Bayesian Social Learning from Consumer Reviews

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Download date: 15 May. 2024

Operations Research

Bayesian Social Learning from Consumer Reviews

Bar Ifrach, Costis Maglaras, Marco Scarsini, Anna Zseleva

Please scroll down for article—it is on subsequent pages

http://pubsonline.informs.org/journal/opre

Crosscutting Areas

Bayesian Social Learning from Consumer Reviews

Bar Ifrach,[a] Costis Maglaras,[b] Marco Scarsini,[c] Anna Zseleva[d]

[a] Uber Technologies, San Francisco, California 94103; [b] Columbia Business School, Columbia University, New York, New York 10027;
[c] Dipartimento di Economia e Finanza, LUISS, 00196 Roma, Italy; [d] School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Contact: barifrach@gmail.com, https://orcid.org/0000-0003-1897-4631 (BI); c.maglaras@gsb.columbia.edu,
https://orcid.org/0000-0002-4283-2177 (CM); marco.scarsini@luiss.it, https://orcid.org/0000-0001-6473-794X (MS);
zseleva.anna@gmail.com, https://orcid.org/0000-0001-5530-1183 (AZ)

Abstract. Motivated by the proliferation of user-generated product-review information and its widespread use, this note studies a market where consumers are heterogeneous in terms of their willingness to pay for a new product. Each consumer observes the binary reviews (like or dislike) of consumers who purchased the product in the past and uses Bayesian updating to infer the product quality. We show that the learning process is successful as long as the price is not prohibitive, and therefore at least some consumers, with sufficiently high idiosyncratic willingness to pay, will purchase the product irrespective of their posterior quality estimate. We examine some structural properties of the dynamics of the posterior beliefs. Finally, we study the seller's pricing problem, and we show that, if the set of possible prices is finite, then a stationary optimal pricing policy exists. If it costs the seller a constant amount for each additional unit sold, then under the optimal policy learning fails with positive probability.

Keywords: social learning • Bayesian update • reviews • optimal pricing

1. Introduction

Online review sites are playing an increasingly large role in consumers' purchasing decisions. The hotel and hospitality industry has been transformed through the availability and impact of reviews on travelers' decisions.[1] Similarly, other industries, such as online retail, motion pictures, and restaurants, have seen consumers' decisions increasingly influenced by reviews. The proliferation of smartphones is making access to such review sites easier than ever. This note studies the problem of Bayesian social learning from online reviews.

Specifically, a monopolist seller introduces a new product with unknown quality to a market of heterogenous consumers who have access to reviews generated by consumers that purchased the product in the past, and where (a) consumers use Bayesian updating to infer from past reviews the product quality; (b) the mechanism by which consumers report their reviews resembles that of online review sites, albeit in a simplified way; and (c) consumers have heterogeneous preferences (willingness to pay) for the product. The new product (or service) features observed attributes (e.g., location) and unobserved attributes that we denote with the term *quality*, which,

to facilitate the Bayesian inference, can either be high (H) or low (L). The quality experienced by a consumer who purchases the product is equal to the true quality of the product plus some random perturbation (e.g., due to variability in the service delivery process).

Consumers differ in their preferences for the observed product attributes, which is akin to having a heterogenous willingness to pay for a base good of low quality. Consumers arrive sequentially over time and make a Bayesian inference about the product quality based on the information available in the market, as described at the end of this paragraph. Following that, consumers make a once-and-for-all decision of whether to purchase or to forgo the product, depending on their posterior quality distribution and on their idiosyncratic preference for the observed attributes that jointly determine their willingness to pay. Specifically, if a consumer perceives that the expected quality is $\hat{Q}$, her idiosyncratic preference for the other attributes is Θ, and the price is p, then the consumer purchases the product if her expected net utility is nonnegative (i.e., if $\Theta + \hat{Q} - p \geq 0$); we assume that the no-purchase option produces zero utility. The heterogeneity in preferences is captured by consumers'

types that are private information. Buyers submit a binary review that takes the form of a "like" if their ex-post net utility was nonnegative, and a "dislike" if their ex-post net utility was negative. The review is based on the true quality of the product plus the effect of the random perturbation experienced upon consumption of the product. It is only partially informative owing to the heterogeneity in preferences that remains unseen and the fluctuations in the experienced quality of the product. Each agent observes the ordered sequence of consumer reviews.

In the first part of our paper we assume that the price is static and nonprohibitive, in the sense that even if consumers perceived the quality to be low, there are still some consumers with sufficiently high type that will choose to purchase and write a review. Proposition 1 shows that the conditional beliefs of the above-mentioned Bayesian learning process converge to a point mass distribution on the true state of the world (i.e., eventually consumers learn the true quality of the product almost surely). No-purchase decisions do not contribute any information to the learning process; therefore, the fact that they are not observed has no influence on the decisions of the following consumers.

Contrasting to the literature on social learning and, specifically, Banerjee (1992) and Bikhchandani et al. (1992), in our model, before buying the product, consumers have no private information (signals) on the state of the world and are heterogeneous in their preferences; see also Chamley (2003). Information does not aggregate by observing the sequence of purchasing actions and making inferences about past beliefs and signals, but rather information accumulates through the effect of online reviews. In common with these papers, consumers are Bayesian and the unknown quality parameter takes on two possible values, namely H or L. Our assumption on the price and type distribution ensures that some consumers will always purchase and generate new reviews. This ultimately drives the learning result. Our assumption is analogous to the one introduced in Smith and Sørensen (2000), who show that asymptotic learning holds if agents' signals have unbounded strength (i.e., may try the product irrespective of the observed actions of all predecessors). In our model heterogeneous willingness to pay drives experimentation, which in turn drives learning, whereas in the model with signals a sufficient number of consumers have a sufficiently accurate quality information so as to always follow their own signal in their decision. Moreover, in the model of signals, learning implies that all consumers make the same decision. This is not true in our model, where after the market has learned, there will be a set of consumers that choose to

purchase and a set of consumers that will choose not to purchase, and even among the purchasers there will be some fraction that will ultimately dislike the product owing to the quality noise when they experience the good. One essential novelty of the paper is that only agents whose preference is above a threshold take an action, and thus future agents need to filter out this bias.

Focusing on the learning trajectory, Proposition 2 shows that the posterior after the review sequence (like, dislike) will be greater than the posterior after the review sequence (dislike, like), highlighting that the order of reviews affects the dynamics of consumers' beliefs and the importance of early reviews on the product's demand. As a corollary, one can show that the likelihood that the next review will be positive is decreasing with the belief, or in other words that reviews tend to be negative following high quality expectation.

Most papers on Bayesian social learning assume a structure where consumers are endowed with private signals on the unknown quality of the product. Acemoglu et al. (2011) consider agents who are embedded on a general social network and find conditions on the social network and signal structure for asymptotic learning. Herrera and Hörner (2013) consider a model where, as in this paper, agents make a binary choice—"buy" and "not buy"—but only the "buy" decision can be observed by predecessors. They show that asymptotic learning occurs when signals have unbounded strength and agents observe the time elapsed since the product launch. In the model considered by Acemoglu et al. (2014), agents can collect information by forming costly communication links or by delaying their irreversible action. Goeree et al. (2006) consider a model where agents make choices sequentially and their payoff depends not only on the state of the world and their action but also on an idiosyncratic privately observed shock.

Social learning with Bayesian consumers who read reviews before purchasing has been considered by Besbes and Scarsini (2018) in a setting where the reviews are not binary. A recent paper by Acemoglu et al. (2017) shows that, for a fixed price, under some conditions, learning stops with positive probability. These authors study the speed of convergence of learning, when it is achieved.

Next, we explore the effect of social learning from reviews on the seller's pricing decisions. We assume that there is a fixed cost c per unit sold in the market, so, specifically, pricing at $p > c$ may lead to profitable sales for the seller, whereas pricing at $p < c$ results in a loss if the respective units sold at such prices. We allow the seller to use state-dependent pricing and consider the setting where the seller is herself

uninformed about the true quality of the product and learns from the reviews posted by consumers. Under the latter assumption, which is common in the related literature, pricing has no signaling effect on the true product quality.[2] Theorem 1 shows that ex-ante the seller benefits from social learning in the sense that her ex-ante expected revenue increases if consumers engage in social learning; however, ex-post (i.e., after the true quality has been revealed), the seller is better off if the quality is high, and worse off otherwise. Bose et al. (2006) have argued that social learning benefits the seller in a setting with signals, and the above result justifies their conclusion in a setting with reviews and heterogenous preferences.

