)$ at any price. Then, since the cost function is monotonic, a uniform finite limit on $\frac{|C(b_L(p+\delta)) - C(b_L(p))|}{\delta}$ also implies a uniform limit on $\frac{|b_L(p+\delta) - b_L(p)|}{\delta}$, hence the derivative.

Thus, we have established that

$$\lim_{\bar{p} \to v_L} \frac{d\pi_L}{d\bar{p}} = 1 > 0. \tag{21}$$

Given that the L-type's profit from deviating to $p_L = v_L$ is equal to v_L, (21) implies that prices sufficiently close to v_L should deliver the L-type a profit above v_L.

Finally, suppose that $v_L > 0$. The profit of the L-type firm for any pooling strategy profile is lower than $\bar{p}(1-F(\bar{x}-b_L-v_H))$, which converges to 0 as $\bar{p}\to v_H$ by Lemma 4. This implies that the L-type firm's profit for a pooling strategy profile would be lower than v_L, which violates the incentive compatibility condition for a pooling equilibrium. ■

Lemma 6 *Suppose that Assumptions 1, 2 and 3 hold. $v_L \geq v_H/2$ implies that no price $\bar{p}\in[p_L, v_H]$ can be supported as a pooling PBE outcome, where $p_L < \mathbb{E}(v)$ is such that $\bar{x}=b_L+v_L$.*

Proof of Lemma 6 By definition of p_L, for any price $\bar{p}\geq p_L$, $\bar{x}\geq b_L+v_L$. Therefore, L-type's demand in a pooling PBE with price $\bar{p}\geq p_L$ is given by $1-F(\bar{x}-b_L-v_L)\leq 0.5$. Thus, the profit for the L-type $\bar{p}(1-F(\bar{x}-b_L-v_L))-C(b_L)$ is less than $v_H/2$. When $v_L\geq v_H/2$, the L-type has an incentive to deviate from any pooling PBE that supports any price $\bar{p}\geq p_L$ to setting a price equal to v_L. ■

Proof of Theorem 2 We will start with L-type's effective manipulation. L-type does effective manipulation if and only if $b_L > \bar{x}-\underline{x}$, which is equivalent to $\underline{x} > \bar{x}-b_L$. By Corollary 1, this is equivalent to $\mathbb{E}(v|\bar{x}-b_L) < \bar{p}$ if manipulation is restricted to be 0. That is, when there is no manipulation, a consumer who receives a signal that is equal to $\bar{x}-b_L$, should expect that the quality is lower than $\bar{p}$. The posterior expectation of such a consumer is calculated as

$$\begin{gathered}\frac{(1-g)v_H f(\bar{x}-b_L-v_H)+g v_L f(\bar{x}-b_L-v_L)}{(1-g)f(\bar{x}-b_L-v_H)+g f(\bar{x}-b_L-v_L)} < \bar{p}.\\ (1-g)(v_H-\bar{p})f(\bar{x}-b_L-v_H)-g(\bar{p}-v_L)f(\bar{x}-b_L-v_L)<0\\ =(1-g)(v_H-\bar{p})f(\bar{x}-b_H-v_H)-g(\bar{p}-v_L)f(\bar{x}-b_L-v_L)\Leftrightarrow\\ f(\bar{x}-b_L-v_H)<f(\bar{x}-b_H-v_H)\Leftrightarrow|\bar{x}-b_L-v_H|>|\bar{x}-b_H-v_H|.\end{gathered}$$

In sum,

$$b_L > \bar{x}-\underline{x}\Leftrightarrow|\bar{x}-b_L-v_H|>|\bar{x}-b_H-v_H|. \tag{22}$$

We will show below that this equivalence condition for L-type's effective manipulation holds when $\bar{p}<(1-g)v_H+g v_L$ or when $\bar{p}>(1-g)v_H+g v_L$ and $\bar{x}>(b_L+b_H)/2+v_H$.

