Econometrica, Vol. 90, No. 6 (November, 2022), 2857–2899

LEARNING FROM REVIEWS: THE SELECTION EFFECT AND THE SPEED OF LEARNING

DARON ACEMOGLU
Department of Economics, Massachusetts Institute of Technology

ALI MAKHDOUMI
Fuqua School of Business, Duke University

AZARAKHSH MALEKIAN
Rotman School of Management, University of Toronto

ASUMAN OZDAGLAR
Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

This paper develops a model of Bayesian learning from online reviews and investigates the conditions for learning the quality of a product and the speed of learning under different rating systems. A rating system provides information about reviews left by previous customers. observe the ratings of a product and decide whether to purchase and review it. We study learning dynamics under two classes of rating systems: *full history*, where customers see the full history of reviews, and *summary statistics*, where the platform reports some summary statistics of past reviews. In both cases, learning dynamics are complicated by a *selection effect*—the types of users who purchase the good, and thus their overall satisfaction and reviews depend on the information available at the time of purchase. We provide conditions for complete learning and characterize and compare its speed under full history and summary statistics. We also show that providing more information does not always lead to faster learning, but strictly finer rating systems do.

KEYWORDS: Bayesian learning, online reviews, recommendation systems, selection effect, social learning, speed of learning, summary statistics.

1. INTRODUCTION

THE FRACTION OF CONSUMER PURCHASES performed online has reached 16%, and online commerce now accounts for $215 billion. Amazon alone has over 147 million prime customers.[1] The vibrancy of these platforms depends not just on the lower transaction costs they enable but also on successful online information sharing, often based on reviews by past users. Many platforms such as Amazon, Airbnb, eBay, and Uber encourage users to provide reviews and present to new customers/users summaries of the reviews. Despite the centrality of such online reviews, there is relatively little work investigating the specific challenges such systems face and their efficacy in aggregating the dispersed information of diverse users.

Daron Acemoglu: daron@mit.edu
Ali Makhdoumi: ali.makhdoumi@duke.edu
Azarakhsh Malekian: azarakhsh.malekian@rotman.utoronto.ca
Asuman Ozdaglar: asuman@mit.edu

We thank six anonymous referees for constructive suggestions and comments. We also thank participants at various seminars and conferences for useful suggestions. We gratefully acknowledge financial support from the Toulouse Network with Information Technology and Army Research Office, ARO MURI W911NF-12-1-0509.

[1]U.S. Census Bureau News, U.S. Department of Commerce.

We construct a simple benchmark model of Bayesian learning from past reviews. For simplicity, we consider an online platform selling a single product of unknown quality, which is either high or low, and assume that the platform also provides a rating system, consisting of options for reviews for users and a rule for aggregating these reviews. Examples of rating systems include: *full history* (where the reviews of all past users, and their exact sequence, are presented to current users), or more realistically, *summary statistics* (where some summary statistics of past users' reviews are presented to potential new users). The options for reviews include different scores ("like" vs. "dislike" or number of stars) from which users can choose.

Potential users know the rating system of the platform and also their own (ex ante) taste parameter, which determines how likely they are to enjoy the product in question. If they decide to purchase the product, they experience their *material utility*, which depends on the realization of an ex post idiosyncratic preference parameter. They then decide whether to leave a review and the product's rating depending on this material utility. Using this model, we investigate how well (and how rapidly) the information of customers is aggregated by various rating systems.

In addition to developing a benchmark model of learning from online reviews, our analysis has three main contributions. First, we identify a new challenge to learning, which we call the *selection effect*, likely to be particularly relevant in the context of online reviews. Because customers know part of their preferences before purchase, those making a purchase are *selected* depending on their taste parameters. Such selection becomes more pronounced when the information about the product is not very favorable, for example, only those very biased toward a particular type of book or movie would consider buying it if past purchasers have very negative assessments of its quality. The selection effect is different from the difficulty faced by models of observational learning, such as Bikhchandani, Hirshleifer, and Welch (1992), Welch (1992), Banerjee (1992), and Smith and Sørensen (2000). In observational learning models, agents may not be able to learn the underlying state because of "herding"—the possibility that the informative signals of past users cease to affect their behavior because they are themselves following the information of others. In contrast, here, herding issues do not arise because users base their reviews on their own experience, and the information they receive while making their purchase does not directly affect this experience. Instead, the challenge for users is to disentangle past users' preferences concerning the product from their information relayed through reviews. We provide a comprehensive analysis of learning under different rating systems and establish conditions under which Bayesian updating ensures (complete) learning by undoing the implications of the selection effect.

Second, in addition to conditions for complete learning, we investigate *the speed of learning*. Both in the full history and summary statistics cases, we prove that learning is exponentially fast and provide a tight characterization of the speed of learning as a function of the Kullback–Leibler (KL) divergence between the probability distribution of reviews conditional on high versus low quality.[2] Because of the aforementioned selection effect, these probability distributions depend on past users' beliefs. The exact form of this dependence varies between the full history and the summary statistics cases, highlighting how the difficulty of dealing with the selection effect depends on the exact information structure generated by the rating system.

[2]The role of KL divergence in this context is intuitive. Asymptotically, the problem of each individual is similar to a binary hypothesis testing problem (Cover and Thomas (2012, Chapter 11)), though with one important complication: the observations are not conditionally independent. We overcome this complication by using a different approach to prove that the speed of learning still takes the form of KL divergence.

Third, we compare the speed of learning under different rating systems, and show that, in general, providing more information to users does not guarantee faster learning. In particular, we characterize the conditions for observing summary statistics to lead to more rapid learning than the full history of reviews. This is because the more limited information from summary statistics may change equilibrium behavior and inferences.

Throughout, we carry out the analysis separately for the full history case, which enables us to use the martingale convergence theorem and related tools, and for rating systems with summary statistics, which necessitate a different mathematical approach.[3] Complete learning takes place under mild assumptions with full history and requires somewhat more restrictive, but still reasonable, assumptions when the rating system provides summary statistics. This difference arises because the selection effect can be undone when users have access to full history, but not with summary statistics. Conditional on complete learning, however, the speed of learning is given by a similar logic and analogous expressions in the two cases, even though the exact speed of learning differs.

Our results on the speed of learning are not just of methodological interest. A rating system that aggregates the dispersed information of past users accurately but extremely slowly would not be very useful to an online platform that relies on users having this information in real time.

Our work is related to several literatures. The first is the Bayesian observational learning literature, which was mentioned above. Most closely related within this literature are Acemoglu, Dahleh, Lobel, and Ozdaglar (2011), Lobel and Sadler (2015a,b), Mossel, Sly, and Tamuz (2014) and Mossel, Sly, and Tamuz (2015), which study Bayesian observational learning when agents observe a subset of past actions determined according to a stochastic network. Both the difficulty of learning from past information and the techniques used in our analysis are different from those emphasized in these papers. Specifically, the selection effect does not feature in this literature, and in contrast to the Markov-martingale type arguments (e.g., McLennan (1984) and Smith and Sørensen (2000)) or those based on local improvements (e.g., Banerjee and Fudenberg (2004), Acemoglu et al. (2011), and Wolitzky (2018)), our most novel results are based on a characterization of the limiting behavior of dependent stochastic processes (along the lines outlined in footnote 3). Another innovation relative to this literature is our analysis of the speed of learning.

Our paper also relates to several recent papers studying the speed of learning in Bayesian models. Examples include Acemoglu, Dahleh, Lobel, and Ozdaglar (2009) and Rosenberg and Vieille (2019), who characterize the speed of convergence in the baseline observational learning model in some special cases, such as when each agent observes all past actions or just the previous action or one randomly drawn action from the past; Harel, Mossel, Strack, and Tamuz (2014), who study the speed of learning in a setting where finitely many agents repeatedly observe each other's actions; Hann-Caruthers, Martynov, and Tamuz (2018), who compare the speed of convergence in the baseline observational learning model when each agent observes the previous actions versus the case in which they also observe past signals; Dasaratha and He (2019), who study the speed of learning with Gaussian signals; and Vives (1993, 1995), and Amador and Weill (2012), who focus on the speed of learning in rational expectations equilibria where agents learn from prices.

[3]Namely, we construct two distributions, one first-order stochastically dominated (majorized) by the distribution of our summary statistics conditional on high quality, and one first-order stochastically dominating the distribution of our summary statistics conditional on low quality, and prove that these two distributions are asymptotically separated.

Our paper is also related to a few works on online reviews and rating systems. Ifrach, Maglaras, Scarsini, and Zseleva (2019) study a setting similar to the full history version of our model with pricing and investigate the implications of the pricing strategy of the seller on learning. Their model provides guidelines for pricing in this setting, but does not focus on any of our main contributions: analysis of learning with summary statistics, characterization of the speed of learning, and comparison of different rating systems. Besbes and Scarsini (2018) is also closely related. They study a setting similar to ours, but mainly focus on a non-Bayesian learning rule based on averages of utilities reported by previous reviews. They show how this learning rule makes users overestimate quality and how simple heuristics can correct this bias.[4] Other related papers include: Che and Hörner (2018), who investigate the optimal review system to encourage experimentation by early users, Hörner and Lambert (2021), who characterize the trade-off between the informational role of reviews and their impact on the seller's effort on quality, Garg and Johari (2017), who study the implications of pairwise comparisons on online reputation building, and Vellodi (2018) who studies the implications of rating systems on firm incentives to participate in the market.

There is also an emerging empirical literature documenting issues related to the selection effect in the context of online markets. In particular, Hu, Pavlou, and Zhang (2006, 2017), and Hu, Zhang, and Pavlou (2009) document that the distribution of reviews is J-shaped (bimodal) and does not necessarily reveal the product's true quality. They show via online experiments that this is because: (i) only people with higher product valuations purchase a product and they tend to leave positive reviews, and (ii) among people who purchase a product, those with extreme ratings are more likely to express their views. Our model not only provides a theoretical foundation for these effects, but also shows how the rating system impacts the extent of the selection effect, the conditions for complete learning, and the speed of learning.[5]

The rest of the paper is organized as follows. In Section 2, we introduce our model. Sections 3 and 4 provide conditions for complete learning and characterize the speed of learning under full history and summary statistics, respectively. Section 5 compares the speed of learning across a range of rating systems. Section 6 concludes, while proofs are presented in Appendix A and the Online Supplementary Material, Appendix B (Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022)), which also contains several extensions and additional results.

2. ENVIRONMENT

We consider a platform selling, or intermediating the sale of, a product whose quality is unknown to both customers/users and the platform. The platform has a rating system that collects reviews from previous customers and provides a rating of the product (which

---

[4]Other plausible non-Bayesian learning rules and their impact on learning are considered in, among others, Golub and Jackson (2010), Eyster and Rabin (2010), Guarino and Jehiel (2013), Bohren and Hauser (2017), and Frick, Iijima, and Ishii (2020). See Golub and Sadler (2016) for a survey.

[5]Other important empirical works in this area include Chevalier and Mayzlin (2006) and Chua and Banerjee (2016), who examine the impact of reviews on sales, Talwar, Jurca, and Faltings (2007) who document similarity among customer reviews based on their preferences, Li and Hitt (2008), who study the impact of early (positive) reviews on the review trends, and Cai, Chen, and Fang (2009), who undertake an empirical investigation of the baseline observation learning model.

could be a summary statistic of these reviews or their entire history).[6] New customers observe the information from the rating system of the platform and decide whether to purchase the product depending on their ex ante type (valuation). After the purchase decision, their material utility from the product, which depends on true quality, their type, and an additional ex post idiosyncratic preference term, is realized and they decide whether and what review to leave.

2.1. *Customers' Problem and the Rating System*

We assume that the true quality of the product is binary, low or high, and denote it by $Q \in \{0, 1\}$. Without loss of generality, we assume the common prior is that high and low quality are equally likely. A new customer arrives at each time, denoted by her arrival time $t \in \mathbb{N}$, and decides whether to purchase the product. The *material utility* of customer t from the purchase is

$$u_t = \theta_t + \zeta_t + Q - p, \tag{1}$$

where p is the price of the product; θ_t is the ex ante type of the customer, drawn independently from a continuous distribution F_θ; and ζ_t is an ex post idiosyncratic preference term, also drawn independently from a different continuous distribution F_ζ (and even though these variables are independent, it is sometimes convenient to work with their joint distribution, which we denote by $F_{\theta,\zeta}$). The (ex ante) type θ_t captures customer t's valuation based on the features of the product that can be observed before purchase, and the ex post idiosyncratic preference term represents the valuation of the characteristics that customers can evaluate (or experience) only after purchase. Throughout, θ_t and ζ_t are customer t's private information.

Before making the purchase decision, customer t observes the rating of the product provided by the platform based on past reviews (as well as her ex ante valuation θ_t). She then decides whether to purchase the product, which is denoted by $b_t \in \{0, 1\}$. If she purchases ($b_t = 1$), she experiences her material utility and decides whether to leave a review r_t. We assume that the rating system of the platform allows users to leave one of $-\underline{K}, \dots, \overline{K}$ reviews with $\underline{K}, \overline{K} \in \mathbb{N}$, where 0 is interpreted as leaving no review. We let $\mathcal{R} = \{-\underline{K}, \dots, \overline{K}\}$ denote the set of reviews. In what follows, we sometimes refer to the most favorable review $\overline{K}$ as "like" and to the least favorable review $-\underline{K}$ as "dislike." The platform observes the purchase decision b_t as well as the review decision r_t. We denote actions at time t by $a_t \in \mathcal{R} \cup \{N\}$, where $a_t = N$ designates "no purchase" by customer t, that is, $b_t = 0$. The set of actions is denoted by $\mathcal{A} = \mathcal{R} \cup \{N\}$, while the history available to the platform at time t is $h_t = \{a_1, \dots, a_{t-1}\}$, and by convention $h_1 = \emptyset$. The platform has a rating system denoted by $\mathbf{\Omega}$, which at time t maps the history of actions by customers into a rating. We denote the (revealed) rating available to customer t by Ω_t. Examples of rating systems $\mathbf{\Omega}$ are: (i) full history (where $\Omega_t = h_t$), and (ii) systems that report summaries of past reviews, such as fractions of reviews in different bins or certain averages thereof.