Moreover, we show that the unit production cost c plays an important role in the optimal dynamic pricing policy and on whether consumers ultimately learn the true quality of the product. If c is sufficiently small so that the seller can profitably participate in the market even if the product is of low quality, then the seller does not exit the market and learning is achieved under the optimal dynamic pricing. On the other hand, if the production cost c is sufficiently high, specifically such that when the posterior belief is low consumers will not purchase the product unless it is priced below cost, then, first, the seller may indeed choose to price below cost for a short period to incentivize sales and the possibility of reviews; and, second, the seller exits the market when the posterior belief falls below a specified threshold.

A more complete treatment of the dynamic pricing problem in this model with Bayesian learning seems intractable. A natural next step would be to appeal to asymptotic analysis of the underlying stochastic system (e.g., using a mean-field approximation that can be justified under a functional strong law of large numbers in large-scale settings). This approach is pursued in Crapis et al. (2017) in a model where consumers learn from online reviews but use a naïve mechanism to interpret the online reviews.

Even if characterizing the optimal pricing policy is not feasible, we can nevertheless say something about the interaction between pricing and learning, when even the most optimistic consumer would not buy a low-quality product, if sold at cost ($p = c$). In this situation we show that learning can stop with positive probability because the optimal price for the product will be too high. Any price that guarantees some purchase would involve a long-term loss for the seller and is therefore not optimal. Notice that, because the seller is also uninformed about the quality, pricing below cost could be optimal, but only for a short exploring period.

The paper is organized as follows. Section 2 introduces the Bayesian learning model from reviews, and Section 3 establishes the asymptotic learning result and studies the structural properties of the learning trajectory. Section 4 provides some results on the seller's pricing problem.

1.1. Notation

Given any sequence $\{X_t\}$ of independent and identically distributed (i.i.d.) random variables, the distribution function of X_1 is denoted by F_X, its survival function by $\overline{F}_X$, and its density by f_X. The symbol $\mathbb{1}_A$ denotes the indicator of the event A.

2. The Model

A monopolist introduces a product or service of unknown quality to a market of heterogeneous consumers who try to learn about this quality through a social learning mechanism and make their respective purchase decisions accordingly. Specifically, the monopolist introduces a product of intrinsic quality Q that for simplicity is assumed to take one of two possible values, L or H, where $H > L$. The intrinsic quality of the product is determined through a random draw at time $t = 0$ and takes value H with probability π_0 and value L with probability $1 - \pi_0$. The realization of Q is assumed to be unknown to the potential consumers.

Consumers arrive sequentially and are indexed by their arrival time $t \in \{1, 2, \ldots\}$. They are heterogeneous with respect to their preference for the product. Consumer t's preference is represented by her type Θ_t. Types are i.i.d. random variables having a continuous distribution F_Θ, whose support is a (possibly unbounded) interval. We call $\underline{\theta}$ and $\overline{\theta}$ the infimum and supremum of this support, respectively. The type Θ_t is known to consumer t but not to the other consumers. A consumer t who purchases the product will experience a quality level $Q_t = Q + \varepsilon_t$, where ε_t is a random fluctuation around the nominal and initially unknown quality level Q. This fluctuation could be the result of variations in the product itself or even variations in the way individuals experience or perceive quality. The random variables ε_t are i.i.d. with a continuous, zero mean distribution function F_ε, independent of the types Θ_t.

Each consumer t makes a once-and-for-all purchase decision denoted by $B_t \in \{0, 1\}$: she either buys the product ($B_t = 1$) or does not buy it ($B_t = 0$). If a consumer buys the product, her payoff is given by the following simple additive form: $V_t := \Theta_t + Q_t - p$, where p is the price of the product, which, in the first part of this paper, is assumed to be constant over time. If she chooses to forgo the product, her payoff is given by 0, without loss of generality. That is, the payoff of consumer t is given by $B_t V_t$. Whatever the purchase decision is, consumers do not revisit it in later periods.

If $B_t = 1$, once consumer t has bought the product and experienced its quality, she publicly posts a review R_t, where

$$R_t = \begin{cases} \text{👍} & \text{if } B_t = 1, \text{ and } V_t \geq 0, \\ \text{👎} & \text{if } B_t = 1, \text{ and } V_t < 0, \\ \text{✗} & \text{if } B_t = 0, \end{cases}$$

that is, depending on the consumer's ex-post net utility, a review is either positive or negative.[3] Although consumers who do not buy the product do not review it, it is useful to suppose that they report a blank review ✗. We will show that, in contrast to models with private signals, in our model ✗'s are not informative.

Define the time indices of consumers who choose to purchase the product

$$\tau_1 = \min(t \mid B_t = 1) \quad \text{and} \quad \tau_k = \min(t \mid t > \tau_{k-1}, B_t = 1)$$

and let the corresponding review history be, for $\tau_k \leq t < \tau_{k+1}$,

$$h_t = (R_{\tau_1}, \ldots, R_{\tau_k}).$$

Consumer t observes history h_{t-1}. Let $\mathcal{H}_t$ be the set of all histories h_t. Note that the realization of Θ_t and ε_t is never revealed to consumers different from t. Here consumers *generate* information about the quality of the product when they review it, whereas in the literature on social learning with signals, they *reveal* privately held information when making a purchase decision.

The form of the utility function, review decision, information structure, and all the distributions of the relevant random variables are assumed to be common knowledge.

Consumer t chooses either to buy or forgo the product to maximize her expected payoff

$$B_t \mathsf{E}[V_t \mid \Theta_t, h_{t-1}].$$

Note that, conditionally on Q, V_t is independent of the actions of the other consumers, including the ones taken by consumers $1, \ldots, t-1$; past actions affect player t's inference, not her payoff. We call $B = (B_1, B_2, \ldots)$ the sequence of all consumers' decisions. Notice that B_t is h_{t-1}-measurable.

3. Asymptotic Learning
Given history h_t, define

$$\pi(h_t) := \mathsf{P}(Q = H \mid h_t).$$

We frequently use the shorthand notation $\pi_t := \pi(h_t)$. The belief determines the buyers' purchase decision. Consumer t will buy the product if and only if (iff) the expected net utility from buying is greater than zero:

$$\mathsf{E}[V_t \mid h_{t-1}, \Theta_t] = \Theta_t + \mathsf{E}[Q_t \mid h_{t-1}] - p$$
$$= \Theta_t + \pi_{t-1} H + (1 - \pi_{t-1}) L - p \geq 0,$$

or, alternatively, iff $\Theta_t \geq \theta(\pi_{t-1}, p)$, where

$$\theta(\pi, p) := p - (\pi H + (1 - \pi)L) = p - \mathsf{E}_\pi[Q]. \qquad (1)$$

Note that ε_t does not affect the purchase decision because it has zero mean and is independent of the history and Θ_t. Given the purchase criterion, in each period t the seller faces an expected demand function $\overline{F}_\Theta(\theta(\pi_{t-1}, p))$.

Assumption 1.
 a. $\mathrm{supp}(\varepsilon_t) = \mathbb{R}$.
 b. *Given the fixed product price* p, $\overline{F}_\Theta(p - L) > 0$.

Assumption 1(a) ensures that there is always a positive probability that a consumer will derive positive or negative net utility from buying the product, irrespective of whether the intrinsic quality Q is high or low and of the value of the consumer type. Assumption 1(b) will prove necessary for social learning to occur. It ensures that under the firm's price, there will be some consumers who have sufficiently high idiosyncratic type Θ and therefore will choose to buy regardless of the product true quality. The assumption trivially holds when the support of Θ_t is unbounded, and it is reminiscent of assumption 3 in Goeree et al. (2006).

The belief π_t is a random variable in $[0, 1]$. Because of the binary nature of the quality, we can see that $\pi((h_t, r))$ depends on h_t only through $\pi(h_t)$; therefore. the updating procedure is stationary and independent of t. Lemma A.1 (in the appendix) demonstrates some basic properties of the belief. First, starting from any history h_t, the belief increases after a 👍 and decreases after a 👎,

$$\pi((h_t, \text{👎})) \leq \pi(h_t) \leq \pi((h_t, \text{👍})),$$

and this inequality is weak only when $\pi(h_t) \in \{0, 1\}$. This is expected, because the ex-post net utility of a buyer is higher when the intrinsic quality is high, which itself increases the probability of a positive review. The opposite result holds in case of 👎.

Following a no-buy, the posterior will not change, because the sequence of reviews remains unchanged. However, this would have been the case even if ✗ reviews had been observable. A no-buy decision of consumer $t + 1$ merely reveals that her type is lower than $\theta(\pi(h_t), p)$. This carries no information about the quality of the product. This observation is in sharp contrast with the literature on social learning from signals, where any action can be informative by revealing the agent's private signal. Our main result is the following:

Proposition 1. *Let Assumption* 1 *hold. If* $\pi_0 \in (0, 1)$, *then* $\pi_t \to \mathbb{1}_{\{Q=H\}}$ *with probability* 1.