Now, when $\bar{p}<(1-g)v_H+g v_L$, $b_L > b_H$ by Proposition 2. Then, by first-order conditions (5),

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

since $f(\cdot)$ is unimodal. Then, by Lemma 2, the condition above is equivalent to

$$\begin{gathered}\bar{x}<b_H+v_H\\ \bar{x}-b_H-v_H<\bar{x}-b_L-v_L<b_H+v_H-\bar{x}\Rightarrow\\ \bar{x}<\frac{b_L+b_H}{2}+\frac{v_L+v_H}{2}<\frac{b_L+b_H}{2}+v_H.\end{gathered}$$

On the other hand, when $b_L > b_H$, condition (22) is equivalent to

$$\begin{gathered}\bar{x} < b_L + v_H\\ \bar{x} - b_L - v_H < \bar{x} - b_H - v_H < b_L + v_H - \bar{x} \Leftrightarrow\\ \bar{x} < \frac{b_L + b_H}{2} + v_H,\end{gathered}$$

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 2, condition (22) becomes equivalent to

$$\begin{gathered}\bar{x} > b_L + v_H\\ b_L + v_H - \bar{x} < \bar{x} - b_H - v_H < \bar{x} - b_L - v_H \Leftrightarrow\\ \bar{x} > \frac{b_L + b_H}{2} + v_H.\end{gathered}$$

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

$$\begin{gathered}b_H > \bar{x} - \underline{x} \Leftrightarrow \underline{x} > \bar{x} - b_H\\ (v_H - \bar{p})(1-g)f(\bar{x} - b_H - v_H) - (\bar{p} - v_L)gf(\bar{x} - b_H - v_L) < 0\\ = (v_H - \bar{p})(1-g)f(\bar{x} - b_H - v_H) - (\bar{p} - v_L)gf(\bar{x} - b_L - v_L) \Leftrightarrow\\ f(\bar{x} - b_L - v_L) < f(\bar{x} - b_H - v_L) \Leftrightarrow |\bar{x} - b_L - v_L| > |\bar{x} - b_H - v_L|.\end{gathered}$$

In sum,

$$b_H > \bar{x} - \underline{x} \Leftrightarrow |\bar{x} - b_L - v_L| > |\bar{x} - b_H - v_L|. \tag{23}$$

When $\bar{p} < (1-g)v_H + gv_L$, $b_H < b_L$, therefore condition (23) is equivalent to

$$\begin{gathered}\bar{x} < b_L + v_L\\ \bar{x} - b_L - v_L < \bar{x} - b_H - v_L < b_L + v_L - \bar{x} \Leftrightarrow\\ \bar{x} < v_L + \frac{b_L + b_H}{2}.\end{gathered}$$

When $\bar{p} > (1-g)v_H + gv_L$, $b_H > b_L$. Therefore condition (23) is equivalent to

$$\bar{x} > b_L + v_L$$
$$b_L + v_L - \bar{x} < \bar{x} - b_H - v_L < \bar{x} - b_L - v_L \Leftrightarrow$$
$$\bar{x} > v_L + \frac{b_L + b_H}{2},$$

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 4 and 5. To show that $p_L < (1-g)v_H + gv_L$, note that

$$\bar{x} = v_L + \frac{b_L + b_H}{2} \Rightarrow$$
$$\bar{x} - b_L - v_L = \frac{b_H - b_L}{2} \quad \& \quad \bar{x} - b_H - v_H = -(v_H - v_L) + \frac{b_L - b_H}{2}$$
$$\Rightarrow |\bar{x} - b_H - v_H| > |\bar{x} - b_L - v_L| \Rightarrow f(\bar{x} - b_H - v_H) < f(\bar{x} - b_L - v_L)$$
$$\Rightarrow b_H < b_L,$$

by the first-order conditions and the fact that the noise distribution is unimodal and symmetric around 0. Then, by Proposition 2, any price level p_L which induces a purchase threshold $\bar{x} = v_L + \frac{b_L+b_H}{2}$ must satisfy $p_L < (1-g)v_H + gv_L$.