Customers' problems can be broken down into two steps: *purchase* and *review*. Let us start with the purchase decision. Customer t observes the realized rating Ω_t and forms her belief regarding the quality of the product, and we assume that she does so using Bayes' rule taking the strategies of other players as given. The purchase decision of customer t,

[6] Throughout, we refer to the scores or other information left by customers as "review," to the aggregate of these reviews provided by the platform as "rating" and to the rules for reporting these reviews as "rating system."

$B_t : \Omega_t \times \mathbb{R}^3 \to \{0, 1\}$, maps the information provided by the rating system at time t, the ex ante type θ_t, the ex post idiosyncratic preference ζ_t, and the price p into a purchase decision. A collection of purchase decision strategies $\mathbf{B} = \{B_t\}_{t=1}^{\infty}$ constitute a Bayes–Nash equilibrium if, for all $t \in \mathbb{N}$,

$$B_t(\Omega_t, \theta_t) = 1 \quad \text{if and only if} \quad \theta_t + \mathbb{E}[\zeta_t] - p + q_t \geq 0,$$

where Bayesian updating gives the belief of customer t as

$$q_t = \mathbb{P}_{\{(\theta_s, \zeta_s)\}_{s=1}^{t-1}}[Q = 1 \mid \Omega_t]. \tag{2}$$

In our baseline analysis, the rating system Ω_t reveals the number of previous customers. In extensions, we consider rating systems that provide information only about a subset of reviews $T \subseteq \mathcal{R}$ and, therefore, the customers cannot directly observe the number of previous customers. Letting τ denote the number of reviews in T observed by a customer in this case, her equilibrium purchase decision in a Bayes–Nash equilibrium depends on her posterior belief about the true quality as given in (2). Customers do not observe any other event and form a belief over the times at which there have been previous purchases. We assume that their prior is that the index of each customer (with respect to calendar time) is drawn from an improper uniform prior. This implies that user posteriors for any history of observations will be that the number of users between any two purchase decisions is uniformly distributed.[7]

2.2. *Review Decisions*

We next turn to review decisions. If customer t purchases the product ($b_t = 1$), then her material payoff, (1), is realized (whether she directly observes her material utility or the underlying quality $Q \in \{0, 1\}$ and the ex post idiosyncratic preference term ζ_t is immaterial). At this point, the customer decides whether to leave a review of the product and what review to leave. We assume that all customers have thresholds denoted by $\lambda_{-\underline{K}} \leq \dots \lambda_{-1} \leq \lambda_1 \leq \dots \lambda_{\overline{K}} \in \mathbb{R}$, and their reviews will be determined by the location of their material utility relative to these thresholds. In particular, customer t chooses review r_t such that

$$r_t = \begin{cases} -\underline{K} & \text{if } u_t < \lambda_{-\underline{K}}, \\ i & \text{if } \lambda_{i-1} \leq u_t < \lambda_i, -\underline{K} < i < 0, \\ 0 & \text{if } \lambda_{-1} \leq u_t < \lambda_1, \\ i & \text{if } \lambda_i \leq u_t < \lambda_{i+1}, 0 < i < \overline{K}, \\ \overline{K} & \text{if } u_t \geq \lambda_{\overline{K}}. \end{cases} \tag{3}$$

The next assumption imposes that there is sufficient "richness" in the distribution of valuations so that such high (low) enough material utilities will be realized with positive probability.

[7]A similar property and all of our results hold so long as each user has a "memoryless" prior over her index in the sequence of all users—so that the number of users between the purchasing user and the previous purchase has a geometric distribution which is independent of the length of history (see Appendices B.3.3 and B.3.4).

ASSUMPTION 1—Richness: *The random variables θ and ζ have continuous and strictly increasing cumulative distribution functions over their supports*, $[\underline{\theta}, \bar{\theta}]$ *and* $[\underline{\zeta}, \bar{\zeta}]$, *respectively. The support of ζ is wide enough so that $\bar{\zeta} + \bar{\theta} - p > \lambda_{\overline{K}}$ (which guarantees that when material utility is near $\bar{\zeta} + \bar{\theta} - p$, the review decision will be $\overline{K}$) and $\underline{\zeta} + \bar{\theta} - p + 1 < \lambda_{-\underline{K}}$ (which ensures that when material utility is near $\underline{\zeta} + \bar{\theta} - p + 1$, the review decision will be $-\underline{K}$).*

Note that the conditions for both the most favorable and the least favorable reviews are in terms of the customer with the highest ex ante valuation, $\bar{\theta}$, because lower-valuation customers may not purchase the product, and thus their review decisions might not be relevant to observed ratings. Moreover, the condition for the most favorable review is evaluated at $Q = 0$, while the condition for the least favorable review is evaluated at $Q = 1$, which ensures that these reviews are not fully revealing about the quality of the product.

REMARK 1: In Appendix A.2, we provide microfoundations for review decisions. We first show that when users have *expressive overall utility*—meaning that their review decisions reflect their wish to express their satisfaction/dissatisfaction—then (3) applies under a natural single crossing condition. We also derive a similar decision rule from a model of *consequentialist utility*, where agents leave reviews to influence the decisions of others, and generalize our main results, Theorems 1–4, to this setting.

3. FULL HISTORY

With full history, the rating system reveals all past actions, that is, $\Omega_t = h_t$. In particular, we are initially assuming that the history also includes "no purchase" and "no review" decisions. This assumption is adopted to simplify the notation in our baseline analysis and is relaxed at the end of the section. We first prove that, as long as customers do not stop purchasing the product, there will be complete learning under full history and then characterize the speed of learning.

3.1. *Learning Dynamics*

When $\Omega_t = h_t$, the belief of customer t, q_t, defined in (2) becomes

$$q_t = \mathbb{P}_{\{(\theta_s, \zeta_s)\}_{s=1}^{t-1}}[Q = 1 \mid h_t].$$

Because in this case history h_t is available to all future customers, we follow the observational learning literature and refer to q_t as *public belief*. We also denote the associated likelihood ratio by

$$l_t = \frac{q_t}{1 - q_t} = \frac{\mathbb{P}_{\{(\theta_s, \zeta_s)\}_{s=1}^{t-1}}[Q = 1 \mid h_t]}{\mathbb{P}_{\{(\theta_s, \zeta_s)\}_{s=1}^{t-1}}[Q = 0 \mid h_t]}. \tag{4}$$

Throughout the paper, we use $\pi(a; F_{\theta,\zeta}, Q, q)$ to denote the probability of action $a \in \mathcal{A}$ given the joint distribution of the valuations θ and ζ, the price of the product p and the thresholds $\lambda_{-\underline{K}}, \ldots, \lambda_{\overline{K}}$ when the true quality is Q and the belief is q. This probability can be written as

$$\pi(a; F_{\theta,\zeta}, Q, q)$$
$$= \begin{cases} \mathbb{P}_{\theta,\zeta}\big[q+\theta+\mathbb{E}[\zeta]-p<0\big], & \text{for } a=N, \\ \mathbb{P}_{\theta,\zeta}\big[q+\theta+\mathbb{E}[\zeta]-p\geq 0, \theta+\zeta+Q-p<\lambda_{-\underline{K}}\big], & \text{for } a=-\underline{K}, \\ \mathbb{P}_{\theta,\zeta}\big[q+\theta+\mathbb{E}[\zeta]-p\geq 0, \lambda_{a-1}\leq\theta+\zeta+Q-p<\lambda_a\big], & \text{for } -\underline{K}<a<0, \\ \mathbb{P}_{\theta,\zeta}\big[q+\theta+\mathbb{E}[\zeta]-p\geq 0, \lambda_{-1}\leq\theta+\zeta+Q-p<\lambda_1\big], & \text{for } a=0, \\ \mathbb{P}_{\theta,\zeta}\big[q+\theta+\mathbb{E}[\zeta]-p\geq 0, \lambda_a\leq\theta+\zeta+Q-p<\lambda_{a+1}\big], & \text{for } 0<a<\overline{K}, \\ \mathbb{P}_{\theta,\zeta}\big[q+\theta+\mathbb{E}[\zeta]-p\geq 0, \theta+\zeta+Q-p\geq\lambda_{\overline{K}}\big], & \text{for } a=\overline{K}. \end{cases}$$

We also let $\boldsymbol{\pi}(F_{\theta,\zeta}, Q, q)$ be the vector of probabilities of all actions $a \in \mathcal{A}$:

$$\boldsymbol{\pi}(F_{\theta,\zeta}, Q, q) = \big(\pi(a; F_{\theta,\zeta}, Q, q) : a \in \mathcal{A}\big).$$

The likelihood ratio at time $t \geq 2$ is thus the product of likelihood ratios of past actions:

$$\begin{aligned} l_t = \frac{q_t}{1-q_t} &= \frac{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}[Q=1 \mid h_t]}{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}[Q=0 \mid h_t]} = \frac{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}[h_t \mid Q=1]}{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}[h_t \mid Q=0]} \\ &= \prod_{s=1}^{t-1} \frac{\pi(a_s; F_{\theta,\zeta}, Q=1, q_s)}{\pi(a_s; F_{\theta,\zeta}, Q=0, q_s)}, \end{aligned} \tag{5}$$

where we used the fact that a priori $Q = 0$ and $Q = 1$ are equally likely.

We next derive the evolution of the public belief. Customer $t+1$ observes a_t (and not θ_t and ζ_t) and she updates her belief as

$$l_{t+1} = l_t \times \frac{\pi(a_t; F_{\theta,\zeta}, Q=1, q_t)}{\pi(a_t; F_{\theta,\zeta}, Q=0, q_t)}.$$

Therefore, the dynamics of public belief are presented by the following stochastic process:

$$l_{t+1} = l_t \times \frac{\pi(a_t; F_{\theta,\zeta}, Q=1, q_t)}{\pi(a_t; F_{\theta,\zeta}, Q=0, q_t)}, \quad \text{w.p. } \pi(a_t; F_{\theta,\zeta}, Q, q_t), a \in \mathcal{A} \text{ for } t \geq 1,$$

where Q is the true quality of the product and $l_1 = 1$. Note that l_t is a sufficient statistic of h_t for (estimating) Q. Moreover, given the thresholds $\lambda_{-\underline{K}}, \dots, \lambda_{\overline{K}}$ and the price p, the law of motion of the likelihood ratio l_t is determined by exogenously-specified distributions.

3.2. *Complete Learning*

The next theorem provides necessary and sufficient conditions for complete learning—almost sure convergence of q_t to the true Q.

THEOREM 1: *Suppose Assumption* 1 *holds*.

1. *If* $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$, *then*, *starting from any initial belief* $q_1 \in (0, 1)$, *there is complete learning with full history*, *that is*, $q_t \to Q$ *almost surely*.
2. *If* $\bar{\theta}+\mathbb{E}[\zeta]-p < 0$, *then starting from any initial belief* $q_1 \in (0, 1)$, *learning is incomplete with positive probability*.

The proof of this theorem, like those of our other main results, Theorems 2–4, is presented in Appendix A.1.

The theorem shows that the condition $\bar{\theta}+\mathbb{E}[\zeta]-p\geq 0$ is sufficient for complete learning starting from any initial belief. When this condition does not hold, for sufficiently pessimistic beliefs about the quality of the product Q, all customers stop buying it. This can be seen by noting that at the time of the purchase, the most positive assessment of expected utility will be from a customer with the highest ex ante valuation $\bar{\theta}$ and is thus $\bar{\theta}+\mathbb{E}[\zeta]+q-p$, where q is the public belief at the time of purchase. When $\bar{\theta}+\mathbb{E}[\zeta]-p<0$, there exists a sufficiently low value of q such that $\bar{\theta}+\mathbb{E}[\zeta]+q-p<0$, implying that once beliefs reach this pessimistic level, even customers with the most positive ex ante valuation stop purchasing, and consequently beliefs remain stuck at q. Conversely, however, when condition $\bar{\theta}+\mathbb{E}[\zeta]-p\geq 0$ holds, even for very pessimistic beliefs about quality, some customers purchase the product and this generates sufficient information for complete learning. When $\bar{\theta}+\mathbb{E}[\zeta]-p=0$, the analysis is more nuanced because for $Q=0$, as the public belief, q_t, converges to Q, the probability of purchase converges to 0, and we prove that complete learning still happens.[8]

The condition $\bar{\theta}+\mathbb{E}[\zeta]-p\geq 0$ plays a role analogous to the unbounded likelihood assumption in baseline models of observational learning (e.g., McLennan (1984) and Smith and Sørensen (2000)). In these models, unbounded likelihood ratio ensures that learning never comes to an end (because there is always the possibility of a very informative signal), and this precludes "herding" where all agents follow the action favored by the public belief, disregarding their own information. This condition similarly rules out "herding in purchase decisions," whereby on the basis of a negative public belief purchasing stops. However, in our framework, there is no "herding in review decisions," because customers leave reviews after experiencing the true quality and what they believed before the purchase decision is irrelevant. But which types of users purchase the product depends on the public belief at the time of purchase, underpinning our selection effect (which we discuss further in Section 3.4).

In the rest of the paper, we impose the following.

ASSUMPTION 2: $\bar{\theta}+\mathbb{E}[\zeta]-p>0$.

Assumption 2 ensures that there is complete learning, so that we can study the speed of learning. It also rules out the edge case where $\bar{\theta}+\mathbb{E}[\zeta]-p=0$, where our characterization of the speed of learning does not apply.

3.3. *Speed of Learning*

We next characterize the speed of learning under full history. For this purpose, we introduce the Kullback–Leibler (KL) divergence between two distributions.