As long as consumers purchase the product, the drift of the belief process is positive when the quality

is high and negative when it is low. As the number of reviews grows large, the posterior will converge and correctly identify the intrinsic quality Q of the product.

Although the learning result seems intuitive it is not trivial, because learning happens only when consumers buy and the section of consumers who buy is not a random representative of the whole population of consumers. The bias that the purchasing induces has to be filtered out to achieve asymptotic learning. Moreover the observations are not exchangeable, and no finite-dimensional sufficient statistic exists. Although the observations are the result of the censoring, due to the purchasing decision of the consumers, the usual models of inference under censored observations cannot be applied because the censoring is endogenous and not independent of the types. The following corollaries will clarify the conditions for learning to occur.

Corollary 1. *If $\overline{F}_\Theta(p - L) = 0$, then with positive probability consumers stop buying at some finite time t, so no learning occurs.*

Corollary 1 shows that it is possible that consumers do not learn the quality of the object, even asymptotically. This is because at some point they stop buying. The reason for this is that their type distribution is bounded above, so when the probability that the quality is high goes below a certain level, nobody has an incentive to buy.

There is a connection and a difference with respect to the classic literature on social learning based on private signals à la Smith and Sørensen (2000). In both cases the boundedness of a distribution makes the difference between learning and not learning. In the private-signal model not learning corresponds to consumers getting stuck on a dominated action. In our model it just implies that consumers do not update their posterior any longer and they stop buying. Nevertheless not buying may be the right decision. This will never be discovered.

Corollary 2. *If $\varepsilon_t = 0$ for all t, then $\pi(h_t) = 0$ if h_t contains at least one 👎. Otherwise $\pi(h_t)$ converges monotonically to 1.*

Corollary 2 says that when the quality of the object is exactly revealed to buyers, learning either happens fast or is monotone. If a consumer posts a 👎, then learning is immediate, because this betrays that the quality is low. Otherwise the probability that the quality is high converges to 1 monotonically and every consumer makes the right decision.

Corollary 3. *If $\mathsf{P}(-\eta \le \varepsilon_1 \le \eta) = 1$, with $\eta < H - L$, then, if $Q = L$, there exist a finite t and a history h_t such that $\mathsf{P}(h_t) > 0$ and $\pi(h_t) = 0$.*

Corollary 3 deals with the case in which the quality is revealed almost exactly. Here the phenomenon of

immediate learning (of the bad quality) can happen with positive probability.

We conclude this section with a couple of structural properties of the dynamics of the learning process. Specifically, in the spirit of comparative statics analysis, we compare outcomes of the social learning process under a single change in model parameters or review histories.

Identical reviews may carry different information, because the reviewers observed different past review histories. Thus, reviews are not exchangeable random variables and potentially carry different weights on the posterior distributions of a future consumer. Do earlier or later reviews carry more weight in forming a posterior belief? How does self-selection bias due to a low belief (after negative reviews) or a high belief (after many positive reviews) affect the likelihood of positive or negative reviews, respectively? The next proposition answers these two questions. It compares the belief resulting from two pairs of reviews: one where a positive review is followed by a negative one, $h_{👍,👎} := (👍, 👎)$, and the reverse sequence, $h_{👎,👍} := (👎, 👍)$. The resulting structural result then offers insight onto the effect of self-selection bias.

Proposition 2.

a. *If f_ε is log-concave, then for any two finite sequences of reviews h' and h'', we have*

$$\pi(h', h_{👍,👎}, h'') \ge \pi(h', h_{👎,👍}, h'').$$

b. *For any $\pi \in (0, 1)$ and $q \in \{L, H\}$ we have that $\mathsf{P}(R_{t+1} = 👍 \,|\, (B_{t+1}, \pi_t, Q) = (1, \pi, q))$ is decreasing in π and $\mathsf{P}(R_{t+1} = 👎 \,|\, (B_{t+1}, \pi_t, Q) = (1, \pi, q))$ is increasing in π.*

Part (a) shows that under some weak assumption on the distribution of ε_t, the earlier review is more influential.[4] This result holds for any distribution of types Θ. Proposition 2(a) should be understood in the context of consumers' self-selection given their unobservable types (e.g., see Li and Hitt 2008). Because reviews vary over time, the corresponding cutoff for purchase, $\theta(\pi_t, p)$, varies as well. Holding Q fixed, a consumer who purchases when the cutoff $\theta(\pi_t, p)$ is low is more likely to be disappointed than a consumer who purchases when the cutoff is high, because the latter is more likely to be a high type. As a result, 👎 is a weaker negative signal in $h_{👍,👎}$ than in $h_{👎,👍}$. Similarly, a purchasing consumer with a low cutoff is less likely to be satisfied, because on average her type is low, resulting is a stronger positive effect of the 👍 in $h_{👍,👎}$ than in $h_{👎,👍}$. In conclusion, earlier reviews have a higher effect on the posterior belief.

Part (b) shows another effect of self-selection bias: namely, that the likelihood that the next review will be positive is decreasing with the belief, or with the

expected quality. Actually it is decreasing with $\pi_{t-1}H + (1 - \pi_{t-1})L - p$. This shows that not only the likelihood of a consumer writing a good review is decreasing in π, it is also increasing in the price p. When either π goes up or p goes down, the probability that a consumer buys increases. With this the probability that a consumer will be unsatisfied increases. All the above is in the sense of comparative statics, that is, ceteris paribus if in situation A the probability π_t is higher (or the price p is lower) than in situation B, then the probability of seeing a favorable review from consumer t is lower in A than in B.

4. Learning and Pricing

In this section we consider a model in which the price is optimally and dynamically chosen by the seller. If the price p_t changes over time, then histories are now richer than in the model considered in the previous section. In particular, we consider the following (price, review) histories for $\tau_k \leq t < \tau_{k+1}$:

$$h_t = (p_{\tau_1}, R_{\tau_1}, \ldots, p_{\tau_k}, R_{\tau_k}), \tag{2}$$

$$h_t^+ = (h_t, p_{t+1}) = (p_{\tau_1}, R_{\tau_1}, \ldots, p_{\tau_k}, R_{\tau_k}, p_{t+1}). \tag{3}$$

At each period t the seller observes h_{t-1}, whereas consumer t observes history h_{t-1}^+. Let $\mathcal{H}_t$ be the set of all histories h_t and $\mathcal{H}_t^+$ the set of all histories h_t^+. As before, we define $\pi_t := \pi(Q = H \mid h_t)$.

Remark 1. (Seller's Information). The seller does not hold any private information about the quality of the product or value of the disturbances around it, but rather is as informed as consumers are on the realization of the product's quality (i.e., the seller does not know ex-ante whether the product is of high quality). Under this assumption, the price is not itself a signal of quality that consumers could use to learn Q. The analysis of the paper would readily extend to a setting in which the seller may in fact know Q but owing to external factors, such as competition or other marketing considerations, the price of the product is again not indicative of the realization of Q.

We assume that there is a cost of producing and marketing an item. This will have relevant implications for the optimal pricing.

Assumption 2. *There is a cost $c > 0$ of putting one item on the market. This cost is constant over time and linear in the number of items.*

At every time t, on the basis of the history h_{t-1}, the seller chooses the price for that period, denoted by p_t. Hence, the seller's strategy is a measurable function $\phi : \mathcal{H} \to A \subset \mathbb{R}_+$, where $\mathcal{H} := \cup_{t \geq 0} \mathcal{H}_t$ is the set of all possible histories. A strategy is called stationary if it depends only on the posterior belief and not on time, so, with an abuse of notation, if ϕ is stationary, we write

$$\phi(h_t) = \phi(\pi_t). \tag{4}$$

We call Φ the set of stationary strategies. The seller seeks to maximize her expected discounted profit, given by

$$\mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t (p_t - c) B_t\right], \tag{5}$$

where $\beta \in (0, 1)$ is the discount factor. If we call

$$\Psi(\phi) := \mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t (\phi(\pi_{t-1}) - c) B_t\right], \tag{6}$$

then the seller program is

$$\max_{\phi \in \Phi} \Psi(\phi)$$

and we define

$$\phi^* \in \arg\max_{\phi \in \Phi} \Psi(\phi),$$

if the maximum is achieved.