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 $d\bar{x}/d\bar{p} > 0$, $\bar{x}_2 > \bar{x}_1$, and therefore

$$b_{L1} + b_{H1} < b_{L2} + b_{H2}.$$

Since $f(\cdot)$ is unimodal and $\bar{x} = v_L + \frac{b_L+b_H}{2} < b_L + v_L$ by the fact that $b_L > b_H$,

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

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 4 First, we show that *ex ante* consumer surplus is always greater than or equal to zero in any pooling or partially separating PBE. Note that for any pooling PBE that supports the price $\bar{p}$

$$E(CS_{\text{pool}}) = g(v_L - \bar{p})(1 - F_L) + (1 - g)(v_H - \bar{p})(1 - F_H) \geq 0 \Leftrightarrow$$
$$\frac{1 - F_L}{1 - F_H} \leq \frac{(1 - g)(v_H - \bar{p})}{g(\bar{p} - v_L)} = \frac{f_L}{f_H} \Leftrightarrow$$
$$\frac{f_L}{1 - F_L} \geq \frac{f_H}{1 - F_H},$$

where, the equality in the second condition above follows from consumer indifference condition for the pooling PBE. 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 last inequality as desired. The proof for partially-separating PBE is similar. The proof of the second part of the proposition follows from the fact that there is some demand distortion due to asymmetric information in any pooling or partially-separating PBE, while there is none if the firm can commit to a revealing strategy and so can charge v_j for each type j to extract all of the consumer surplus. Since consumer surplus is positive for any pooling or partially-separating PBE, the hypothesis follows. ■

APPENDIX B. PARTIALLY-SEPARATING EQUILIBRIA

Definition 1 (Perfect Bayesian Equilibrium) *A strategy profile consists of the bias level chosen by each type of the firm* (b_L, b_H)*, a possibly mixed pricing decision* (β_L, β_H) *and the purchasing decision* $s(x, p)$ *by 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*

$$\forall p \in supp(\beta_L^*) \quad b_L^*(p) \in \underset{b_L\in[0,\infty)}{\operatorname{argmax}}\, pS(b_L, b_H^*, \beta_L^*, \beta_H^*, p, v_L) - C(b_L)$$
$$\forall p \in supp(\beta_H^*) \quad b_H^*(p) \in \underset{b_H\in[0,\infty)}{\operatorname{argmax}}\, pS(b_L^*, b_H, \beta_L^*, \beta_H^*, p, v_H) - C(b_H)$$

$$\beta_L^* \in \underset{\beta_L\in\Delta([v_L,v_H])}{\operatorname{argmax}} \int_{p\in supp(\beta_L)} pS(b_L^*, b_H^*, \beta_L, \beta_H^*, p, v_L)\beta_L(p)dp - C(b_L^*)$$
$$\beta_H^* \in \underset{\beta_H\in\Delta([v_L,v_H])}{\operatorname{argmax}} \int_{p\in supp(\beta_H)} pS(b_L^*, b_H^*, \beta_L^*, \beta_H, p, v_H)\beta_H(p)dp - C(b_H^*),$$

where the strategy and beliefs of consumers are defined as

$$S(b_L^*, b_H^*, \beta_L^*, \beta_H^*, p, v_j) = \int_{-\infty}^{\infty} s^*(x, p) f(x - v_j - b_j^*)dx$$
$$s^*(x,p) = \begin{cases} 1 & if \quad \sum_j (v_j - p)\mu(v_j|x,p) \geq 0 \\ 0 & otherwise, \end{cases}$$

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

$$\mu(v_j|x,p) = \frac{Pr(x|v = v_j, p)Pr(v = v_j)}{gPr(x|v = v_L, p) + (1-g)Pr(x|v = v_H, p)},$$