DEFINITION 1—KL Divergence: For two strictly positive distributions $\boldsymbol{\mu}=(\mu_1,\dots,\mu_m)$ and $\boldsymbol{\nu}=(\nu_1,\dots,\nu_m)$ defined on a finite set $\{1,\dots,m\}$, KL divergence is defined as

$$D(\boldsymbol{\mu}\parallel\boldsymbol{\nu})=\sum_{i=1}^{m}\mu_i\log\left(\frac{\mu_i}{\nu_i}\right).$$

[8]In Appendix B.3.1, we strengthen part 2 of Theorem 1 and establish that if $\bar{\theta}+\mathbb{E}[\zeta]-p<0$, then for $Q=0$ almost surely learning is incomplete.

In the rest of the paper, we use the following definition of the speed of learning.

DEFINITION 2—Speed of Learning: For a rating system with exponentially fast (complete) learning, the speed of learning is $\lim_{t\to\infty}\frac{1}{t}\log q_t$ when $Q=0$, and $\lim_{t\to\infty}\frac{1}{t}\log(1-q_t)$ when $Q=1$. We say that a rating system has faster learning than another one if its speed of learning is greater for both $Q=0$ and $Q=1$.

THEOREM 2: *Suppose Assumptions 1 and 2 hold. Then learning is exponentially fast. That is, q_t almost surely converges exponentially to Q. In particular, for $Q=0$, we almost surely have*

$$\lim_{t\to\infty}\frac{1}{t}\log q_t = -D\big(\boldsymbol{\pi}(F_{\theta,\zeta}, Q=0, q=0)\parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q=1, q=0)\big),$$

and for $Q=1$, we almost surely have

$$\lim_{t\to\infty}\frac{1}{t}\log(1-q_t) = -D\big(\boldsymbol{\pi}(F_{\theta,\zeta}, Q=1, q=1)\parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q=0, q=1)\big).$$

This theorem establishes that learning under full history is exponentially fast, and moreover, its exact rate is governed by the KL divergence between the probability distribution of possible actions (i.e., $a\in\mathcal{A}$) when the underlying quality is Q and the public belief is $q=Q$ and the probability distribution when the underlying quality is $1-Q$ (while still $q=Q$). There are three components to the intuition for this result. First, the fact that the learning is exponentially fast follows from the ability of users to overcome the selection effect and combine (the independent components of) past reviews (see Cover and Thomas (2012, Chapter 11)). They can achieve this because they know the distribution of past reviews and can draw the correct inferences from them.

Second, that the speed of learning is given by KL divergence is intuitive as well. We can think of the problem of distinguishing $Q=0$ from $Q=1$ as a binary hypothesis testing problem. The best error exponent for a binary hypothesis testing problem from independently-drawn samples is given by the KL divergence between the probability distributions of these samples conditional on the two hypotheses.[9] The subtlety in our case is that, because of the selection effect, reviews are not conditionally independent—the current belief affects the distribution of types that will purchase, and thus the probability distribution of reviews. Nevertheless, as $q\to Q$ almost surely, we can bound the effects of this dependence and still derive the KL divergence as the measure of the distance between the two relevant probability distributions determining the speed of learning.[10]

[9] See Cover and Thomas (2012, Theorem 11.8.3) for the theory and Glosten and Milgrom (1985) for an application in the context of learning from prices.

[10] More specifically, note that if we could apply the strong law of large numbers, then when $Q=0$, (5) would imply

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{t}\log l_t &= \sum_{a\in\mathcal{A}}\pi(a; F_{\theta,\zeta}, Q=0, q=0)\log\left(\frac{\pi(a; F_{\theta,\zeta}, Q=1, q=0)}{\pi(a; F_{\theta,\zeta}, Q=0, q=0)}\right)\\ &= -D\big(\boldsymbol{\pi}(F_{\theta,\zeta}, Q=0, q=0)\parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q=1, q=0)\big).\end{aligned}$$

Though we cannot directly apply the strong law of large numbers, we can bound the departure from independence by sandwiching l_t between two stochastic processes with independent increments, both converging at the rate stated in the theorem.

Third, note that both probability distributions in the KL divergence condition on $q = Q$. This is because, under full history, each customer correctly reasons about previous users' beliefs, which are converging to Q. It is this feature of the full history rating system that enables the effective filtering out of selection effect and contrasts with our analysis of the case of summary statistics, presented in the next section.

3.4. *The Selection Effect*

The selection effect refers to the fact that because the composition of customers purchasing the product is influenced by information at time t (summarized by q_t), the distribution of reviews depends on this information. For example, with full history, when the public belief q_t is very low, only customers with very high θ purchase and these customers are much more likely to enjoy a high material utility and leave a positive review than the average customer. The next example illustrates the selection effect.

EXAMPLE 1: Consider a rating system with $p = 0$, $\zeta = 0$, $\theta \sim \mathcal{U}[-1, 1]$ and $Q = 0$. Let us assume that customers leave a positive review ("like") when $Q + \theta \geq 0$. Suppose first that the public belief is $q \approx 1$, and so all customers purchase the good because $\theta + q \approx \theta + 1$ is greater than zero for almost all θ. Then the ex ante valuation of customers who have purchased the product is uniformly distributed over $[-1, 1]$, and thus half of the reviews will be positive. In contrast, when the public belief is $q \approx 0$, then only customers with positive θ purchase the product and the conditional distribution of ex ante valuations is uniform over $[0, 1]$. Consequently, all reviews will be positive.

The selection effect, or more generally any additional information, naturally impacts the speed of learning. Suppose that customer t (in addition to h_t) observes the ex ante valuation of previous customers, that is, θ_s for $s = 1, \dots, t-1$. In this setting, the extra information completely removes the selection effect because current customers can condition on past customers' valuations. In Appendix B.3.2, we show that learning is faster in this case than in Theorem 2, and establish a more general result, Proposition B-2: providing any extra information regarding the distribution of previous customers' preferences (weakly) increases the speed of learning.

4. SUMMARY STATISTICS

In this section, we characterize the conditions for complete learning and its speed for more realistic rating systems where the platform provides summary statistics of reviews by past customers.

4.1. *Learning Dynamics*

With summary statistics, customers will not see the full history (i.e., the sequence of reviews) and only observe a vector of statistics $\mathbf{S}$ that includes the fraction of reviews in a subset of all reviews. Formally, we consider a nonempty subset of all possible actions $\mathcal{A}$ denoted by T. A rating system is represented by a partition of T, $\{T_1, \dots, T_m\}$ (i.e., $T = \bigcup_{i=1}^{m} T_i$ and $T_i \cap T_j = \emptyset$, $i \neq j \in [m]$) such that for any $i > j$, all the reviews in the set T_i are more positive than the reviews in the set T_j.[11] We also use the notation τ to denote

[11]This ordering assumption is to make the definitions that will follow (and in particular, positive and negative selections) easier to explain and does not substantially change our analysis. Also, the "no purchase" action can

the times t at which an action in T occurs. In our leading case, τ will be the number of reviews left so far. With this convention, customers observe the number of actions in the set T, and thus know their own τ (and not their exact place in the original sequence, over which they have uniform priors). Therefore, for user τ, the rating system takes the form $\Omega_\tau = \mathbf{S}_{\tau-1}$ and reports information from the first $\tau - 1$ reviews.

A summary statistic is a vector $\mathbf{S}_\tau \in \mathbb{R}^m$ with its ith entry representing fraction of past reviews in T_i, that is,

$$\mathbf{S}_\tau(i) = \frac{1}{\tau}\sum_{s=1}^{\tau}\mathbf{1}\{a_s \in T_i\}, \quad i \in [m].$$

Examples of rating systems with summary statistics include:

(1) The fractions of each one of $\underline{K}+\overline{K}$ reviews are reported. In this case, $\mathbf{S}_\tau \in \mathbb{R}^m$ where $m = \underline{K} + \overline{K}$, and $T = \mathcal{R} \setminus \{0\}$, $T_1 = \{-\underline{K}\}, \dots, T_{\underline{K}} = \{-1\}$, $T_{\underline{K}+1} = \{1\}, \dots, T_{\underline{K}+\overline{K}} = \{\overline{K}\}$.
(2) Averages of the scores of past reviews (see Appendix B.3.5).
(3) "Likes" among all reviews meaning that the rating system reports only the fraction of reviews that give the highest score, $a = \overline{K}$, out of the available $\underline{K} + \overline{K}$ options. In this case, $T = \mathcal{R} \setminus \{0\}$, $T_2 = \{\overline{K}\}$, $T_1 = T \setminus T_2$ and the rating system is represented by $\mathbf{S}_\tau \in \mathbb{R}^2$ where

$$\mathbf{S}_\tau(2) = \frac{1}{\tau}\sum_{s=1}^{\tau}\mathbf{1}\{a_s = \overline{K}\}, \quad \text{and} \quad \mathbf{S}_\tau(1) = 1 - \mathbf{S}_\tau(2). \tag{6}$$

The key object in our analysis is again the expectation of quality Q, conditional on the information available from the rating system at time t, q_t (as defined in (2)). Since in this case the relevant information is summarized by the vector $\mathbf{S}_\tau$ after τ purchases, we write this as $q_\tau = \mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{\tau-1}}[Q = 1 \mid \mathbf{S}_\tau]$. Critically, in contrast to the full history case, q_τ is no longer the public belief because $\mathbf{S}_\tau$ is only observed by the customer making the τth purchase. The likelihood ratio implied by belief q_τ takes an analogous form to (4):

$$l_\tau = \frac{q_\tau}{1-q_\tau} = \frac{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{\tau-1}}[Q = 1 \mid \mathbf{S}_\tau]}{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{\tau-1}}[Q = 0 \mid \mathbf{S}_\tau]} = \frac{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{\tau-1}}[\mathbf{S}_\tau \mid Q = 1]}{\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{\tau-1}}[\mathbf{S}_\tau \mid Q = 0]}. \tag{7}$$

Because future customers do not observe $\mathbf{S}_\tau$ and cannot compute q_τ, the likelihood ratio l_τ is no longer a martingale, and we develop a different approach to study its asymptotic properties.

We denote the probability of observing an action profile a in the set T_i conditional on the action profile being in T by

$$\pi(i; F_{\theta,\zeta}, Q, q, T) = \mathbb{P}_{\theta,\zeta}[a \in T_i \mid a \in T, q, Q], \quad \forall i \in [m].$$

The governing stochastic process for $\mathbf{S}_\tau$ given the true quality Q can then be written as

$$\mathbf{S}_{\tau+1} = \frac{\tau}{\tau+1}\mathbf{S}_\tau + \frac{1}{\tau+1}\mathbf{Y}_{\tau+1}, \quad \forall \tau \geq 0, \tag{8}$$

belong to T in which case it can be included in any of the sets T_j for $j \in [m]$, noting that the information content of the rating system and its speed of learning depends on the set T_j that contains "no purchase."

where $\mathbf{Y}_{\tau+1} \in \mathbb{R}^m$, and

$$\mathbf{Y}_{\tau+1} = \mathbf{e}_i, \quad \text{w.p. } \pi(i; F_{\theta,\zeta}, Q, q_{\tau+1}, T),$$

where $\mathbf{e}_i \in \mathbb{R}^m$ is the ith canonical basis vector, $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$.

In computing (7), agents take into account that there may be unobserved actions (those not purchasing or not leaving reviews), but because the information observed by an agent who did not leave a review is the same as the information available to the next agent, Bayesian customers can consistently compute posterior probabilities (see Lemma A-1 in Appendix A.1).

4.2. *Negative and Positive Selection*

The notions of *negative selection* and *positive selection* play a central role in our analysis of learning with summary statistics. Intuitively, negative selection corresponds to the case where, as belief q_t becomes more favorable to $Q = 1$, reviews become less likely to be positive—they are "negatively selected." Positive selection, conversely, corresponds to the case where as belief q_t becomes more favorable to $Q = 1$, reviews are more likely to be positive. We next formally introduce these notions.

DEFINITION 3—Negative and positive selections: Consider a rating system with m review options $T_1, \dots, T_m$ such that, for any $i' < i$, $\mathbb{P}[a \in \bigcup_{j=i}^m T_j \mid a \in \bigcup_{j=i'}^m T_j, q, Q = 1] > \mathbb{P}[a \in \bigcup_{j=i}^m T_j \mid a \in \bigcup_{j=i'}^m T_j, q, Q = 0]$.[12]

- The rating system has *negative selection* if, for any $i' < i$, $\mathbb{P}[a \in \bigcup_{j=i}^m T_j \mid a \in \bigcup_{j=i'}^m T_j, q, Q]$ is decreasing in q, that is, if the probability of more favorable reviews is decreasing in belief q.
- The rating system has *positive selection* if, for any $i' < i$, $\mathbb{P}[a \in \bigcup_{j=i}^m T_j \mid a \in \bigcup_{j=i'}^m T_j, q, Q]$ is increasing in q, that is, if the probability of more favorable reviews is increasing in belief q.

In both negative and positive selection, as q_t becomes more favorable, customers with lower ex ante valuations (lower θ) become more likely to purchase the product. With negative selection, these additional purchases decrease the likelihood of favorable reviews. With positive selection, on the other hand, they increase the likelihood of favorable reviews.

Whether a rating system features negative or positive selection depends on both the review options available to customers and the distribution of the random variables θ and ζ. The next proposition presents simple sufficient conditions for positive and negative selection in terms of the primitives of the model, given a rating system with $m = \overline{K} - \ell + 1 \geq 2$ review options reporting the fraction of all reviews that are more favorable than ℓ for some $\ell \in \{-\underline{K} + 1, \dots, \overline{K} - 1\}$ (i.e., the set T is $\{\ell, \dots, \overline{K}\}$ and we have $T_1 = \{\ell\}, \dots, T_m = \{\overline{K}\}$).

[12]This condition is not very restrictive. It requires that the probability of a more positive review, $a \in \bigcup_{j=i}^m T_j$, conditional on the information that $a \in \bigcup_{j=i'}^m T_j$ (where $i' < i$) is greater when $Q = 1$ than when $Q = 0$. It always holds when θ has a uniform distribution with a sufficiently wide range and $T_1, \dots, T_m$ is a partition of the set of review options (see Proposition 1). It may be violated in some rare cases where $T_1, \dots, T_m$ suppresses information on some middling reviews.