Proposition 3. *If the set A of possible prices is finite, then there exists a stationary optimal pricing strategy for the firm.*

As a consequence of Proposition 3, we will make the following assumption:

Assumption 3. *The seller's pricing strategy at time t is a measurable function $\phi : [0, 1] \to \mathbb{R}_+$, whose argument is the posterior probability π_t.*

We consider first the seller's problem of maximizing the expected revenue from a single consumer with belief π. This will be useful in the consideration of the dynamic pricing. Concretely, in the case of a single consumer, the seller is interested in maximizing the revenue function

$$W(\pi, p) := (p - c) D(\pi, p),$$

where

$$D(\pi, p) = \overline{F}_\Theta(\theta(\pi, p)) = \overline{F}_{\Theta + \pi H + (1 - \pi)L}(p) \tag{7}$$

is the demand function. Having defined W, we can rewrite the seller's objective (5) as

$$\mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t W(\pi_{t-1}, p_t)\right].$$

The next assumption is standard in the revenue management literature (see, e.g., Lariviere and Porteus 2001).

Assumption 4. *The revenue function $W(\pi, p)$ has a unique global maximum in p for all $\pi \in [0, 1]$.*

Lariviere (2006) shows that W has a unique global maximizer if the generalized failure rate of $\Theta + \pi H + (1 - \pi)L$ is increasing, where the generalized failure rate of X computed at x is $x f_X(x)/\overline{F}_X(x)$.

Let $p^*(\pi) := \arg\max_{p \in \mathbb{R}_+} W(\pi, p)$ be an optimal price and $W^*(\pi) := W(\pi, p^*(\pi))$ the corresponding maximal revenue.

Lemma 1. *Under Assumption 4 the function $W^*(\pi)$ is convex in π.*

This result is the counterpart of Bose et al. (2006, proposition 3) who show convexity of the revenue function for a model with homogeneous preferences and learning from signals.

Now we consider the problem of dynamic optimal pricing. First we show that social learning is beneficial for the seller.

Theorem 1. *Under Assumptions 2–4 the expected discounted revenue of the seller under social learning is greater than the expected discounted revenue when consumers do not engage in social learning; that is,*

$$\sup_{\phi \in \Phi} \mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t W(\pi_{t-1}, \phi(\pi_{t-1}))\right] \geq \mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t W^*(\pi(h_{t-1}))\right]$$

$$\geq \mathsf{E}\left[\sum_{t=0}^{\infty} \beta^t W^*(\pi_0)\right] = \frac{W^*(\pi_0)}{1 - \beta}.$$

Theorem 1 shows that social learning is beneficial for the seller ex-ante, before knowing the true quality of the product. Ex-post the seller loses when the quality is low (with probability $1 - \pi_0$) and gains when it is high (with probability π_0). To gain intuition, compare the maximal revenue extracted from consumer 1, $W^*(\pi_0)$, and the expected maximal revenue extracted from consumer 2, $\mathsf{E}[W^*(\pi_1)]$. Combining Lemmas 1 and A.4 and Jensen's inequality, we can see that $\mathsf{E}[W^*(\pi_1)] \geq W^*(\mathsf{E}[\pi_1]) = W^*(\pi)$. Thus, on average the seller extracts more revenue from the second consumer than from the first, because of social learning. This argument can be repeated to establish Theorem 1.

This result formalizes and demonstrates the claim of Bose et al. (2006), who argue that social learning benefits the seller in a setting with signals. We point out here that our result does not depend on our particular learning model; it would hold for any learning process with the same preference model as long as π_t is a martingale. In particular, it could be adapted to a setting where consumers are heterogeneous in preferences and learn from signals.

In Proposition 1 we showed that learning is achieved under the hypothesis that some buyers (an infinite number of them) would buy the product even knowing that the quality is low. Assumption 1(b) clarifies which values of the price guarantee this. Under dynamic pricing, if for all $t \geq 1$ we have $\overline{F}_\Theta(p_t - L) > 0$, then a simple modification of the proof of Proposition 1 shows that learning occurs.

We consider now a situation in which buying may stop at some point. As mentioned in Section 2, types are i.i.d. random variables whose support is an interval. In the following proposition we consider a case in which this support is bounded.

Proposition 4. *Assume that $L + \overline{\theta} + \beta(H - L) < c$. If π_t is sufficiently low, then the optimal price is so high that no consumer buys the product.*

Under the assumption of Proposition 4 there is a range of π_t when it is not profitable for the monopolist to sell the product. In that case the consumers do not learn the quality.

When Assumption 1(b) does not hold, learning may fail. In particular, this happens when the types Θ_t are bounded above and the price is high enough as to be sustainable only if the quality is high. In this situation, it may be the case that, at the beginning, the expected quality is sufficiently high for some consumers to purchase the product. If dislikes start piling up, then it may happen that, at some time t, the conditional expected quality $\mathsf{E}[Q|\pi_{t-1}]$ is such that $\overline{F}_\Theta(\theta(\pi_{t-1}, p_t)) = 0$, so no consumer will want to buy the product at price p_t. At this point learning will stop. This may happen when the quality is low (actually in this case it will happen almost surely). This scenario is not so bad for the consumers, because it represents a situation in which initial uncertainty is solved and, as a consequence, consumers make the right decision of not buying. Most of the consumers make the decision that they would have made, had they known for sure the quality of the product. Notice that purchasing may stop also when the actual quality is high, if, by chance, some of the initial reviews happen to be dislikes and the conditional expected quality is pushed down below the critical threshold where nobody wants to buy. In this case learning will stop for the wrong reason and many consumers will not buy the product, whereas they would have bought it, had they had perfect information about the quality. This shows that, with positive probability, an infinite number of consumers will make the wrong decision. This phenomenon is akin to the one observed in bandit models, the difference being that a classic bandit model involves only one decision maker, whereas here we have an infinite sequence of them. A model with a similar flavor has been studied by Papanastasiou et al. (2018).

Definition 1. Call

$$\hat{\pi} := \frac{c - \overline{\theta} - L}{H - L} \tag{8}$$

the threshold below which no consumer buys the product when its price equals the cost c.

The following proposition gives a sufficient condition for subsidizing. That is, under this condition it is profitable for the monopolist to sell the product at a price below cost for a certain period.

We have

$$P(R_t = r \mid \pi(h_t) = \pi, Q = q, p_t = p)$$

$$= \begin{cases} \int_{\theta(\pi,p)}^{\infty} \overline{F}_{\varepsilon}(p - q - x) \, dF_{\Theta}(x) & \text{for } r = \text{👍}, \\ \int_{\theta(\pi,p)}^{\infty} F_{\varepsilon}(p - q - x) \, dF_{\Theta}(x) & \text{for } r = \text{👎}, \\ F_{\Theta}(\theta(\pi,p)) & \text{for } r = \text{✗}. \end{cases}$$

Define

$$G(r, \pi, q, p) := P(R_t = r \mid \pi_t = \pi, Q = q, p),$$

$$G(r, \pi, p) := \pi G(r, \pi, H, p) + (1 - \pi) G(r, \pi, L, p),$$

$$\eta_t := \kappa_{t-1} H + (1 - \kappa_{t-1}) L + \underline{\theta}, \tag{9}$$

where $\kappa_0 \in (0, 1)$ is given and, for all $t \in \{1, 2, \ldots\}$, κ_t is a Bayesian update of κ_{t-1} after a like review at price η_t. For any natural number $n \geq 2$ let

$$P(n) = (\eta_1 - c) + \sum_{t=2}^{n} \left(\beta^{t-1}(\eta_t - c) \prod_{s=0}^{t-2} G(\text{👍}, \kappa_s, \eta_{s+1}) \right). \tag{10}$$

Proposition 5. *Assume that $\pi_0 < \hat{\pi}$. If there exists a natural number $m \geq 2$ such that $P(m) > 0$ with $\kappa_0 = \pi_0$, then $\phi^*(\pi_0) < c$, that is, the optimal policy ϕ^* subsidizes in period 1.*

The proof of Proposition 5 is based on the construction of a suboptimal pricing strategy that subsidizes in period 1 and achieves a positive expected discounted payoff when $P(m) > 0$. Then a comparative argument shows that a fortiori the optimal policy subsidizes in period 1, as well.

When $\pi_0 < \hat{\pi}$ the monopolist can choose between pricing below cost or not selling at all, because consumers would not purchase the product at a price above cost. Assume that in the first period the seller sets such a low price that the probability of purchasing is 1, and she only continues selling if her product receives a like review. Under the given condition $P(m) > 0$, even with this strategy, the seller receives a positive expected payoff, so it is profitable to sell the product. However, when $\pi_0 < \hat{\pi}$, she can only sell if she prices below cost in the first period. An example in the proof shows that the existence of m such that $P(m) > 0$ is not a vacuous hypothesis.