$Pr(v = v_j)$ *is the prior probability that the firm is of type* j*, and* $Pr(x|v = v_j, p)$ *is the probability that a consumer receives private signal* x *given that the product quality is* v_j *and price is* p*.*

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.[37]

[37] This result would not change if H-type had a higher fixed marginal cost of production. To see this note that the only source of heterogeneity among consumers is the private information. Therefore, in any separating PBE

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 that $v_L \leq p_L < p_H \leq v_H$. 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 pure strategy separating PBE with $p_H > v_H$ would be strict in the sense that the firm would make zero profit in equilibrium as there are no sales. ■

Proof of Theorem 3 The proof consists of two steps. In the first step, we will argue that equations (13) and (14) describe a partially-separating PBE. Then in the second step, we will prove that equations (13) and (14) 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 (13), 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 private 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 in (5).

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

$$\frac{v_H(1-g)f(x_i - v_H - b_H) + v_L\bar{\alpha}gf(x_i - v_L - b_L)}{(1-g)f(x_i - v_H - b_H) + \bar{\alpha}gf(x_i - v_L - b_L)}.$$

Consumers are indifferent when they receive the threshold signal $\bar{x}$. After some rearrangement, the indifference condition can be written as in equation (14).

To establish the existence, first recall that the equilibrium conditions (17) for pooling PBE induce a solution $(b_L, b_H, \bar{x})$ for any $\bar{p} \in (v_L, v_H)$ by Lemma 3. Now, suppose that the low-type firm's corresponding profit is smaller than v_L; that is,

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

so that a pure-strategy pooling PBE does not exist for the price $\bar{p}$. It is possible to show the

the demand would be rectangular. Therefore, even if the high type has a different and higher variable cost of production, there would not be any incentive-compatible way of discouraging the low type from mimicking the behavior of the high type.

existence of a solution $(b_L, b_H, \bar{x})$ to the equilibrium conditions (13) and (14) for each α and $\bar{p}$ and each of the variables of the solution is continuously differentiable in α by Implicit Function Theorem. The low-type firm's corresponding profit is

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

When $\alpha = 1$, these two set of equilibrium conditions coincide. On the other hand, as $\alpha \to 0$ the posterior expected value

$$\frac{v_H(1-g)f(x_i-v_H-b_H)+v_L\alpha g f(x_i-v_L-b_L)}{(1-g)f(x_i-v_H-b_H)+\alpha g f(x_i-v_L-b_L)},$$

converges v_H. Since b_L and b_H can be uniformly bounded by Assumption 3 on the cost function, this implies that $\bar{x}$ diverges to $-\infty$. This implies that the corresponding profit of the low-type firm converges to $\bar{p} > v_L$. Since the solution $(b_L, b_H, \bar{x})$ is continuous in α, there exists $\bar{\alpha} \in (0,1)$ such that

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

■

Proof of Proposition 3 Equation (14) implies that when $\bar{p} = \hat{p}_{\bar{\alpha}}$,

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

which implies that $b_L = b_H$ by the first-order conditions (5). Moreover, by a comparison, $b_L > b_H$ if and only if $\bar{p} < \hat{p}_{\bar{\alpha}}$.

To show that the L-type always effectively manipulates, compare the indifference condition (13) to the corresponding one in the no-manipulation benchmark. Since

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

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

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

which completes the proof. ■

Figure 5 illustrates the relationship between manipulation levels and the price level across partially-separating PBE.