PROPOSITION 1: *For any* $\ell \in \{-\underline{K}+1, \ldots, \overline{K}-1\}$, *consider a rating system that reports the fraction of all reviews that are more favorable than* ℓ. *If* θ *has a uniform distribution and* $\bar{\theta} \geq \max\{p - \mathbb{E}[\zeta], \lambda_{\overline{K}} - \underline{\zeta} + p\}$, *then for any* $i' < i$, $\mathbb{P}[a \in \bigcup_{j=i}^{m} T_j \mid a \in \bigcup_{j=i'}^{m} T_j, q, Q=1] > \mathbb{P}[a \in \bigcup_{j=i}^{m} T_j \mid a \in \bigcup_{j=i'}^{m} T_j, q, Q=0]$, *and we have*:

1. *If the hazard rate* $\frac{f_\zeta(x)}{1-F_\zeta(x)}$ *is decreasing in* x, *then this rating system features negative selection*.
2. *If the hazard rate* $\frac{f_\zeta(x)}{1-F_\zeta(x)}$ *is increasing in* x, *then this rating system features positive selection*.

The next example illustrates Proposition 1 and presents examples of negative and positive selection.

EXAMPLE 2: Negative Selection: Consider a rating system with two review options reporting the fraction of the most favorable review, $\overline{K}$ (i.e., $T = \{\overline{K}-1, \overline{K}\}$, $T_1 = \{\overline{K}-1\}$, $T_2 = \{\overline{K}\}$ with $\overline{K} \geq 2$). This rating system features negative selection when ζ has decreasing hazard rate (e.g., when the distribution of ζ is Pareto) and θ is uniform with $\bar{\theta} \geq \max\{p - \mathbb{E}[\zeta], \lambda_{\overline{K}} - \underline{\zeta} + p\}$.

Positive Selection: Consider the same rating system, but now suppose that ζ has increasing hazard rate (e.g., when the distribution of ζ is uniform) and θ is again uniform with $\bar{\theta} \geq \max\{p - \mathbb{E}[\zeta], \lambda_{\overline{K}} - \underline{\zeta} + p\}$.

A simpler, even if less realistic, example of positive selection is a rating system with two review options that reports the fraction of "likes" among all customers, that is, $T = \{-1, 0, 1\} \cup \{N\}$ and $T_1 = \{-1, 0\} \cup \{N\}$, $T_2 = \{1\}$ with any distribution of θ and ζ (see Appendix B.2.2).

The selection effect becomes more pronounced when customers have access to summary statistics (rather than the full history as in the previous section) and also when there is negative selection. The former is because customers do not know the exact belief with which the previous actions were taken. To understand the latter claim, suppose the rating system exhibits negative selection and $Q = 0$. Then as q_t approaches 0, only customers with very high θ purchase the product, and they tend to leave more positive reviews. This makes it more difficult for q_t to converge to 0. This intuition also explains why with positive selection, the selection effect will be less burdensome for learning and less harmful to the speed of learning; in this case, as q_t approaches 0, more favorable reviews become less likely, and this helps faster convergence to 0.

4.3. *Complete Learning*

We now define the notion of *separation*, which plays a critical role in our analysis of learning, because it enables customers to filter out the selection effect for rating systems with summary statistics. Intuitively, this condition requires the distribution of at least one review option under $Q = 1$ to be nonoverlapping with its distribution under $Q = 0$. Formally, we have the following.

DEFINITION 4—Separation: A rating system $(T_1, \ldots, T_m)$ satisfies the weak separation condition if there exists a subset of reviews $\mathcal{S} \subseteq [m]$ such that the range of functions $\sum_{i\in\mathcal{S}} \pi(i; F_{\theta,\zeta}, Q=0, q, T)$ and $\sum_{i\in\mathcal{S}} \pi(i; F_{\theta,\zeta}, Q=1, q, T)$ (as functions of q) are weakly separate. Formally, we have either

$$\min_{q\in[0,1]} \sum_{i\in\mathcal{S}} \pi(i; F_{\theta,\zeta}, Q=0, q, T) \geq \max_{q\in[0,1]} \sum_{i\in\mathcal{S}} \pi(i; F_{\theta,\zeta}, Q=1, q, T)$$

or

$$\min_{q\in[0,1]} \sum_{i\in\mathcal{S}} \pi(i; F_{\theta,\zeta}, Q=1, q, T) \geq \max_{q\in[0,1]} \sum_{i\in\mathcal{S}} \pi(i; F_{\theta,\zeta}, Q=0, q, T).$$

A rating system satisfies the strict separation condition if the above inequalities are strict.

THEOREM 3: *Suppose Assumptions 1 and 2 hold, and consider a rating system with summary statistics.*

1. *The strict separation condition is sufficient for complete learning.*
2. *For $m=2$ reviews and negative selection, the weak separation condition is necessary and sufficient for complete learning.*

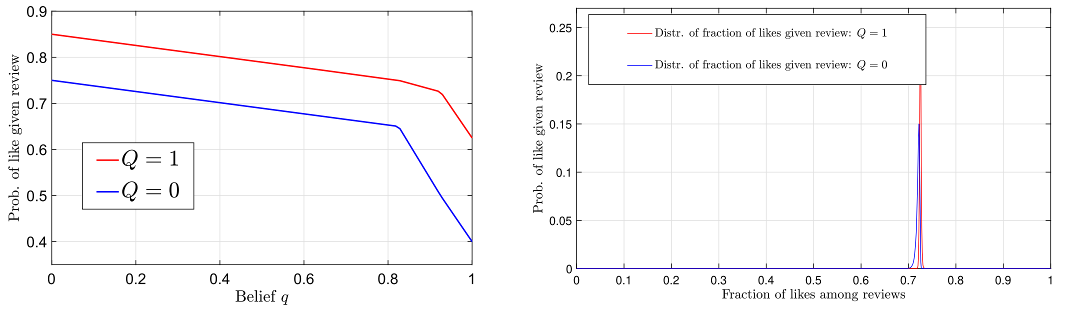
The first part of the theorem shows that, with summary statistics, the strict separation is sufficient for complete learning. The second part establishes a partial converse to this result: for rating systems with $m=2$ reviews and negative selection, the weak separation condition is necessary as well as being sufficient for complete learning. We also note that if Assumption 2 did not hold, then purchases would stop with positive probability and there would be no complete learning, as in Theorem 1.

The proof of Theorem 3 is based on a different approach than those commonly used in this literature. We first provide a recursive characterization of the stochastic process for reviews conditional on the underlying quality of the product, which highlights that the processes under low and high quality are coupled. We then construct two independent distributions, one first-order stochastically dominated by the distribution of summary statistics conditional on low quality and the other one first-order stochastically dominating the distribution conditional on high quality. We finally prove that, under strict separation, these two distributions are asymptotically separated on at least one dimension, enabling us to establish complete learning. This intuition also explains the role of the separation condition. Without this condition, the probability distributions of summary statistics conditional on both $Q=0$ and $Q=1$ would assign positive probabilities to the same asymptotic events, making it impossible for users to learn the true quality. Conversely, when the two probability distributions are separated, for example, for some $i\in[m]$, then $\pi(i; F_{\theta,\zeta}, Q=1, q)$ would be different than $\pi(i; F_{\theta,\zeta}, Q=0, q)$, and thus as τ grows, the marginal distributions of $\mathbf{S}_\tau(i)$ conditional on $Q=0$ and $Q=1$ overlap with lower and lower probability, ensuring complete learning.

The next example illustrates how absence of weak separation leads to failure of complete learning with negative selection, but not necessarily with positive selection.

EXAMPLE 3: Consider a rating system with two review options, "like" and "dislike" denoted by $\overline{K}$ and $-\underline{K}$. Suppose $p=0$ and a customer leaves review $\overline{K}$ when her material utility is positive. Also, let the distribution of θ be close to a distribution with two equally likely point masses at $-7/8$ and $1/2$, and the distribution of ζ be close to four equally likely point masses at $-7/4$, $-1/4$, $1/4$, and $7/4$ (the exact distributions are given in Appendix B.2.3).

For a rating system reporting the fraction of "likes" ($\overline{K}$) among reviews, the probabilities of "like" given reviews as a function of belief q for $Q=0$ and $Q=1$ are depicted in



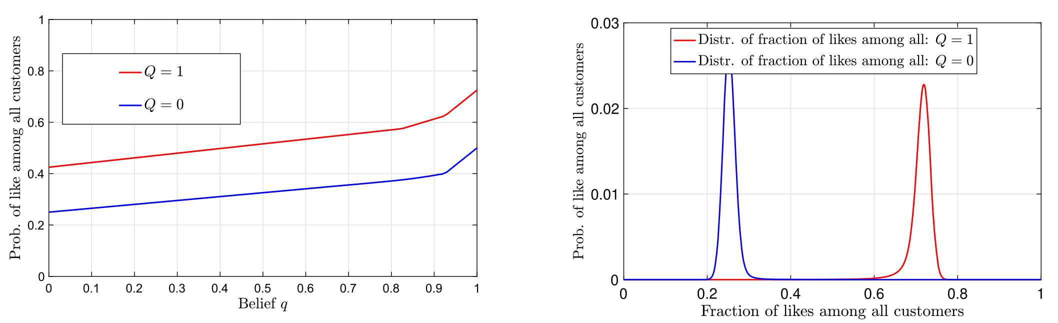
(a) (b)

FIGURE 1.—(a) Probability of "like" given review as a function of $q \in [0, 1]$ for $Q = 1$ and $Q = 0$ (b) distribution of fraction of "likes" for $Q = 1$ and $Q = 0$ for $\tau = 1000$.

panel (a) of Figure 1. The presence of negative selection in this case can be seen from the fact that the curves are downward sloping in panel (a). That there is no weak separation in this case can be seen from the ranges of $\pi(\overline{K}; F_{\theta,\zeta}, Q = 0, q)$ and $\pi(\overline{K}; F_{\theta,\zeta}, Q = 1, q)$ being overlapping. Panel (b) of Figure 1 shows the distribution of the number of "likes" among 1000 reviews for both $Q = 1$ and $Q = 0$ and illustrates that complete learning fails: for beliefs in the overlapping range customers cannot identify the underlying quality.

If, in contrast, the rating system reports fraction of "likes" among all (potential) customers, then the probability of "like" as a function of belief q for both $Q = 0$ and $Q = 1$ is depicted in panel (a) of Figure 2. In this case, the curves are upward sloping in panel (a), indicating positive selection. Even though weak separation again fails (the ranges of $\pi(\overline{K}; F_{\theta,\zeta}, Q = 0, q)$ and $\pi(\overline{K}; F_{\theta,\zeta}, Q = 1, q)$ are overlapping), the distributions of the number of "likes" among 1000 customers for $Q = 1$ and $Q = 0$, shown in panel (b) of Figure 2, are distinct, which ensures complete learning.

The separation condition is stated in terms of probabilities of the realization of different review combinations, the $\pi(i; F_{\theta,\zeta}, Q = 0, q, T)$'s. The next lemma presents a simple property of the distribution of ex ante valuations that is sufficient for this condition to be satisfied.



(a) (b)

FIGURE 2.—(a) Probability of "like" as a function of $q \in [0, 1]$ for $Q = 1$ and $Q = 0$ (b) distribution of fraction of "likes" among all customers for $Q = 1$ and $Q = 0$ for $t = 1000$.

LEMMA 1: *Consider a rating system that reports the fraction of each review among all reviews, that is,* $T = \mathcal{R}$ *and* $T_1 = \{-\underline{K}\}, \ldots, T_m = \{\overline{K}\}$ *where* $m = \overline{K} + \underline{K} + 1$. *For any distribution on* ζ, *if the hazard rate* $\frac{f_\theta(x)}{1-F_\theta(x)}$ *is monotonically increasing in* x, *then the strict separation condition holds.*

A consequence of Lemma 1 is that, for a rating system that reports the fraction of each review among all reviews, strict separation is satisfied for a wide range of distributions of θ, including uniform and normal (see, e.g., Thomas (1971) for a list of distributions with increasing monotone hazard rate).

4.4. *Speed of Learning*

The next theorem shows that, conditional on complete learning, the speed of learning under summary statistics is governed by a KL divergence measure closely related to the full history case.

THEOREM 4: *For a given rating system* $(T_1, \ldots, T_m)$, *suppose Assumptions 1 and 2 and the strict separation condition hold. Then learning is exponentially fast and in particular, for* $Q = 0$, *we almost surely have*

$$\lim_{\tau\to\infty} \frac{1}{\tau} \log q_\tau = -D\big(\boldsymbol{\pi}(F_{\theta,\zeta}, Q=0, q=0, T) \,\|\, \boldsymbol{\pi}(F_{\theta,\zeta}, Q=1, q=1, T)\big),$$

and for $Q = 1$, *we almost surely have*

$$\lim_{\tau\to\infty} \frac{1}{\tau} \log(1-q_\tau) = -D\big(\boldsymbol{\pi}(F_{\theta,\zeta}, Q=1, q=1, T) \,\|\, \boldsymbol{\pi}(F_{\theta,\zeta}, Q=0, q=0, T)\big).$$

The intuition for why KL divergence determines the speed of learning is similar to the full history case. There is a major difference worth noting, however. Let us consider $Q = 0$. While with full history the probability distribution under the alternative hypothesis was still conditioned on $q = 0$, it is now conditioned on $q = 1$. This difference is a consequence of the selection effect: under full history, when drawing inferences from past reviews, customers know the public belief at each point and correct for the selection effect by conditioning on the public belief at the time the review was left. This is not possible with summary statistics. This forces a user into the following inference: if $Q = 1$, then the belief of all other customers, which she does not observe, is very likely to have also converged to $q = 1$. More specifically, consider customer t's learning problem for some large t. In both full history and summary statistics settings, customer t is facing a binary hypothesis testing problem. For both settings, the null hypothesis is conditioned on, say, $Q = 0$ and the alternative is conditioned on $Q = 1$. What about the belief of customer t regarding previous customers' beliefs? With full history, she observes their public beliefs and conditions on them. With summary statistics, however, she does not observe their beliefs and draws inferences about their belief conditioning on her expectation that q is converging to Q, that is, $q \approx 0$ under the null hypothesis and $q \approx 1$ under the alternative hypothesis. This explains why in Theorem 4 the distributions being compared have $q = 0$ and $q = 1$, while in Theorem 2, under full history, these distributions both condition on the same q. In Section 5, we investigate the implications of this difference on the speed of learning under full history and summary statistics.[13]

REMARK 2: We note that the results in Theorems 3 and 4 apply when we consider more general summary statistics (see Appendix B.3.5). In particular, a general summary statistics can be represented by a mapping from the fraction of each of the $k = \underline{K} + \overline{K} + 1$ reviews among previous customers, that is, $f : \Delta^k \to \mathbb{R}^m$, where $\Delta^k = \{(x_1, \ldots, x_k) \in [0, 1]^k : \sum_{i=1}^k x_i = 1\}$.