Notice that, whenever learning does not happen under dynamic optimal pricing, it does not happen under any fixed profitable price $p > c$. Proposition 4 shows that under an optimal pricing with positive probability learning may fail. This is because, to keep an active stream of purchases, the price p_t should be set much below cost, so as to be unprofitable in

expectation. When the price is fixed and profitable ($p > c$) learning stops always before it would have stopped under an optimal dynamic pricing, because, by Proposition 5, to keep the stream of purchases active, the optimal price p_t can fall below c, and so below p, before learning stops. In other words, if π_t falls below a certain threshold, consumers would not buy at price p, but they would for some $p_t < p$.

Acknowledgments

Appendix. Proofs
Proofs of Section 3

Given random variables X, its failure rate is denoted by λ_X and its reverse failure rate by ρ_X, that is,

$$\lambda_X(x) = \frac{f_X(x)}{\overline{F}_X(x)} \quad \text{and} \quad \rho_X(x) = \frac{f_X(x)}{F_X(x)}.$$

Notice that

$$P(R_t = r \mid \pi(h_t) = \pi, Q = q)$$

$$= \begin{cases} \int_{\theta(\pi,p)}^{\infty} \overline{F}_{\varepsilon}(p - q - x) \, dF_{\Theta}(x) & \text{for } r = \text{👍}, & \text{(A.1a)} \\ \int_{\theta(\pi,p)}^{\infty} F_{\varepsilon}(p - q - x) \, dF_{\Theta}(x) & \text{for } r = \text{👎}, & \text{(A.1b)} \\ F_{\Theta}(\theta(\pi,p)) & \text{for } r = \text{✗}, & \text{(A.1c)} \end{cases}$$

where $\theta(\pi, p)$ is defined as in (1). We can therefore define the shorthand notation

$$G(r, \pi, q) := P(R_t = r \mid \pi_t = \pi, Q = q), \tag{A.2}$$

$$G(r, \pi) := \pi G(r, \pi, H) + (1 - \pi) G(r, \pi, L). \tag{A.3}$$

Lemma A.1 illustrates some properties of the Bayesian updating.

Lemma A.1. *Let Assumption 1 hold.*
  a. *For all $\pi \in (0, 1)$, $G(\text{👍}, \pi, H) > G(\text{👍}, \pi, L)$.*
  b. *For all $\pi \in (0, 1)$, $G(\text{👎}, \pi, H) < G(\text{👎}, \pi, L)$.*
  c. *There exist $\underline{G}, \overline{G} \in (0, 1)$ such that $\underline{G} \leq G(r, \pi, q) \leq \overline{G}$ for $r \in \{\text{👍}, \text{👎}\}$, $q \in \{H, L\}$, $\pi \in (0, 1)$.*
  d. *For all $q \in \{H, L\}$ and $\pi \in (0, 1)$ we have $G(\text{👎}, \pi, q) + G(\text{👍}, \pi, q) = \overline{F}_{\Theta}(\theta(\pi, p))$.*
  e. *For any history h_{t-1}, we have $\pi(h_{t-1}, \text{✗}) = \pi(h_{t-1})$.*
  f. *Whenever $\pi(h_{t-1}) \in (0, 1)$ we have*

$$\pi(h_{t-1}, \text{👎}) < \pi(h_{t-1}) < \pi(h_{t-1}, \text{👍}).$$

Proof.

a. Because $\overline{F}_\varepsilon$ is nonincreasing, we have

$$G(👍, \pi, H) = \int_{\theta(\pi, p)}^{\infty} \overline{F}_\varepsilon(p - H - x)\, dF_\Theta(x)$$
$$> \int_{\theta(\pi, p)}^{\infty} \overline{F}_\varepsilon(p - L - x)\, dF_\Theta(x) = G(👍, \pi, L).$$

b. Because F_ε is nondecreasing, we have

$$G(👎, \pi, H) = \int_{\theta(\pi, p)}^{\infty} F_\varepsilon(p - H - x)\, dF_\Theta(x)$$
$$< \int_{\theta(\pi, p)}^{\infty} F_\varepsilon(p - L - x)\, dF_\Theta(x) = G(👎, \pi, L).$$

c. This follows from Assumption 1, because there exists a fraction of consumers that would always choose to buy the product, and because the support of ε is large enough.

d. Just add (A.1a) and (A.1b) and consider that given π_t the probability of buying is independent of Q.

e. In general, by Bayes' rule,

$$\pi(h_{t-1}, r) = \frac{\mathsf{P}(R_t = r \mid h_{t-1}, Q = H)\pi(h_{t-1})}{\mathsf{P}(R_t = r \mid h_{t-1}, Q = H)\pi(h_{t-1}) + \mathsf{P}(R_t = r \mid h_{t-1}, Q = L)(1 - \pi(h_{t-1}))}. \quad (A.4)$$

Hence

$$\pi(h_{t-1}, ✗) = \frac{G(✗, \pi(h_{t-1}), H)\pi(h_{t-1})}{G(✗, \pi_t, H)\pi(h_{t-1}) + G(✗, \pi(h_{t-1}), L)(1 - \pi(h_{t-1}))}$$
$$= \pi(h_{t-1}),$$

because $G(✗, \pi, q) = F_\Theta(\theta(\pi, p))$ for all $q \in \{H, L\}$.

f. We have

$$\pi(h_{t-1}, 👎)$$
$$= \frac{G(👎, \pi(h_{t-1}), H)\pi(h_{t-1})}{G(👎, \pi(h_{t-1}), H)\pi(h_{t-1}) + G(👎, \pi(h_{t-1}), L)(1 - \pi(h_{t-1}))}$$
$$< \frac{G(👎, \pi(h_{t-1}), H)\pi(h_{t-1})}{G(👎, \pi(h_{t-1}), H)\pi(h_{t-1}) + G(👎, \pi(h_{t-1}), H)(1 - \pi(h_{t-1}))}$$
$$= \pi(h_{t-1})$$
$$= \frac{G(👍, \pi(h_{t-1}), H)\pi(h_{t-1})}{G(👍, \pi(h_{t-1}), H)\pi(h_{t-1}) + G(👍, \pi(h_{t-1}), H)(1 - \pi(h_{t-1}))}$$
$$< \frac{G(👍, \pi(h_{t-1}), H)\pi(h_{t-1})}{G(👍, \pi(h_{t-1}), H)\pi(h_{t-1}) + G(👍, \pi_t, L)(1 - \pi(h_{t-1}))}$$
$$= \pi(h_{t-1}, 👍).$$

where the first inequality follows from (b) and the second from (a). □

The following lemma is needed to prove Proposition 1.

Lemma A.2. *Define the function*

$$g(x, y, z) = \log\!\left(\frac{x}{y}\right)x + \log\!\left(\frac{z - x}{z - y}\right)(z - x). \quad (A.5)$$

Then, $0 < x \le y < z$ implies $g(x, y, z) \ge 0$ with equality iff $x = y$.

Proof. All variables in the proof are assumed to satisfy the condition in the statement of the lemma. We will show that g is strictly monotonically decreasing in x and that $g(x, x, z) = 0$. We first show that g is convex with respect to the first argument,

$$\frac{\partial g(x, y, z)}{\partial x} = \log\!\left(\frac{x}{y}\right) - \log\!\left(\frac{z - x}{z - y}\right),$$

and,

$$\frac{\partial^2 g(x, y, z)}{\partial x^2} = \frac{1}{x} + \frac{1}{z - x} > 0.$$

It is easy to see that for $\delta > 0$ small,

$$\left.\frac{\partial g(x, y, z)}{\partial x}\right|_{x=y} = 0 \quad \text{and} \quad \left.\frac{\partial g(x, y, z)}{\partial x}\right|_{x=y-\delta} < 0.$$

By convexity we have that $g(x, y, z) \ge g(y, y, z) = 0$, with equality iff $x = y$. □

Proof of Proposition 1. For $t \ge 1$, let $h_t^{\text{full}} := (R_1, \ldots, R_t)$ be the full history including ✗ reviews; recall that in contrast the observed history h_t only includes information from the consumers that bought $\tau_1, \ldots, \tau_k$. Define $\pi_t^{\text{full}} = \pi(h_t^{\text{full}})$. We have

$$\pi_t = \pi_t^{\text{full}} = \mathsf{P}(Q = H \mid h_t^{\text{full}}) \to \pi_\infty,$$

using the martingale convergence theorem (see, for instance, Karlin and Taylor 1975). Because $\pi_t \in [0, 1]$ for all $t \ge 0$, we further conclude that $\pi_\infty \in [0, 1]$ and that $\pi_0 = \mathsf{E}[\pi_t] = \mathsf{E}[\pi_\infty]$.