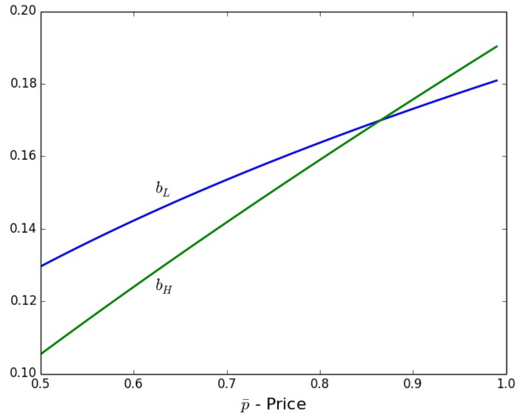

Figure 5: Bias levels as price increases

Proposition 6 *Consider the set of partially-separating PBE indexed by prices $\bar{p}$ over the interval $(\tilde{p}, v_H)$, 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 $\partial\pi_H/\partial\bar{p} > 0$. Moreover, as $\bar{p} \to v_H$, $\alpha \to 0$, and the limit of $\bar{x}$ is finite.*

Proof of Proposition 6 Implicitly differentiating the incentive compatibility condition for the L-type, equation (13) 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 (5) yields

$$\frac{\partial\bar{x}}{\partial\bar{p}} = \frac{1 - F(\bar{x} - b_L - v_L)}{\bar{p}f(\bar{x} - b_L - v_L)} > 0. \tag{24}$$

The behavior of b_L and π_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., $f'F - f^2 < 0 \Leftrightarrow f^2 > f'F$. 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 (5) yields

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

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 (5) becomes

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

which always holds by Lemma 2.

Finally, the last observation is a straightforward corollary to the final argument in the proof of Theorem 3. ■

REFERENCES

Aköz, K. K. and Arbatlı , C. E., 2016, 'Information Manipulation in Election Campaigns,' *Economics & Politics*, 28(2), pp. 181–215.

An, M. Y., 1998, 'Logconcavity versus Logconvexity: A Complete Characterization,' *Journal of Economic Theory*, 80(2), pp. 350–369.

Anderson, E. T. and Simester, D. I., 2014, 'Reviews Without a Purchase: Low Ratings, Loyal Customers, and Deception,' *Journal of Marketing Research*, 51(3), pp. 249–269.

Anderson, S. P. and Renault, R., 2006, 'Advertising Content,' *American Economic Review*, 96(1), pp. 93–113.

Bagwell, K. and Riordan, M. H., 1991, 'High and Declining Prices Signal Product Quality,' *American Economic Review*, 81(1), pp. 224–239.

Caselli, F.; Cunningham, T.; Morelli, M. and Barreda, I. M., 2014, 'The Incumbency Effects of Signalling,' *Economica*, 81(323), pp. 397–418.

Corts, K. S., 2013, 'Prohibitions on False and Unsubstantiated Claims: Inducing the Acquisition and Revelation of Information through Competition Policy,' *Journal of Law and Economics*, 56(2), pp. 453–486.

Corts, K. S., 2014, 'Finite Optimal Penalties for False Advertising,' *Journal of Industrial Economics*, 62(4), pp. 661–681.

Corts, K. S., 2014, 'The Social Value of Information on Product Quality,' *Economics Letters*. 122(2), pp. 140–143.

Dellarocas, C., 2006, 'Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms,' *Management Science*, 52(10), pp. 1577–1593.

Drugov, M. and Troya-Martinez, M., 2019, 'Vague Lies and Lax Standards of Proof: On the Law and Economics of Advice,' *Journal of Economics & Management Strategy*, 28(2), pp. 298–315.

Edmond, C., 2013, 'Information Manipulation, Coordination, and Regime Change,' *The Review of Economic Studies*, 80(4), pp. 1422–1458.

Finucane, T. E. and Boult, C. E., 2004, 'Association of Funding and Findings of Pharmaceutical Research at a Meeting of a Medical Professional Society,' *The American Journal of Medicine*, 11(1), pp. 842–845.

Fudenberg, D. and Tirole, J. A, 1986, 'A "Signal-Jamming" Theory of Predation,' *The RAND Journal of Economics*, 17(3), pp. 366–376.

Fudenberg, D. and Tirole, J., 1991, 'Perfect Bayesian Equilibrium and Sequential Equilibrium,' *Journal of Economic Theory*, 53(2), pp. 236–260.