5. COMPARISON OF RATING SYSTEMS

In this section, we compare the speed of learning generated by various different rating systems. We start with a comparison of full history and summary statistics, then move to a characterization of how different review options affect the speed of learning, and subsequently provide examples of fast and slow learning as the rating system is modified.

5.1. *Full History versus Summary Statistics*

A natural conjecture is that full history, which provides more information than summary statistics based on the same history, should lead to faster learning. In this subsection, we show that learning is indeed faster under full history when there is negative selection, but in fact slower when there is positive selection. In the comparisons, we always hold the set of review options the same and only vary whether the platform presents a summary statistic or the full history of these reviews. Our main result is provided in the next proposition.

PROPOSITION 2: *Consider a rating system with summary statistics that reports the fraction of reviews in sets $T_1, \ldots, T_m$ with $T = \bigcup_{i=1}^m T_i$, that is, $\mathbf{S}_\tau(i) = \frac{1}{\tau} \sum_{s=1}^\tau \mathbf{1}\{a_s \in T_i\}$ where τ is the number of reviews in T (i.e., $\tau = \#\{a \in T\}$), and a rating system with full history that reports the sequence of reviews in the sets $T_1, \ldots, T_m$. Suppose that Assumptions 1 and 2 and the strict separation condition hold (so that there is complete learning under both full history and summary statistics). Then:*
1. *If the rating system has negative selection, the speed of learning under full history is greater than under summary statistics.*
2. *If the rating system has positive selection, then the speed of learning under summary statistics is greater than under full history.*

Part 1 of Proposition 2 shows that with negative selection, having access only to summary statistics slows down learning. Paradoxically, however, part 2 of the proposition shows that with positive selection, the opposite result holds.[14]

The intuition for this result is again related to the selection effect. Recall, first, that the speed of learning is determined by the problem of distinguishing the distribution of

[13]The speed of learning in Theorem 4 is with respect to index τ, the number of customers who left a review in the set T, but the speed of learning with respect to calendar time is straightforward to derive by scaling down the present speed of learning with the probability of taking an action in the set T given $q = Q$, which converges to $\sum_{a\in T} \pi(a; F_{\theta,\zeta}, Q, q = Q)$ in both cases (see Appendix B.3.6).

[14]Two additional observations are worth making. First, the comparison between the speeds of learning under negative and positive selection is with respect to τ, the number of reviews. We show in Appendix B.3.6, the comparison is exactly the same with respect to calendar time (because the same asymptotic scaling factor,

reviews under the true state, say $Q = 1$, and their distribution under the alternative state, $Q = 0$. As emphasized in our discussion following Theorems 2 and 4, under full history, because users observe the public belief, we also have $q = 1$. In contrast, with summary statistics, users infer that when $Q = 1$, we must have $q = 1$, and when $Q = 0$, then we must have $q = 0$. Now suppose there is negative selection. Then, under full history, we are trying to distinguish $(Q = 1, q = 1)$ from $(Q = 0, q = 1)$, but under summary statistics, we are trying to disentangle $(Q = 1, q = 1)$ from $(Q = 0, q = 0)$, which exacerbates the selection effect, leading to slower learning. This is because, under negative selection, false favorable reviews are quite likely when $q = 0$, even if $Q = 0$, making the task of distinguishing the two distributions more difficult. Contrast this with the case of positive selection. Now, the combination $(Q = 0, q = 1)$ generates more false favorable reviews than $(Q = 0, q = 0)$, because, by the definition of positive selection, favorable reviews are less likely when $q = 0$. Consequently, under positive selection, distinguishing $(Q = 1, q = 1)$ and $(Q = 0, q = 0)$ is easier than distinguishing $(Q = 1, q = 1)$ from $(Q = 0, q = 1)$, leading to faster learning under summary statistics than full history.

The next example illustrates the results of Proposition 2.

EXAMPLE 4: Consider a rating system with two review options and $T = \{\overline{K} - 1, \overline{K}\}$, $T_1 = \{\overline{K} - 1\}$, $T_2 = \{\overline{K}\}$ that reports the fraction of the most favorable review among the most favorable two review options, that is, $S_\tau(2) = \frac{1}{\tau}\sum_{s=1}^{\tau} \mathbf{1}\{a_s = \overline{K}\}$ where τ is the number of reviews in the set $\{\overline{K} - 1, \overline{K}\}$. Suppose the distribution of θ is uniform and $\bar{\theta} \geq \max\{p - \mathbb{E}[\zeta], \lambda_{\overline{K}} - \underline{\zeta} + p\}$ and the distribution of ζ is Pareto so that this rating system features negative selection (as in Example 2). Suppose that the conditions in Theorem 3 are satisfied so that there is complete learning under both full history and summary statistics. Then the speed of learning under full history is greater than under summary statistics. However, with the same rating system but now ζ having a uniform distribution, there is positive selection, and in this case, the speed of learning under summary statistics is greater than full history.

5.2. *Learning From Refined Rating Systems*

In this subsection, we show that *more refined* rating systems lead to faster learning, both under full history and under summary statistics. Consider a rating system $\mathbf{\Omega}$ with review options $\mathcal{R} = \{-\underline{K}, \dots, \overline{K}\}$ and thresholds $\Lambda = \{\lambda_{-\underline{K}}, \dots, \lambda_{\overline{K}}\}$. We say that $\mathbf{\Omega}'$ is "coarser" than $\mathbf{\Omega}$ if the review options in $\mathbf{\Omega}'$ are fewer and have thresholds $\Lambda' = \{\lambda_{i_1}, \dots, \lambda_{i_m}\}$ where $i_1 < \cdots < i_m$ and $i_j \in \mathcal{R}$ for $j = 1, \dots, m$.

PROPOSITION 3: *Consider a rating system with either full history or summary statistics, and suppose Assumptions* 1 *and* 2 *and the strict separation condition hold* (*so that there is*

$\sum_{a\in T} \pi(a; F_{\theta,\zeta}, Q, q = Q)$, applies to the relationship between the number of reviews and calendar time in rating systems with full history and summary statistics).

Second, in Proposition 2, we are comparing full history and summary statistics with the same partitioning or review options. Using Theorems 2 and 4, we can also compare the speed of learning under full history and summary statistics with different partitioning. In particular, consider a rating system $(T_1, \dots, T_m)$ with full history and the set of actions $T = \bigcup_{i=1}^{m} T_i$. Then we can show that there exists a rating system with summary statistics and the same set of actions T whose speed of learning is greater than the full history if the speed of learning of the rating system that reports the fraction of each of the actions in T separately is higher than the speed of learning under full history.

complete learning). *Then the speed of learning is always faster under a more refined rating system*.

The intuition for this result is that the more refined rating system $\mathbf{\Omega}$ provides strictly more information about the utility of previous users (and, therefore, their preferences) than the less refined $\mathbf{\Omega}'$, and thus makes it easier for customers to distinguish between the probability distributions of reviews induced under the true state of nature and the alternative.

REMARK 3: In Appendix B.3.5, we extend Proposition 3 to rating systems with general summary statistics (as mentioned in Remark 2). In particular, for a general summary statistic $f : \Delta^k \to \mathbb{R}^m$ and a noninjective function $g : \mathbb{R}^m \to \mathbb{R}^{m'}$, $h = g \circ f$ will represent a coarser summary statistic, and Proposition B-5 shows that this general form of coarsening always reduces the speed of learning.

One application of Proposition 3 is to rating systems that report an average score of past reviews rather than reporting detailed fractions of reviews that fall in different categories. When review thresholds remain unchanged, a rating system reporting average scores is less refined than a rating system reporting detailed fractions, and, from Proposition 3, leads to slower learning (see Appendix B.3.5).

Another application is to rating systems that have "targeted information." In particular, platforms such as Amazon offer information about the reviews of groups of customers with certain characteristics. For instance, for a book at the intersection of climate science and economics, Amazon separately depicts reviews among customers who are interested in economics as well as reviews among customers who are interested in climate science. Providing information on reviews by groups of customers is a form of refinement, and we make this intuition precise in Appendix B.3.7, which establishes that this type of targeting always leads to faster learning.

It is worth noting, however, that Proposition 3 and our other results do not imply that more review options always lead to faster learning: when the platform alters the review options, users' thresholds might also change, reducing the informativeness of their reviews. This will be the case, for example, when the review options are noncomparable between two rating systems or when additional options change the users' review thresholds. The next example shows that a greater number of review options may lead to slower learning.

EXAMPLE 5: Assume ζ is uniform over $[-2, 2]$, θ is uniform over $[-1, 1]$, and $p = 0$. Consider the following two rating systems: (i) there are two review options $\{-1, 1\}$, with threshold 0, that is, the review is 1 if and only if the utility is nonnegative, (ii) there are three review options $\{-2, -1, 1\}$, with thresholds $-1/2$ and $1/2$, that is, the customer chooses -2 if her utility is below $-1/2$; -1 if her utility is between $-1/2$ and $1/2$; and 1 if her utility is above $1/2$. It can be verified that even though the second rating system has more review options, it leads to slower learning.

5.3. *Fast and Slow Learning From Reviews*

In Appendix B.2.1, we provide several examples illustrating how different aspects of the rating system and the extent of heterogeneity among customers affect the speed of learning. Example B-1 illustrates how a small refinement of a rating system can lead to a very large change in the speed of learning because the refinement provides a review option that has a much higher likelihood ratio when $Q = 1$ than when $Q = 0$.

Examples B-2 and B-3 show that an increase in ex post heterogeneity (a wider support or greater variance of ζ) reduces the speed of learning because reviews become less informative about the underlying quality of the product. In particular, in Example B-3 we consider a setting where both θ and ζ are normally distributed (with $\mathbb{E}[\zeta] = 0$ and $\text{var}(\theta) = 1$), and then study how the speed of learning changes as a function of $\text{var}(\zeta)$ and $\mathbb{E}[\theta]$. The speed of learning is decreasing in $\text{var}(\zeta)$ and increasing in $\mathbb{E}[\theta]$. The intuition for the former result is that more disperse ex post preferences make reviews less informative. The intuition for the latter result is that higher $\mathbb{E}[\theta]$ implies that customers who purchase the product are more likely to have a positive experience, exacerbating the selection effect and slowing down learning.

6. CONCLUSION

As the number of goods and services sold online continues to grow rapidly, platforms are increasingly relying on rating systems that provide information on both the quality of various products that are difficult to inspect online and the reputation of distant sellers and service providers. Despite their essential role, properties and efficacy of online rating systems have attracted only limited attention in the recent literature.

In this paper, we presented a model of Bayesian learning from online reviews and investigated the conditions for complete learning of the quality of a product and the speed of learning under different rating systems. In addition to building a benchmark model of learning from online reviews, our analysis has three main contributions. First, we identified a new challenge to learning, the selection effect: the distribution of past reviews will depend on the information available to the users at the time, which may not be known by current users. We developed a systematic analysis of learning in the presence of this selection effect, under rating systems with both full history and summary statistics. The latter case necessitated a new approach for analyzing the limiting behavior of reviews and beliefs. Second, we characterized the speed of learning under both full history and summary statistics, showing that in both cases learning is exponentially fast and is characterized by a KL divergence term. Finally, we studied how different rating systems shape the speed of learning, and showed that more information does not necessarily lead to faster learning and full history may cause slower learning than summary statistics, because it changes the behavior of users and impacts how they deal with the selection effect.

We view our paper as a first step in a comprehensive theoretical analysis of learning from online rating systems. Several interesting directions are worth investigating. First, it is important to study how platform decisions, including design of rating systems and pricing, interact with user learning. We take a first step in this direction in Appendix B. In Appendix B.3.8, we show that if the platform would like to maximize participation by users, then it will always choose a rating system that maximizes the speed of learning. In Appendix B.3.9, we show that our learning and speed of learning results generalize to some environments in which the platform also chooses an endogenous sequence of prices to maximize revenue. Second, another interesting direction is to introduce more strategic interactions between users (our analysis of review decisions intended to influence future purchasing behavior in Appendix A.2.2 is one step in this direction). Third, it is important to move beyond Bayesian learning and investigate what types of rating systems robustly aggregate information when agents use simple learning rules. Finally, a fruitful area for future research would be to close the gap between theoretical models of learning and the burgeoning empirical literature on behavior in online markets, which highlights both the role of selection effects and user incentives.

APPENDIX A

This Appendix presents the proof of the main results and two foundations for the threshold review decisions in equation (3). The remaining proofs, additional results, and examples are presented in the Online Supplementary Material, Appendix B.

A.1. *Proofs*

PROOF OF THEOREM 1: At time t, the likelihood ratio is a random variable defined as

$$Z(a_t \mid l_t) = \frac{\pi(a_t; F_{\theta,\zeta}, Q=1, q=q_t)}{\pi(a_t; F_{\theta,\zeta}, Q=0, q=q_t)}, \quad \forall a_t \in \mathcal{A}, \tag{A-1}$$

where $a_t = a$ with probability $\pi(a; F_{\theta,\zeta}, Q, q=q_t)$. We will use this random variable in the proof of this theorem as well as the proof of Theorem 2.