Recall the definitions of G in (A.2) and (A.3). We have

$$\mathsf{E}[\log \pi_t \mid \pi_{t-1}, Q = H]$$
$$= \log\!\left(\frac{G(✗, \pi_t, H)\pi_{t-1}}{G(✗, \pi_t)}\right)G(✗, \pi_t, H)$$
$$+ \log\!\left(\frac{G(👎, \pi_t, H)\pi_{t-1}}{G(👎, \pi_t)}\right)G(👎, \pi_t, H)$$
$$+ \log\!\left(\frac{G(👍, \pi_t, H)\pi_{t-1}}{G(👍, \pi_t)}\right)G(👍, \pi_t, H)$$
$$= \log \pi_{t-1}$$
$$+ \log\!\left(\frac{G(👎, \pi_t, H)}{G(👎, \pi_t)}\right)G(👎, \pi_t, H)$$
$$+ \log\!\left(\frac{G(👍, \pi_t, H)}{G(👍, \pi_t)}\right)G(👍, \pi_t, H)$$
$$= \log \pi_{t-1} + g\!\left(G(👎, \pi_t, H), G(👎, \pi_t), \overline{F}_\Theta(\theta(\pi_t, p))\right), \quad (A.6)$$

where the second equality stems from

$$G(✗, \pi_t, H) = G(✗, \pi_t, L)$$
$$G(✗, \pi_t, H) + G(👎, \pi_t, H) + G(👍, \pi_t, H) = 1,$$

and Lemma A.1(d). The function g was defined in (A.5).

We can show that there exists $\underline{G} > 0$ such that

$$\underline{G} \leq G(👎, \pi_t, H) \leq G(👎, \pi_t) < \overline{F}_\Theta(\theta(\pi_t, p)) < 1.$$

Indeed, the first inequality follows from Lemma A.1(c); the second inequality stems from Lemma A.1(b) and (A.2); the third inequality follows from Lemma A.1(d) and (A.2); the last inequality is a consequence of Assumption 1(b).

Define

$$\gamma(\pi) := g\Big(G(👎, \pi, H), G(👎, \pi), \overline{F}_\Theta(\theta(\pi, p))\Big).$$

Using Lemma A.1(c) and (A.2), we see that

$$G(👎, \pi_t, H) = G(👎, \pi_t)$$

iff $\pi = 1$. Therefore, by Lemma A.2,

$$\gamma(\pi) > 0 \quad \text{for } \pi \in [0, 1) \quad \text{and} \quad \gamma(\pi) = 0 \quad \text{for } \pi = 1.$$

Assume now, by contradiction, that there exist $\delta, \eta > 0$ such that

$$\mathsf{P}(\pi_\infty < 1 - \eta \mid Q = H) > 2\delta.$$

Then there exists an integer T such that for all $t > T$

$$\mathsf{P}(\pi_t < 1 - \eta \mid Q = H) > \delta.$$

The function γ is continuous and strictly positive for all $\pi \in [0, 1 - \eta]$; therefore,

$$\min_{\pi \in [0, 1-\eta]} \gamma(\pi) =: \underline{\gamma} > 0.$$

Finally, using (A.6) iteratively, we get

$$
\begin{aligned}
\mathsf{E}[\log \pi_{T+k} \mid Q = H] &= \log \pi_0 + \sum_{i=1}^{T+k} \mathsf{E}[\gamma(\pi_i) \mid Q = H] \\
&\geq \log \pi_0 + \sum_{i=T}^{T+k} \mathsf{E}[\gamma(\pi_i) \mid Q = H] \\
&\geq \log \pi_0 + \sum_{i=T}^{T+k} \mathsf{E}[\gamma(\pi_i) \mathbb{1}_{\{\pi_i \leq 1-\eta\}} \mid Q = H] \\
&\geq \log \pi_0 + k \underline{\gamma} \delta,
\end{aligned}
$$

and we conclude that $\mathsf{E}[\log(\pi_{T+k}) \mid Q = H] > 0$ by taking k large enough, which contradicts the fact that $\log \pi_t \leq 0$. Therefore, $\mathsf{P}(\pi_\infty < 1 \mid Q = H) = 0$, or $\mathsf{P}(\pi_\infty = 1 \mid Q = H) = 1$.

The same argument can be used to prove that $\mathsf{P}(\pi_\infty = 0 \mid Q = L) = 1$. □

Proof of Corollary 1. By Lemma A.1(d) we have

$$G(👎, \pi, q) + G(👍, \pi, q) = \overline{F}_\Theta(\theta(\pi, p)).$$

Hence, if $\overline{F}_\Theta(\theta(\pi, p)) = 0$, then

$$G(✗, \pi, q) := \mathsf{P}(R_t = ✗ \mid \pi_t = \pi, Q = q) = 1,$$

that is, with probability 1 no consumer buys. Because the event $\theta(\pi_t, p) = p - L$ has positive probability, buying and therefore learning stop with positive probability. □

Proof of Corollary 2. Consumer t buys iff $\Theta_t + \pi_{t-1}H + (1 - \pi_{t-1})L - p \geq 0$. If $\varepsilon_t = 0$, then she dislikes the product iff $\Theta_t + Q - p < 0$. This can happen only if $Q = L$.

By Lemma A.1(f) $\pi(h_{t-1}) < \pi(h_{t-1}, 👍)$; therefore, if no 👎 appears, then the convergence of π_t to 1 is monotone. □

Proof of Corollary 3. If every ε_t falls in the interval $[-\eta, \eta]$ and $R_t = 👎$, then

$$\Theta_t + \pi_{t-1}H + (1 - \pi_{t-1})L - p \geq 0,$$
$$\Theta_t + Q + \varepsilon_t - p < 0,$$

that is,

$$Q - \eta \leq Q + \varepsilon_t < p - \Theta_t \leq \pi_{t-1}H + (1 - \pi_{t-1})L. \qquad \text{(A.7)}$$

If $Q = H$, then (A.7) holds iff $(1 - \pi_t)(H - L) \leq \eta$. Therefore when $\eta < (1 - \pi_t)(H - L)$, the quality Q can only be L. This happens with positive probability. □

Define

$$\Gamma(\pi) = \frac{\pi}{1 - \pi},$$
$$\Lambda(r, \pi) = \frac{G(r, \pi, H)}{G(r, \pi, L)}.$$

Definition A.1. We say that the condition ILR (increasing likelihood ratio) holds if

$$\Lambda(r, \pi)$$

is nondecreasing in π for $r \in \{👍, 👎\}$.

We next discuss some sufficient conditions for ILR, which will depend on the following definitions.

Definition A.2.
a. The distribution of a random variable X is IFR (increasing failure rate) if its failure rate is nondecreasing.
b. The distribution of a random variable X is DRFR (decreasing reverse failure rate) if its reverse failure rate is nonincreasing.

Proposition A.1. *If the distribution of ε is both IFR and DRFR, then condition ILR holds.*

The proof of Proposition A.1 requires some properties of TP_2 (total positivity of order two), for which the reader is referred to Karlin (1968) and Karlin and Rinott (1980).

Proof of Proposition A.1. Notice that ε is IFR iff its survival function $\overline{F}_\varepsilon$ is log-concave. Write

$$
\begin{aligned}
\mathsf{P}(\Theta + \varepsilon > s, \Theta > \theta) &= \int_\theta^\infty \overline{F}_\varepsilon(s - x) \, dF_\Theta(x) \\
&= \int_{-\infty}^\infty \mathbb{1}_{\{(\theta, \infty)\}}(x) \overline{F}_\varepsilon(s - x) \, dF_\Theta(x).
\end{aligned}
$$

Notice that $\overline{F}_\varepsilon$ is log-concave iff $K(s, x) := \overline{F}_\varepsilon(s - x)$ is TP_2. Moreover $L(x, \theta) := \mathbb{1}_{\{(\theta, \infty)\}}(x)$ is TP_2. Therefore the convolution

$$\int K(s, x) L(x, \theta) \, dF_\Theta(x) = \mathsf{P}(\Theta + \varepsilon > s, \Theta > \theta)$$

is TP$_2$. This implies that for $\pi_1 < \pi_2$ we have

$$P(\Theta + \varepsilon > p - H, \Theta > p - \theta(\pi_2)) P(\Theta + \varepsilon > p - L, \Theta > p - \theta(\pi_1)) \geq$$
$$P(\Theta + \varepsilon > p - H, \Theta > p - \theta(\pi_1)) P(\Theta + \varepsilon > p - L, \Theta > p - \theta(\pi_2)).$$

Hence,

$$G(👍, \pi_2, H) G(👍, \pi_1, L) \geq G(👍, \pi_1, H) G(👍, \pi_2, L),$$

that is,

$$\frac{G(👍, \pi, H)}{G(👍, \pi, L)}$$

is nondecreasing in π.