Gardete, P. M., 2013, 'Cheap-talk Advertising and Misrepresentation in Vertically Differentiated Markets,' *Marketing Science*, 32(4), pp. 609–621.

Grunewald, A. and Kräkel, M., 2017, 'Advertising as Signal Jamming,' *International Journal of Industrial Organization*, 55, pp. 91–113.

Hattori, K. and Higashida, K., 2012, 'Misleading Advertising in Duopoly,' *Canadian Journal of Economics / Revue canadienne d'Economique*, 45(3), pp. 1154–1187.

Holmström, B., 1999, 'Managerial Incentive Problems: A Dynamic Perspective,' *The Review of Economic Studies*, 66(1), pp. 169–182.

Hu, N.; Bose I.; Gao, Y. and Liu, L., 2011, 'Manipulation in Digital Word-of-Mouth: A Reality Check for Book Reviews,' *Decision Support Systems*, 50(3), pp. 627–635.

Janssen, M. C. and Roy, S., 2017, 'Regulating False Disclosure,' See Vienna Economics Papers No. 1705 at http://homepage.univie.ac.at/Papers.Econ/RePEc/vie/viennp/vie1705.pdf (University of Vienna, Department of Economics).

Johnson, J. P. and Myatt, D. P., 2006, 'On the Simple Economics of Advertising, Marketing, and Product Design,' *American Economic Review*, 96(3), pp. 756–784.

Judd, K. L. and Riordan, M. H., 1994, 'Price and Quality in a New Product Monopoly,' *The Review of Economic Studies*, 61(4), 773–789.

Kamenica, E. and Gentzkow, M., 2011, 'Bayesian Persuasion,' *American Economic Review*, 101(6), 2590–2615.

Kartik, N., 2009, 'Strategic Communication with Lying Costs,' *The Review of Economic Studies*, 76(4), pp. 1359–1395.

Matthews, S. A. and Mirman, L. J., 1983, 'Equilibrium Limit Pricing: The Effects of Private Information and Stochastic Demand,' *Econometrica*, 51(4), pp. 981–996.

Mayzlin, D., 2006, 'Promotional Chat on the Internet,' *Marketing Science*, 25(2), pp. 155–163.

Mayzlin, D.; Dover, Y. and Chevalier, J., 2014, 'Promotional Reviews: An Empirical Investigation of Online Review Manipulation,' *American Economic Review*, 104(8), pp. 2421–2455.

Miklós-Thal, J. and Zhang, J., 2013, '(De)marketing to Manage Consumer Quality Inferences,' *Journal of Marketing Research*, 50(1), pp. 55–69.

Milgrom, P. and Roberts, J., 1986, 'Price and advertising signals of product quality,' *Journal of Political Economy*, 94(4), pp. 796–821.

Nelson, P., 1970, 'Information and Consumer Behavior,' *Journal of Political Economy*, 78(2), pp. 311–329.

Nelson, P., 1974, 'Advertising as Information,' *Journal of Political Economy*, 82(4), pp. 729–754.

Piccolo, S.; Tedeschi, P. and Ursino, G., 2015, 'How Limiting Deceptive Practices Harms Consumers,' *The RAND Journal of Economics*, 46(3), pp. 611–624.

Piccolo, S.; Tedeschi, P. and Ursino, G., 2017, 'Deceptive Advertising with Rational Buyers,' *Management Science*, 64(3), pp. 1291–1310.

Renault, R., 2016, 'Chapter 4 - Advertising in Markets,' In *Handbook of Media Economics*, Simon P. Anderson; Joel Waldfogel and David Strömberg, Eds., 1, pp. 121–204.

Rhodes, A. and Wilson, C. M., 2018, 'False Advertising,' *The RAND Journal of Economics*, 49(2), pp. 348–369.

Sismondo, S., 2008, 'Pharmaceutical Company Funding and Its Consequences: A Qualitative Systematic Review,' *Contemporary Clinical Trials*, 29(2), pp. 109–113.