Part 1: We first provide the proof when $\bar{\theta} + \mathbb{E}[\zeta] - p > 0$ and then show the proof when $\bar{\theta} + \mathbb{E}[\zeta] - p = 0$. Without loss of generality, we assume $Q = 0$ and then prove that $q_t \to 0$ almost surely. The proof for $Q = 1$ is similar. We first establish that l_t forms a martingale, and thus converges to a limiting random variable, and then show that the limiting random variable must be 0 with probability 1.

Note that the random variables $Z(a \mid l_t)$ are all mean 1 (conditional on history). This is because

$$\begin{aligned}\mathbb{E}_{a\sim\pi(F_{\theta,\zeta},Q=0,q=q_t)}\big[Z(a \mid l_t) \mid h_t\big] &= \mathbb{E}_{a\sim\pi(F_{\theta,\zeta},Q=0,q=q_t)}\left[\frac{\pi(a; F_{\theta,\zeta}, Q=1, q=q_t)}{\pi(a; F_{\theta,\zeta}, Q=0, q=q_t)}\right] \\ &= \sum_{a\in\mathcal{A}} \pi(a; F_{\theta,\zeta}, Q=1, q=q_t) = 1.\end{aligned}$$

This guarantees that l_t forms a martingale. Since $l_t \geq 0$, from the martingale convergence theorem (Chapter 5, Durrett (2010)) we conclude that $l_t \to l_\infty$ almost surely. We next prove that the limiting random variable l_∞ is 0 almost surely.

Given any history (or equivalently its sufficient statistic l_t), we have

$$\begin{aligned}Z(a=\overline{K} \mid l_t) &= \frac{\pi(a=\overline{K}; F_{\theta,\zeta}, Q=1, q=q_t)}{\pi(a=\overline{K}; F_{\theta,\zeta}, Q=0, q=q_t)} \\ &= \frac{\mathbb{P}_{\theta,\zeta}\big[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta + 1 - p \geq \lambda_{\overline{K}}\big]}{\mathbb{P}_{\theta,\zeta}\big[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{\overline{K}}\big]} \\ &= 1 + \frac{\mathbb{P}_{\theta,\zeta}\big[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \in [-1, 0)\big]}{\mathbb{P}_{\theta,\zeta}\big[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \geq 0\big]} \\ &\geq 1 + \frac{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \in [-1, 0)\big]}{\mathbb{P}_{\theta,\zeta}\big[1 + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \geq 0\big]},\end{aligned}$$

where the inequality follows by substituting $q_t = 0$ in numerator and $q_t = 1$ in the denominator. Using $\bar{\theta} + \mathbb{E}[\zeta] - p > 0$ and the fact that by Assumption 1, θ and ζ have continuous and strictly increasing cumulative distributions over their supports and $\bar{\theta} + \bar{\zeta} - p > \lambda_{\overline{K}}$, we have

$$\epsilon = \frac{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \in [-1, 0)\big]}{\mathbb{P}_{\theta,\zeta}\big[1 + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{\overline{K}}\big]} > 0. \tag{A-2}$$

This is because both the numerator and denominator of the above expression are strictly positive. In particular, for $\Delta_1 = \min\{1, \bar{\theta} + \mathbb{E}[\zeta] - p\} > 0$, we have

$$\begin{aligned} &\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \in [-1, 0)\big] \\ &\overset{\text{(a)}}{\geq} \mathbb{P}_{\theta,\zeta}[\theta \geq \bar{\theta} - \Delta_1, p + \lambda_{\overline{K}} - \theta - 1 \leq \zeta \leq p + \lambda_{\overline{K}} - \theta] \overset{\text{(b)}}{>} 0, \end{aligned}$$

where (a) follows from the the choice of Δ_1 and (b) follows from $p + \lambda_{\overline{K}} - \theta - 1 \leq p + \lambda_{\overline{K}} - \bar{\theta} + \Delta_1 - 1 \leq p + \lambda_{\overline{K}} - \bar{\theta} < \bar{\zeta}$ and $p + \lambda_{\overline{K}} - \theta - 1 \geq p + \lambda_{-\underline{K}} - \bar{\theta} - 1 > \underline{\zeta}$ (where we used Assumption 1 in both inequalities), showing the numerator is strictly positive. A similar argument shows that the denominator is strictly positive as well. Therefore, irrespective of the belief, whenever $a = \overline{K}$ the random variable $Z(\cdot \mid l)$ is strictly larger than 1, that is, $\min_l Z(a = \overline{K} \mid l) - 1 \geq \epsilon > 0$.

Again, using Assumption 1 and $\bar{\theta} + \mathbb{E}[\zeta] - p > 0$, for any belief, the probability of $a = \overline{K}$ is positive, that is,

$$\begin{aligned} \eta &= \min_{q \in [0,1]} \mathbb{P}_{\theta,\zeta}\big[q + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{\overline{K}}\big] \\ &\geq \mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{\overline{K}}\big] \\ &\geq \mathbb{P}_{\theta,\zeta}\left[\theta \geq \bar{\theta} - \frac{\Delta_2}{2}, \zeta \geq \bar{\zeta} - \frac{\Delta_2}{2}\right] > 0, \end{aligned} \tag{A-3}$$

where $\Delta_2 = \min\{\bar{\theta} + \mathbb{E}[\zeta] - p, \bar{\theta} + \bar{\zeta} - p - \lambda_{\overline{K}}\} > 0$. With these definitions for ϵ and η, for all t and l_t we have

$$\begin{aligned} \mathbb{P}_{a \sim \boldsymbol{\pi}(F_{\theta,\zeta}, Q=0, q=q_t)}\big[\big|Z(a \mid l_t) - 1\big| \geq \epsilon \mid l_t\big] &\geq \pi(a = \overline{K}; F_{\theta,\zeta}, Q = 0, q = q_t) \\ &\geq \mathbb{P}_{\theta,\zeta}\big[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{\overline{K}}\big] \\ &\geq \eta. \end{aligned}$$

We next prove that $l_\infty = 0$ with probability 1. Using (A-2) and (A-3), for an arbitrary $\delta > 0$ we can write

$$\begin{aligned} \mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^t}\big[|l_{t+1} - l_t| \geq \delta\epsilon\big] &= \mathbb{E}_{\{(\theta_s,\zeta_s)\}_{s=1}^t}\big[\mathbf{1}\big\{\big|l_t\big(Z(\cdot \mid l_t) - 1\big)\big| \geq \delta\epsilon\big\}\big] \\ &\geq \mathbb{E}_{\{(\theta_s,\zeta_s)\}_{s=1}^t}\big[\mathbf{1}\{l_t \geq \delta\}\mathbf{1}\big\{\big|Z(\cdot \mid l_t) - 1\big| \geq \epsilon\big\}\big] \\ &\overset{\text{(a)}}{=} \mathbb{E}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}\big[\mathbf{1}\{l_t \geq \delta\}\mathbb{E}_{\theta_t,\zeta_t}\big[\mathbf{1}\big\{\big|Z(\cdot \mid l_t) - 1\big| \geq \epsilon\big\} \mid l_t\big]\big] \\ &= \mathbb{E}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}\big[\mathbf{1}\{l_t \geq \delta\}\mathbb{P}_{\theta,\zeta}\big[\big|Z(\cdot \mid l_t) - 1\big| \geq \epsilon \mid l_t\big]\big] \\ &\overset{\text{(b)}}{\geq} \eta\mathbb{E}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}\big[\mathbf{1}\{l_t \geq \delta\}\big] = \eta\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}[l_t > \delta], \end{aligned} \tag{A-4}$$

where (a) follows from tower property of expectation and (b) follows from (A-2) and (A-3). Since $l_t \to l_\infty$ almost surely, we have $\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t}}[|l_{t+1} - l_t| \geq \delta\epsilon] \to 0$, which along with (A-4), leads to $\mathbb{P}_{\{(\theta_s,\zeta_s)\}_{s=1}^{t-1}}[l_t > \delta] \to 0$. This proves that $l_t \to 0$ in probability, which together with $l_t \to l_\infty$ almost surely, establishes almost sure convergence, that is, $\mathbb{P}[l_\infty = 0] = 1$.[15]

Finally, note that since $l_t = \frac{q_t}{1-q_t}$, from $l_t \to 0$, we have $q_t \to 0$ almost surely, completing the proof of Part 1 when $\bar{\theta} + \mathbb{E}[\zeta] - p > 0$.

We next present the proof when $\bar{\theta} + \mathbb{E}[\zeta] - p = 0$. Note that l_t still forms a martingale and, therefore, almost surely converges. That is, we have $l_t \to l_\infty$ almost surely. We next prove that $l_\infty = 0$ with probability 1. To obtain a contradiction suppose that $\mathbb{P}[l_\infty = 0] < 1$. This implies that there exists $\omega > 0$ such that $\mathbb{P}[l_\infty \geq \omega] > 0$. Because if $\mathbb{P}[l_\infty \geq \omega] = 0$ for all $\omega > 0$, we have $\mathbb{P}[l_\infty \in [0, \omega)] = 1$ for all $\omega > 0$. Then, the continuity from above of probability measure l_∞ ensures that $1 = \lim_{\omega \to 0} \mathbb{P}[l_\infty \in [0, \omega)] = \mathbb{P}[l_\infty = 0]$, which contradicts $\mathbb{P}[l_\infty = 0] < 1$. Therefore, there exists $\omega > 0$ such that $\mathbb{P}[l_\infty \geq \omega] > 0$. We can bound $Z(a = \overline{K} \mid l)$ for $l \geq \omega$ for some $v > 0$ as follows:

$$\begin{aligned}
\min_{l \geq \omega} Z(a = \overline{K} \mid l) &= \min_{l \geq \omega} 1 + \frac{\mathbb{P}_{\theta,\zeta}\big[q + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \in [-1, 0)\big]}{\mathbb{P}_{\theta,\zeta}\big[q + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \geq 0\big]} \\
&\geq 1 + \frac{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] - p + \dfrac{\omega}{1+\omega} \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \in [-1, 0)\big]}{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] - p + 1 \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \geq 0\big]} \\
&= 1 + v, \qquad\qquad \text{(A-5)}
\end{aligned}$$

where we used Assumption 1 and in particular the facts that θ and ζ have continuous and strictly increasing cumulative distribution functions over their supports and that $Z(a = \overline{K} \mid l)$ is continuous for $l \geq \omega$ and, therefore, the minimum is achievable. That $v > 0$ follows from $\bar{\theta} + \mathbb{E}[\zeta] - p = 0$, with a similar argument to (A-2). Therefore, given $\omega > 0$, the numerator $\mathbb{P}_{\theta,\zeta}[\theta + \mathbb{E}[\zeta] - p + \frac{\omega}{1+\omega} \geq 0, \theta + \zeta - p - \lambda_{\overline{K}} \in [-1, 0)]$ is strictly positive. Now since $\mathbb{P}[l_\infty \geq \omega] > 0$, for $\varrho = \min\{\frac{1}{2}\omega, \frac{v}{2+v}\omega\}$, there exists $l^* \geq \omega + \varrho$ such that $\mathbb{P}[l_\infty \in [l^* - \varrho, l^* + \varrho)] > 0$. The existence of l^* follows from the fact that $[\omega, \infty) = \bigcup_{i=1}^{\infty}[\omega + 2(i-1)\varrho, \omega + 2i\varrho)$. The subadditivity of the probability measure l_∞ implies $\mathbb{P}[l_\infty \in [\omega, \infty)] \leq \sum_{i=1}^{\infty} \mathbb{P}[l_\infty \in [\omega + 2(i-1)\varrho, \omega + 2i\varrho)]$. Because the left-hand side of the above inequality is positive and the right-hand side is the summation of countably many terms, one of the terms must be positive. Letting $i^* \geq 1$ be the index of the positive term, we can take $l^* = \frac{1}{2}((\omega + 2(i^* - 1)\varrho) + (\omega + 2i^*\varrho))$. This implies

$$\big|l\, Z(a = \overline{K} \mid l) - l\big| \overset{\text{(a)}}{\geq} \big(l^* - \varrho\big)v \overset{\text{(b)}}{\geq} 2\varrho, \quad \text{for all } l \in \big[l^* - \varrho, l^* + \varrho\big), \qquad \text{(A-6)}$$

where (a) follows from inequality (A-5), $l^* \geq \omega + \varrho$, and $l \geq l^* - \varrho \geq \omega$ and (b) follows from $l^* \geq \varrho$ and $\varrho \leq \frac{v}{2+v}\omega$. This inequality establishes that if $l_t \in [l^* - \varrho, l^* + \varrho)$ and the review decision at time t is $\overline{K}$, then the likelihood ratio in the next round, l_{t+1}, will not be in the interval $[l^* - \varrho, l^* + \varrho)$. Moreover, using Assumption 1 and a similar argument to

[15]To see this, first note that from $l_t \to l_\infty$ almost surely (a.s.), we have $l_t \to l_\infty$ in probability and the result follows by noting that for a sequence of random variables $\{X_n\}$, if $X_n \to X$ and $X_n \to Y$ in probability, then $X = Y$ a.s.

(A-3), for all $l \in [l^* - \varrho, l^* + \varrho)$ with probability at least $\rho = \mathbb{P}_{\theta,\zeta}[\theta + \mathbb{E}[\zeta] - p + \frac{l^*-\varrho}{1+l^*-\varrho} \geq 0, \theta + \zeta - p \geq \lambda_{\overline{K}}] > 0$, the review decision is $\overline{K}$, and hence using (A-6), we conclude that the likelihood ratio of the next round will not be in the interval $[l^* - \varrho, l^* + \varrho)$ with positive probability.