Next, notice that if ε is DRFR, then its distribution function F_ε is log-concave.

Write

$$P(\Theta + \varepsilon \leq s, \Theta > \theta) = \int_\theta^\infty F_\varepsilon(s - x) \, dF_\Theta(x)$$
$$= \int_{-\infty}^\infty \mathbb{1}_{\{(\theta, \infty)\}}(x) F_\varepsilon(s - x) \, dF_\Theta(x).$$

Notice that F_ε is log-concave iff $K(s, x) := F_\varepsilon(s - x)$ is TP$_2$. Moreover $L(x, \theta) := \mathbb{1}_{\{(\theta, \infty)\}}(x)$ is TP$_2$. Therefore the convolution

$$\int K(s, x) L(x, \theta) \, dF_\Theta(x) = P(\Theta + \varepsilon \leq s, \Theta > \theta)$$

is TP$_2$. This implies that for $\pi_1 < \pi_2$ we have

$$P(\Theta + \varepsilon \leq p - H, \Theta > p - \theta(\pi_2)) P(\Theta + \varepsilon \leq p - L, \Theta > p - \theta(\pi_1)) \geq$$
$$P(\Theta + \varepsilon \leq p - H, \Theta > p - \theta(\pi_1)) P(\Theta + \varepsilon \leq p - L, \Theta > p - \theta(\pi_2)).$$

Hence

$$G(👎, \pi_2, H) G(👎, \pi_1, L) \geq G(👎, \pi_1, H) G(👎, \pi_2, L),$$

that is,

$$\frac{G(👎, \pi, H)}{G(👎, \pi, L)}$$

is nondecreasing in π, and therefore ILR holds. □

A stronger yet simpler sufficient condition is the following.

Corollary A.1. *If the density f_ε is log-concave, then ILR holds.*

Proof. If the density f_ε is log-concave, then both the distribution function F_ε and the survival function $\overline{F}_\varepsilon$ are log-concave; therefore, the proof of Proposition A.1 can be applied. □

Corollary A.1 shows that ILR is a fairly natural assumption on ε given its interpretation as a mean zero noise around the product's quality. For example, ILR holds if ε has a normal or a Gumbel distribution. We can now prove our result.

Proof of Proposition 2(a). We have

$$\Gamma\left(\pi(h_{👍, 👎})\right) - \Gamma\left(\pi(h_{👎, 👍})\right)$$
$$= \Gamma(\pi_0)\Big[\Lambda\left(👍, \pi_0\right)\Lambda\left(👎, \pi(👍)\right)$$
$$- \Lambda\left(👎, \pi_0\right)\Lambda\left(👍, \pi(👎)\right)\Big] \qquad (A.8)$$

We know from Lemma A.1(f) that

$$\pi(👎) \leq \pi_0 \leq \pi(👍).$$

Therefore, by the ILR property (Definition A.1),

$$\Lambda\left(👍, \pi_0\right) \geq \Lambda\left(👍, \pi(👎)\right) \text{ and } \Lambda\left(👎, \pi(👍)\right) \geq \Lambda\left(👎, \pi_0\right),$$

which implies that the right hand side of (A.8) is nonnegative. Nonnegativity of the left-hand side provides the result $\pi(h_{👍, 👎}) \geq \pi(h_{👎, 👍})$.

Note that h' is summarized in π_0 in (A.8). Because the result holds for all π_0, it will hold for all prior histories h'. By monotonicity of the Bayesian update in the belief, the inequality is preserved after history h''. This completes the proof of part (a).

Proof of Proposition 2(b). Note that

$$P\left(R_{t+1} = 👍 \mid (B_{t+1}, \pi_t, Q) = (1, \pi, q)\right) = G(👍, \pi, q)/\overline{F}_\Theta(\theta(\pi, p))$$

and

$$P\left(R_{t+1} = 👎 \mid (B_{t+1}, \pi_t, Q) = (1, \pi, q)\right) = G(👎, \pi, q)/\overline{F}_\Theta(\theta(\pi, p)).$$

For ease of notation, we omit the arguments of the cutoff function $\theta(\pi, p)$ in this proof and write θ instead. Consider the derivative

$$\frac{\partial P\left(R_{t+1} = 👍 \mid (B_{t+1}, \pi_t, Q) = (1, \pi, q)\right)}{\partial \pi}$$

$$= (H - L)(\overline{F}_\Theta(\theta))^{-2}\bigg[F_\varepsilon(p - q - \theta)f_\Theta(\theta)\overline{F}_\Theta(\theta)$$

$$- f_\Theta(\theta)\int_{\theta(\pi, p)}^\infty \overline{F}_\varepsilon(p - q - x)f_\Theta(x) \, dx\bigg]$$

$$= (H - L)f_\Theta(\theta)(\overline{F}_\Theta(\theta))^{-2}\bigg[F_\varepsilon(p - q - \theta)\overline{F}_\Theta(\theta)$$

$$- \int_\theta^\infty \overline{F}_\varepsilon(p - q - \theta)f_\Theta(x) \, dx\bigg]$$

$$< (H - L)f_\Theta(\theta)\overline{F}_\varepsilon(p - q - \theta)(\overline{F}_\Theta(\theta))^{-2}$$

$$\cdot \bigg[\overline{F}_\Theta(\theta) - \int_\theta^\infty f_\Theta(x) \, dx\bigg]$$

$$= 0,$$

where $(H - L) = -\partial\theta(\pi, p)/\partial\pi$, the inequality follows from the fact that $\overline{F}_\Theta$ is decreasing in $x \in [\theta, \infty)$, and the final equality from the definition of survival function. Note that

$$P\left(R_{t+1} = 👍 \mid (B_{t+1}, \pi_t, Q) = (1, \pi, q)\right)$$
$$+ P\left(R_{t+1} = 👎 \mid (B_{t+1}, \pi_t, Q) = (1, \pi, q)\right) = 1,$$

and so

$$\frac{\partial P\left(R_{t+1} = 👎 \mid (B_{t+1}, \pi_t, Q) = (1, \pi, q)\right)}{\partial \pi}$$

$$= -\frac{\partial P\left(R_{t+1} = 👍 \mid (B_{t+1}, \pi_t, Q) = (1, \pi, q)\right)}{\partial \pi}$$

$$> 0,$$

which concludes the proof. □

Proofs of Section 4

The proof of Proposition 3 rests on Blackwell (1965, theorem 7(b)), which, for the sake of completeness, we restate here together with its hypotheses. Call $Q(Y \mid X)$ the set of conditional probabilities on Y given X, that is the set of functions $q(\cdot \mid \cdot)$ such that, for each $x \in X$, $q(\cdot \mid x)$ is a probability on Y and for each Borel subset $B \subset Y$, $q(B \mid \cdot)$ is a Baire function on X.

Lemma A.3. (Blackwell 1965). *Let the following hypotheses hold:*
1. *The state space S is a non-empty Borel set.*
2. *The action space A is finite.*
3. *The conditional probability $q \in Q(S \mid S \times A)$.*
4. *The reward function r is a bounded Baire function on $S \times A \times S$.*
5. *The discount factor $\beta \in [0, 1)$.*
Then there exists an optimal stationary strategy.

Proof of Proposition 3. We now show that the hypotheses of Lemma A.3 are satisfied here.
1. If $\pi_0 \in (0, 1)$, then the state space is an open interval, hence a Borel set.
2. The set of possible prices (i.e., the action set) is finite.
3. Because $\mathsf{P}(R_t = r \mid \pi_{t-1} = \pi, Q = q)$ is continuous in π and q, so are the functions $G(r, \pi, q)$ in (A.2) and $G(r, \pi)$ in (A.3). Therefore the conditional probability π_t is continuous, hence Baire.
4. The daily reward $(p_t - c)B_t$ is a Baire function in (π_{t-1}, p_t). Because the set A of possible prices is finite, the function is also bounded.
5. The discount factor $\beta \in (0, 1)$.