Since $\mathbb{P}[l_\infty \in [l^* - \varrho, l^* + \varrho)] > 0$, with positive probability there exists t_0 such that for $t \geq t_0$, the likelihood ratio l_t is in the interval $[l^* - \varrho, l^* + \varrho)$. But we just proved that if $l_t \in [l^* - \varrho, l^* + \varrho)$, then l_{t+1} is not in the interval $[l^* - \varrho, l^* + \varrho)$ with probability at least $\rho > 0$. Letting E_n be the event that l_{t_0+n} does not fall into interval $[l^* - \varrho, l^* + \varrho)$, we have $\sum_{n=1}^{\infty} \mathbb{P}[E_n] = \infty$. Using the second Borel–Cantelli lemma, we can conclude that with probability 1 the events $\{E_n\}_{n=1}^{\infty}$ occur infinitely often, which contradicts the fact that for $t \geq t_0$, l_t is in the interval $[l^* - \varrho, l^* + \varrho)$. This completes the proof of Part 1.

Part 2: We break the proof into two steps. In the first step, we prove that with a positive probability, starting from any initial belief $q_1 \in (0, 1)$, in finite time q_t becomes very small. In the second step, we show that once this happens and given the assumption of part 2, which is $\bar{\theta} + \mathbb{E}[\zeta] - p < 0$, no purchase takes place and learning stops.

Step 1: Let $\Delta = -(\bar{\theta} + \mathbb{E}[\zeta] - p)$, where by the assumption of part 2, $\Delta > 0$. First, note that if $\Delta > 1$ then $\bar{\theta} + \mathbb{E}[\zeta] - p + q \leq -\Delta + 1 < 0$, which implies that when $\Delta > 1$, no purchase happens. Therefore, we only need to consider the case where $\Delta \leq 1$. In this case, we prove that there exists a finite t such that, starting from any belief, with positive probability we have $q_t < \Delta$.

For any $q \geq \Delta$, using Assumption 1 and a similar argument to that of (A-2), we have

$$\max_{q \in [\Delta, 1]} \frac{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta + 1 - p \leq \lambda_{-\underline{K}}\big]}{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-\underline{K}}\big]} < 1.$$

This is because a necessary condition to satisfy the first constraint in the numerator and denominator is to have $\theta \geq \bar{\theta} - (1 - \Delta)$. Using this condition, we have $\underline{\zeta} < \lambda_{-\underline{K}} + p - \bar{\theta} - 1 \leq \lambda_{-\underline{K}} + p - \theta - 1 \leq \lambda_{\overline{K}} + p - (\bar{\theta} - (1 - \Delta)) - 1 < \bar{\zeta}$, which implies that the difference between denominator and numerator is bounded from below by $\mathbb{P}_{\theta,\zeta}[\theta \geq \bar{\theta} - (1 - \Delta), \lambda_{-\underline{K}} + p - \theta - 1 \leq \zeta \leq \lambda_{-\underline{K}} + p - \theta] > 0$. Therefore, we obtain

$$\lim_{t \to \infty} \left(\max_{q \in [\Delta, 1]} \left(\frac{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta + 1 - p \leq \lambda_{-\underline{K}}\big]}{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-\underline{K}}\big]} \right) \right)^t = 0.$$

Let T_0 be the smallest number such that for all $t \geq T_0$, we have

$$\frac{q_1}{1 - q_1} \left(\max_{q \in [\Delta, 1]} \left(\frac{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta + 1 - p \leq \lambda_{-\underline{K}}\big]}{\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-\underline{K}}\big]} \right) \right)^t < \frac{\Delta}{1 - \Delta}. \tag{A-7}$$

Notice that such T_0 exists because $q_1 \in (0, 1)$. Let us also define

$$\rho = \left(\min_{q \in [\Delta, 1]} \left(\mathbb{P}_{\theta,\zeta}\big[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-\underline{K}}\big] \right) \right)^{T_0} > 0,$$

where ρ denotes a lowerbound on the probability of having T_0 consecutive dislikes when belief is above Δ. Note that under this condition, (A-7) implies that $q_{T_0} < \Delta$.

We next prove that, starting from initial belief $q_1 \in (0, 1)$, with probability at least ρ there exists a time $t \in [1, T_0]$ such that $q_t < \Delta$. Let $\mathcal{E}$ be the event that in the interval

$[1, T_0]$, the belief goes below Δ and $\mathcal{E}_1$ be the event that in the interval $[1, T_0 - 1]$, the belief goes below Δ. We can write

$$\mathbb{P}[\mathcal{E}] = \mathbb{P}[\mathcal{E} \mid \mathcal{E}_1]\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E} \mid \mathcal{E}_1^c]\mathbb{P}[\mathcal{E}_1^c] = \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[q_{T_0} < \Delta \mid q_t \geq \Delta, 1 \leq t \leq T_0 - 1]\mathbb{P}[\mathcal{E}_1^c]$$
$$\geq \mathbb{P}[\mathcal{E}_1] + \rho\mathbb{P}[\mathcal{E}_1^c] \geq \rho,$$

where the probabilities are over random variables $\{(\theta_s, \zeta_s)\}_{s=1}^{T_0}$. Therefore, with probability $\rho > 0$, there exists $t^* \in [1, T_0]$ such that $q_{t^*} < \Delta$.

Step 2: For all $t \geq t^*$, we have $q_t = q_{t^*}$ and the limiting belief becomes $q_{t^*} \neq Q$.

We establish Step 2 by induction on t. It holds for $t = t^*$. Since $q_{t^*} - p + \theta + \mathbb{E}[\zeta] < \Delta + \bar{\theta} + \mathbb{E}[\zeta] - p = 0$, purchase does not happen at time t^*. Since purchase does not happen at time t^*, the belief at time $t^* + 1$ is the same as q_{t^*} because $\frac{\mathbb{P}_{\theta,\zeta}[a=N|Q=1,l_{t^*}]}{\mathbb{P}_{\theta,\zeta}[a=N|Q=0,l_{t^*}]} = \frac{\mathbb{P}_{\theta,\zeta}[\theta+\mathbb{E}[\zeta]+q_{t^*}-p<0]}{\mathbb{P}_{\theta,\zeta}[\theta+\mathbb{E}[\zeta]+q_{t^*}-p<0]} = 1$. Therefore, no purchase occurs at time $t^* + 1$. By repeating this argument, no purchase occurs for any $t \geq t^*$ and $q_t = q_{t^*}$. Finally, note that q_{t^*} is away from $Q = 0$ (similarly from $Q = 1$) because using Assumption 1 if purchase occurs, then the probability of any review is nonzero. Hence, the likelihood ratio at each time is multiplied by a number bounded away from zero and cannot become 0 in finite time. *Q.E.D.*

PROOF OF THEOREM 2: We prove the theorem for $Q = 0$ as the proof for $Q = 1$ can be obtained by a similar argument. Recall that $\{l_t\}_{t=0}^{\infty}$ is the sequence of likelihood ratio of beliefs. The proof of this theorem follows by starting from the sequence $\{l_t\}_{t=0}^{\infty}$ and then defining a coupled new sequence $\{\bar{l}_t\}_{t=0}^{\infty}$ that is larger (and similarly smaller) than l_t and has updates with i.i.d increments.

We index the set of actions and let $\mathcal{A} = \{1, \dots, m\}$ denote the set of all actions including "no purchase." First, note that using Assumptions 1 and 2, both $Z(a \mid l)$ and $\mathbb{P}_{\theta,\zeta}[a \mid l, Q]$ are continuous, where $Z(a \mid l)$ is defined in (A-1) denoting the likelihood ratio of action a with belief $q = \frac{l}{l+1}$. Since $q_t \to 0$ almost surely (equivalently $l_t \to 0$ almost surely), for any ϵ we can choose N such that $l_t \leq \epsilon$ for $t \geq N$. For any $a \in \mathcal{A}$ and ϵ, we define $\epsilon_a = \arg\max_{l \in [0,\epsilon]} Z(a \mid l)$, denoting the likelihood ratio in $[0, \epsilon]$ that results in the highest belief when taking action a. Using this definition, we almost surely have

$$Z(a \mid l_t) \leq Z(a \mid l = \epsilon_a), \quad \forall a \in \mathcal{A}, t \geq N. \tag{A-8}$$

We can reindex the set of actions and without loss of generality we suppose

$$Z(a = 1 \mid l = \epsilon_1) \leq \cdots \leq Z(a = m \mid l = \epsilon_m). \tag{A-9}$$

Let $p_a = \min_{l \leq \epsilon} \mathbb{P}_{\theta,\zeta}[a \mid l, Q = 0]$ for all $a = 1, \dots, m - 1$, and define $p_m = 1 - \sum_{a=1}^{m-1} p_a$, so that $\sum_{a \in \mathcal{A}} p_a = 1$. Note that

$$\sum_{a=1}^{m-1} p_a \leq \min_{l \leq \epsilon} \left\{ \sum_{a=1}^{m-1} \mathbb{P}_{\theta,\zeta}[a \mid l, Q = 0] \right\} \leq 1.$$

Therefore, $p_m = 1 - \sum_{a=1}^{m-1} p_a \geq 0$ and $p_1, \dots, p_m$ form a probability mass function. With this choice of $p_1, \dots, p_m$, we have

$$\lim_{\epsilon \to 0} \sum_{a \in \mathcal{A}} p_a \log\bigl(Z(a \mid l = \epsilon_a)\bigr) \overset{\text{(a)}}{=} \sum_{a \in \mathcal{A}} \mathbb{P}[a \mid Q = 0, l = 0] \log\left(\frac{\mathbb{P}[a \mid Q = 1, l = 0]}{\mathbb{P}[a \mid Q = 0, l = 0]} \right) \overset{\text{(b)}}{<} 0,$$

where (a) follows from the continuity of $Z(a \mid l)$ and $\mathbb{P}_{\theta,\zeta}[a \mid l, Q]$ and (b) follows from the fact that for $a = \overline{K}$, using Assumption 1, $\underline{\zeta} + \bar{\theta} - p + 1 < \lambda_{-\underline{K}} \leq \lambda_{\overline{K}}$, therefore, $\mathbb{P}[a = \overline{K} \mid Q = 1, l = 0] > \mathbb{P}[a = \overline{K} \mid Q = 0, l = 0]$ and the fact that KL divergence is strictly positive if the two distributions are not identical. Therefore, for small enough ϵ we have

$$\sum_{a \in \mathcal{A}} p_a \log\bigl(Z(a \mid l = \epsilon_a)\bigr) < 0, \tag{A-10}$$

and from the choice of $p_1, \dots, p_m$ we have

$$\sum_{i=1}^{a} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0] \geq \sum_{i=1}^{a} p_i, \quad \forall a \in \mathcal{A}, \text{for } t \geq N. \tag{A-11}$$

Recall that $l_{t+1} = l_t Z(a_t \mid l_t)$ where the action at time t, a_t, is equal to $a \in \mathcal{A}$ with probability $\mathbb{P}_{\theta,\zeta}[a \mid l_t, Q = 0]$. Alternatively, we can define $\{l_t\}$ as follows. Let Σ denote a uniform distribution with support $[0, 1]$. At any time t, we draw an independent sample $\sigma_t \sim \Sigma$. We then find action a for which $\sigma_t \in (\sum_{i=1}^{a-1} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0], \sum_{i=1}^{a} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0]]$, and set a_t equal to this action a (for $a = 1$, we define $\sum_{i=1}^{a-1} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0] = 0$). Finally, we update $l_{t+1} = l_t Z(a \mid l_t)$.

Note that since $\sigma_t \in [0, 1]$, there exists $a \in \mathcal{A}$ such that $\sigma_t \in (\sum_{i=1}^{a-1} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0], \sum_{i=1}^{a} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0]]$. By definition, the sequence $\{l_t\}$ defined above evolves as the likelihood ratio of the public belief. This is because we have

$$\mathbb{P}_{\Sigma}\left[\sigma_t \in \left(\sum_{i=1}^{a-1} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q], \sum_{i=1}^{a} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q]\right]\right] = \mathbb{P}_{\theta,\zeta}[a \mid l_t, Q].$$

We next define a sequence $\bar{l}_t$ for $t \geq N$ and the corresponding sequence of actions $\bar{a}_t$ that couples with sequence $\{l_t\}$. We initialize this sequence at $t = N$, by defining $\bar{l}_N = l_N$. At any time $t > N$, let σ_t denote the drawn random variable in sequence l_t. We then find an action $a \in \mathcal{A}$ such that $\sigma_t \in (\sum_{i=1}^{a-1} p_i, \sum_{i=1}^{a} p_i]$ and set $\bar{a}_t = a$ (again, for $a = 1$, we define $\sum_{i=1}^{a-1} p_i = 0$). Finally, we update $\bar{l}_{t+1} = \bar{l}_t Z(a \mid l = \epsilon_a)$.

Since $\sigma_t \in [0, 1]$, there exists $a \in \mathcal{A}$ such that $\sigma_t \in (\sum_{i=1}^{a-1} p_i, \sum_{i=1}^{a} p_i]$. Since $\{\sigma_t\}$ is an i.i.d. sequence, the strong law of large numbers implies that, almost surely,

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t} \log \bar{l}_t &= \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t-1} \sum_{a=1}^{m} \mathbf{1}\left\{\sigma_s \in \left(\sum_{i=1}^{a-1} p_i, \sum_{i=1}^{a} p_i\right]\right\}\bigl(\log\bigl(Z(a \mid l = \epsilon_a)\bigr)\bigr) \\ &= \sum_{a \in \mathcal{A}} p_a \log\bigl(Z(a \mid l = \epsilon_a)\bigr).\end{aligned}$$

Combined with (A-10), this proves that $\bar{l}_t$ converges to zero at the rate $-\sum_{a \in \mathcal{A}} p_a \log Z(a \mid l = \epsilon_a)$.