If $\pi_0 \in \{0, 1\}$. Then $\pi_t = \pi_0$ for all t, so it is enough to solve a single-period optimization problem. Therefore, even in this case there is an optimal stationary strategy. □

Lemma A.4. *We have*

$$\mathsf{E}[\pi_{t+1} \mid h_t] = \pi_t. \tag{A.9}$$

Proof of Lemma A.4. For $t \geq 1$ call

$$h_t^{\mathrm{full}} := (p_1, R_1, \ldots, p_t, R_t) \quad \text{and} \quad h_t^{\mathrm{full}+} := (p_1, R_1, \ldots, p_t, R_t, p_{t+1})$$

the full histories including ✗ reviews. As before, $\pi(h_t^{\mathrm{full}}) = \pi(h_t^{\mathrm{full}+})$. Define

$$\pi_t^{\mathrm{full}} = \pi(h_t^{\mathrm{full}}).$$

We have

$$\mathsf{E}[\pi_{t+1} \mid h_t] = \mathsf{E}[\pi_{t+1}^{\mathrm{full}} \mid h_t^{\mathrm{full}}] = \pi_t^{\mathrm{full}} = \pi_t,$$

where the first and last equality follow from Lemma A.1(e) and the second from fact that π_t^{full} is a Doob martingale. □

Proof of Theorem 1. From Lemma A.4 we have that $\mathsf{E}[\pi_{t+1} \mid h_t] = \pi_t$. Therefore for all $t \geq 1$ we have that $\pi(h_t)$ is a dilation of $\pi(h_{t-1})$, hence $\mathsf{E}[\psi(\pi(h_t))] \geq \mathsf{E}[\psi(\pi(h_{t-1}))]$ for all convex functions ψ. A direct implication of Lemma 1 is that the the expected revenue extracted from consumer t is increasing in t, that is, $\mathsf{E}[W^*(\pi(h_{t+1}))] \geq \mathsf{E}[W^*(\pi(h_t))]$. Iteration of the argument proves the result. □

Proof of Proposition 4. By assumption

$$0 < \frac{c - L - \overline{\theta} - \beta(H - L)}{(1 - \beta)(H - L)}.$$

Take a π_t such that

$$\pi_t < \frac{c - L - \overline{\theta} - \beta(H - L)}{(1 - \beta)(H - L)}. \tag{A.10}$$

The highest price that the monopolist can set in period $t + 1$ is $\overline{\theta} + \pi_t H + (1 - \pi_t)L$. Imagine the following optimistic scenario. The next consumer buys the product at period $t + s$ for this highest possible price, and from period $t + s + 1$ the monopolist sells the product every time period at the highest possible price $\overline{\theta} + H$. Then the profit of the monopolist would be

$$\beta^{t+s-1} \left(\overline{\theta} + (\pi_t H + (1 - \pi_t)L) - c + (\overline{\theta} + H - c)\frac{\beta}{1 - \beta} \right),$$

and because of (A.10) this is negative. If even under this optimistic scenario the profit is negative, then under any outcome the profit would be negative. Thus in case of (A.10) the monopolist would have to price so much below cost that there is no scenario that would make it profitable to sell the product. □

Proof of Proposition 5. Given a posterior κ_{t-1}, the price η_t defined in (A.7) guarantees that the probability of a purchase is 1, so the review R_1 is either 👍 or 👎. Now we consider the following suboptimal history-dependent pricing strategy. In period 1 the seller sets price η_1. For any period $t \in \{2, 3, \ldots\}$, if all the previous reviews were 👍, then the seller sells at price η_t. Otherwise, she stops selling forever.

Given this history-dependent pricing strategy, the expected profit of the first m periods is $P(m)$, given in (A.8), which is assumed to be positive. Note that whenever a 👎 appears, the seller's expected payoff from future periods is 0, so $P(m)$ is computed considering only the probability of having a string of positive reviews. Because the expected profit is positive, it is profitable for the firm to sell. Clearly a higher expected payoff could be achieved through a different pricing strategy. However, because $\pi_0 < \hat{\pi}$, consumers will only buy if the first period price is below cost c. So for any pricing strategy with a positive expected payoff, the first period price must be below cost. This means that the seller is forced to subsidize to achieve a positive profit.

To show that the assumptions are not vacuous, we provide a set of parameters for which the condition holds. Let $\beta = 0.999$, $F_\varepsilon \sim N(0, 1)$, $F_\Theta \sim U(0, 1)$, $L = 0$, $H = 100$, $c = 99.1$, and $m = 3$. That is, with the above-mentioned strategy the expected profit within three periods is positive. So the optimal strategy must also provide a positive expected payoff, therefore it is profitable to subsidize in the first period. □

Endnotes

[1] See, for example, the blog and articles at the website https://www.reviewpro.com/blog.

[2] Apart from the relative analytical tractability it yields, it could be motivated by the practical reality that while consumers are learning about some initially unobserved product quality attribute (e.g., the quality of service provided by the hotel staff), the seller is simultaneously learning the impact of that quality on the consumers'

willingness to pay. That is, in many practical settings when a new product is introduced in the market, there is a joint learning problem whereby the seller is learning the demand model while consumers are learning about the product quality. The revenue management literature has studied the seller's problem of learning the demand model. A separate motivation is to consider that the learning interaction is primarily between the review aggregator platform and consumers, and where the platform itself is uninformed, similarly to the consumers it interacts with.

[3] All our results extend to the case when each buying consumer writes a review with fixed positive probability, as long as this is independent of the reviews of past consumers and the experienced utility V_t.

[4] As will become clear in its proof, this result could be stated under more general assumptions. We present it this way for the sake of simplicity.

References

Acemoglu D, Bimpikis K, Ozdaglar A (2014) Dynamics of information exchange in endogenous social networks. *Theoret. Econom.* 9(1):41–97.

Acemoglu D, Dahleh MA, Lobel I, Ozdaglar A (2011) Bayesian learning in social networks. *Rev. Econom. Stud.* 78(4):1201–1236.

Acemoglu D, Makhdoumi A, Malekian A, Ozdaglar A (2017) Fast and slow learning from reviews. NBER Working Paper No. 24046, National Bureau of Economic Research, Cambridge, MA.

Banerjee AV (1992) A simple model of herd behavior. *Quart. J. Econom.* 107(3):797–817.

Besbes O, Scarsini M (2018) On information distortions in online ratings. *Oper. Res.* 66(3):597–610.

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change in informational cascades. *J. Political Econom.* 100(5):992–1026.

Blackwell D (1965) Discounted dynamic programming. *Ann. Math. Statist.* 36:226–235.

Bose S, Orosel G, Ottaviani M, Vesterlund L (2006) Dynamic monopoly pricing and herding. *RAND J. Econom.* 37(4):910–928.

Chamley C (2003) *Rational Herds: Economic Models of Social Learning* (Cambridge University Press, Cambridge, UK).

Crapis D, Ifrach B, Maglaras C, Scarsini M (2017) Monopoly pricing in the presence of social learning. *Management Sci.* 68(11): 3586–3608.

Goeree JK, Palfrey TR, Rogers BW (2006) Social learning with private and common values. *Econom. Theory* 28:245–264.

Herrera H, Hörner J (2013) Biased social learning. *Games Econom. Behav.* 80:131–146.

Karlin S (1968) *Total Positivity*, vol. I (Stanford University Press, Stanford, CA).

Karlin S, Rinott Y (1980) Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* 10(4):467–498.

Karlin S, Taylor HM (1975) *A First Course in Stochastic Processes*, 2nd ed. (Academic Press, New York).

Lariviere MA (2006) A note on probability distributions with increasing generalized failure rates. *Oper. Res.* 54(3):602–604.

Lariviere MA, Porteus EL (2001) Selling to the newsvendor: An analysis of price-only contracts. *Manufacturing Service Oper. Management* 3(4):293–305.

Li X, Hitt LM (2008) Self-selection and information role of online product reviews. *Inform. Systems Res.* 19(4):456–474.

Papanastasiou Y, Bimpikis K, Savva N (2018) Crowdsourcing exploration. *Management Sci.* 64(4):1727–1746.

Smith L, Sørensen P (2000) Pathological outcomes of observational learning. *Econometrica* 68(2):371–398.

Bar Ifrach heads the Marketplace team at Uber Freight. Before Uber Freight, Bar led data science for Airbnb's Homes business unit, across analytics, inference, and algorithms. Bar holds a BA in economics from Tel Aviv University (summa cum laude) and a PhD in operations management from Columbia Business School.

Costis Maglaras is the David and Lyn Silfen Professor of Business at Columbia University. He served as chair of the Decision, Risk & Operations division at the Business School through June 2018 and is a member of the executive committee of Columbia University's Data Science Institute. His research lies on the interface of stochastic modeling and data science.

Marco Scarsini is a professor in the Department of Economics and Finance at LUISS, Rome, Italy. His research interests include applied probability and game theory, with a current focus on congestion games and social learning.

Anna Zseleva is a research fellow at the School of Mathematical Sciences at Tel Aviv University. Her research interests include game theory and decision theory.