We next prove that $\bar{l}_t \geq l_t$ for $t \geq N$. We establish this inequality by induction on t. It evidently holds for $t = N$. Suppose at time t, we have $\sigma_t \in (\sum_{i=1}^{a-1} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0], \sum_{i=1}^{a} \mathbb{P}_{\theta,\zeta}[i \mid l_t, Q = 0]]$, that is, $a_t = a$. Using (A-11) and assuming $a_t = a$ leads to $\bar{a}_t \geq a$. Therefore, we have

$$\bar{l}_{t+1} = Z(\bar{a}_t \mid l = \epsilon_{\bar{a}_t})\bar{l}_t \overset{\text{(a)}}{\geq} Z(\bar{a}_t \mid l = \epsilon_{\bar{a}_t}) l_t \overset{\text{(b)}}{\geq} Z(a \mid l = \epsilon_a) l_t \overset{\text{(c)}}{\geq} Z(a \mid l_t) l_t \overset{\text{(d)}}{=} l_{t+1},$$

where (a) follows from induction hypothesis, (b) follows from $\bar{a}_t \geq a$ and (A-9), (c) follows from (A-8), and (d) follows from the definition of sequence $\{l_t\}$. This establishes that almost surely $\lim_{t\to\infty} \frac{1}{t} \log l_t \leq \sum_{a\in\mathcal{A}} p_a \log(Z(a \mid l = \epsilon_a))$. Since this inequality holds for all small enough ϵ, letting $\epsilon \to 0$ (and consequently $\epsilon_a \to 0$ and $p_a \to \mathbb{P}[a \mid l = 0, Q = 0]$ for all $a \in \mathcal{A}$) leads to

$$\begin{aligned}\lim_{t\to\infty} \frac{1}{t} \log l_t &\leq \lim_{\epsilon\to 0} \sum_{a\in\mathcal{A}} p_a \log\bigl(Z(a \mid l = \epsilon_a)\bigr) = \sum_{a\in\mathcal{A}} \mathbb{P}_{\theta,\zeta}[a \mid l = 0, Q = 0] \log\bigl(Z(a \mid l = 0)\bigr) \\ &= -D\bigl(\boldsymbol{\pi}(F_{\theta,\zeta}, Q = 0, q = 0) \parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q = 1, q = 0)\bigr). \end{aligned} \tag{A-12}$$

To prove the other direction of inequality (A-12), we define $\tilde{\epsilon}_a = \arg\max_{l\in[0,\epsilon]} Z(a \mid l)$, re-index the set of actions and suppose without loss of generality that $Z(a = 1 \mid l = \tilde{\epsilon}_1) \geq \cdots \geq Z(a = m \mid l = \tilde{\epsilon}_m)$. We also define $\tilde{p}_a = \min_{l\leq\epsilon} \mathbb{P}_{\theta,\zeta}[a \mid l, Q = 0]$, for all $a = 1, \ldots, m-1$, and $\tilde{p}_m = 1 - \sum_{a=1}^{m-1} \tilde{p}_a$.

Using this new ordering and by replacing ϵ_a and p_a with $\tilde{\epsilon}_a$ and $\tilde{p}_a$ respectively, a similar procedure used in defining sequence $\{\bar{l}_t\}_{t=N}^{\infty}$ defines coupled sequences $\{l_t\}_{t=N}^{\infty}$ and $\{\underline{l}_t\}_{t=N}^{\infty}$ such that $l_t \geq \underline{l}_t$ for all $t \geq N$ almost surely and $\lim_{t\to\infty} \frac{1}{t} \log \underline{l}_t = \sum_{a\in\mathcal{A}} \tilde{p}_a \log(Z(a \mid l = \tilde{\epsilon}_a))$. Therefore, by letting $\epsilon \to 0$ (and consequently $\tilde{\epsilon}_a \to 0$ and $\tilde{p}_a \to \mathbb{P}[a \mid l = 0, Q = 0]$), we almost surely have

$$\lim_{t\to\infty} \frac{1}{t} \log l_t \geq -D\bigl(\boldsymbol{\pi}(F_{\theta,\zeta}, Q = 0, q = 0) \parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q = 1, q = 0)\bigr). \tag{A-13}$$

Combining (A-12) and (A-13) leads to

$$\lim_{t\to\infty} \frac{1}{t} \log l_t = -D\bigl(\boldsymbol{\pi}(F_{\theta,\zeta}, Q = 0, q = 0) \parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q = 1, q = 0)\bigr).$$

Finally, we have

$$\lim_{t\to\infty} \frac{1}{t} \log q_t = \lim_{t\to\infty} \frac{1}{t} \log l_t = -D\bigl(\boldsymbol{\pi}(F_{\theta,\zeta}, Q = 0, q = 0) \parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q = 1, q = 0)\bigr),$$

which completes the proof. *Q.E.D.*

PROOF OF THEOREM 3: A similar argument to the proof of Theorem 1 shows that under Assumptions 1 and 2 for any belief $q \in [0, 1]$ and any quality $Q \in \{0, 1\}$ all actions have nonzero probability. We first derive some relations that will be useful in the rest of the proof. A summary statistic $\mathbf{S}_\tau$ is a vector $(k_1/\tau, \ldots, k_m/\tau)$ such that $\sum_{j=1}^{m} k_j = \tau$ and $k_j \geq 0$, $j \in [m]$. Any summary statistic can be equivalently expressed as $(k_1, \ldots, k_m)$ at time τ, and we define

$$\begin{aligned} p_1(k_1, \ldots, k_m, \tau) &= \mathbb{P}\bigl[\mathbf{S}_\tau = (k_1, \ldots, k_m) \mid Q = 1\bigr], \\ p_0(k_1, \ldots, k_m, \tau) &= \mathbb{P}\bigl[\mathbf{S}_\tau = (k_1, \ldots, k_m) \mid Q = 0\bigr]. \end{aligned}$$

We next determine the stochastic evolution of $p_1(k_1, \ldots, k_m, \tau)$ and $p_0(k_1, \ldots, k_m, \tau)$. Note that both sequences $\{p_1(\cdot, \tau)\}$ and $\{p_0(\cdot, \tau)\}$ depend on $q(\cdot, \tau)$, which is in turn determined by the sequences themselves. The next lemma shows that despite this dependence and the fact that the behavior and number of customers who have taken actions that are not in T are not observed, Bayesian updating implies that these probabilities satisfy an intuitive recursion under our assumption of uniform priors over the number of past people who joined the platform.

To simplify the notation, let us use $\pi_1(i, q)$ (resp., $\pi_0(i, q)$) to denote the probability of $\mathbf{Y}_\tau = \mathbf{e}_i$ given belief q and $Q = 1$ (resp., $Q = 0$). Therefore,

$$\begin{aligned} \pi_1(i, q) &= \pi(i; F_{\theta,\zeta}, Q = 1, q), \quad i \in [m], \\ \pi_0(i, q) &= \pi(i; F_{\theta,\zeta}, Q = 0, q), \quad i \in [m], \end{aligned} \tag{A-14}$$

where $[m]$ denotes the set $\{1, \ldots, m\}$. Note that since $T_1, \ldots, T_m$ form a partition of T, for any q, we have $\sum_{i=1}^m \pi_1(i, q) = \sum_{i=1}^m \pi_0(i, q) = 1$. Moreover, for any subset $\mathcal{R} \subseteq \{1, \ldots, m\}$, we write

$$\pi_1(\mathcal{R}, q) = \sum_{i \in \mathcal{R}} \pi_1(i, q), \qquad \pi_0(\mathcal{R}, q) = \sum_{i \in \mathcal{R}} \pi_0(i, q).$$

LEMMA A-1: *The sequences* $\{p_1(k_1, \ldots, k_m, \tau)\}$ *and* $\{p_0(k_1, \ldots, k_m, \tau)\}$ *satisfy*

$$\begin{aligned} & p_1(k_1, \ldots, k_m, \tau) \\ &\quad = \sum_{j=1}^m p_1(k_1, \ldots, k_{j-1}, k_j - 1, k_{j+1}, \ldots, k_m, \tau - 1) \\ &\qquad \times \pi_1\big(j, q(k_1, \ldots, k_{j-1}, k_j - 1, k_{j+1}, \ldots, k_m, \tau - 1)\big), \\ & p_0(k_1, \ldots, k_m, \tau) \\ &\quad = \sum_{j=1}^m p_0(k_1, \ldots, k_{j-1}, k_j - 1, k_{j+1}, \ldots, k_m, \tau - 1) \\ &\qquad \times \pi_0\big(j, q(k_1, \ldots, k_{j-1}, k_j - 1, k_{j+1}, \ldots, k_m, \tau - 1)\big), \\ & q(k_1, \ldots, k_m, \tau) = \frac{p_1(k_1, \ldots, k_m, \tau)}{p_1(k_1, \ldots, k_m, \tau) + p_0(k_1, \ldots, k_m, \tau)}. \end{aligned} \tag{A-15}$$

PROOF: For $\tau = 1$, the belief without any observation is $1/2$. Therefore, we have $p_1(\mathbf{e}_i, 1) = \pi_1(i, \frac{1}{2})$ and $p_0(\mathbf{e}_i, 1) = \pi_0(i, \frac{1}{2})$ for all $i \in [m]$. We let random variable $h_{\tau-1:\tau}$ denote the history of actions in between $\tau - 1$-th and τth actions in T. Given customers have uniform prior on the number of people who joined the platform denoted by C, and the fact that all customers in this interval observe the same history, and hence form the same belief, we can write

$$\mathbb{P}[h_{\tau-1:\tau} \mid \mathbf{S}_{\tau-1}, Q] = \sum_{c=0}^{\infty} \mathbb{P}[h_{\tau-1:\tau} \mid C = c, \mathbf{S}_{\tau-1}, Q] = \sum_{c=0}^{\infty} \mathbb{P}[a \notin T \mid \mathbf{S}_{\tau-1}, Q]^c. \tag{A-16}$$

Using (A-16), we next show the update rule for $\tau \geq 2$. We have

$$\begin{aligned} & p_1(k_1, \ldots, k_m, \tau) \\ &\quad = \mathbb{P}\big[\mathbf{S}_\tau(1) = k_1, \ldots, \mathbf{S}_\tau(m) = k_m \mid Q = 1\big] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{m} \sum_{h_{\tau-1:\tau}} \mathbb{P}\big[\mathbf{S}_{\tau-1} = (k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m) \mid Q = 1\big] \\
&\quad \times \mathbb{P}\big[h_{\tau-1:\tau} \mid Q = 1, \mathbf{S}_{\tau-1} = (k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m)\big] \\
&\quad \times \mathbb{P}\big[a_\tau \in T_j \mid h_{\tau-1:\tau}, Q = 1, \mathbf{S}_{\tau-1} = (k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m)\big] \\
&= \sum_{j=1}^{m} \mathbb{P}\big[\mathbf{S}_{\tau-1} = (k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m) \mid Q = 1\big] \\
&\quad \times \sum_{c=0}^{\infty} \mathbb{P}\big[a \notin T \mid q(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m), Q = 1\big]^c \\
&\quad \times \mathbb{P}\big[a \in T_j \mid q(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m), Q = 1\big] \\
&= \sum_{j=1}^{m} \mathbb{P}\big[\mathbf{S}_{\tau-1} = (k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m) \mid Q = 1\big] \\
&\quad \times \frac{\mathbb{P}\big[a \in T_j \mid q(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m), Q = 1\big]}{1 - \mathbb{P}\big[a \notin T \mid q(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m), Q = 1\big]} \\
&= \sum_{j=1}^{m} p_1(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m, \tau - 1) \\
&\quad \times \pi_1\big(j, q(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m, \tau - 1)\big).
\end{aligned}
$$

We can write a similar recursion for $p_0(k_1, \dots, k_m, \tau)$. The proof concludes by using Bayes' rule to find $q(k_1, \dots, k_m, \tau)$ as in (A-15). *Q.E.D.*

We now provide the proof of the theorem. We first prove Part 1: strict separation is sufficient for complete learning.

Part 1: We prove that strict separation is sufficient for complete learning in three steps. Under strict separation, there exists $\hat{\mathcal{R}} \subseteq [m]$ such that the range of $\sum_{i\in\hat{\mathcal{R}}} \pi_1(i, \cdot)$ and $\sum_{i\in\hat{\mathcal{R}}} \pi_0(i, \cdot)$ are separated. Without loss of generality, suppose the range of $\sum_{i\in\hat{\mathcal{R}}} \pi_1(i, \cdot)$ is above the range of $\sum_{i\in\hat{\mathcal{R}}} \pi_0(i, \cdot)$, that is, $\min_q \sum_{i\in\hat{\mathcal{R}}} \pi_1(i, q) > \max_q \sum_{i\in\hat{\mathcal{R}}} \pi_0(i, q)$. We let

$$
\pi_1 = \left(\min_q \sum_{i\in\hat{\mathcal{R}}} \pi_1(i, q)\right) - \frac{1}{4}\Delta_3, \quad \text{and} \quad \pi_0 = \left(\max_q \sum_{i\in\hat{\mathcal{R}}} \pi_0(i, q)\right) + \frac{1}{4}\Delta_3, \tag{A-17}
$$

where $\Delta_3 = (\min_q \sum_{i\in\hat{\mathcal{R}}} \pi_1(i, q)) - (\max_q \sum_{i\in\hat{\mathcal{R}}} \pi_0(i, q)) > 0$.

Step 1: Let $\mathbf{S}_\tau(\hat{\mathcal{R}}) = \sum_{i\in\hat{\mathcal{R}}} \mathbf{S}_\tau(i)$ denote the fraction of reviews in $\hat{\mathcal{R}}$ after τ reviews. Then

$$
\mathbf{S}_\tau(\hat{\mathcal{R}}) \succ \text{Binomial}(\pi_1, \tau) \quad \text{for } Q = 1 \quad \text{and} \quad \mathbf{S}_\tau(\hat{\mathcal{R}}) \prec \text{Binomial}(\pi_0, \tau) \quad \text{for } Q = 0.
$$

Equivalently, using the notation $k_{\hat{\mathcal{R}}} = \sum_{i\in\hat{\mathcal{R}}} k_i$, we have

$$\begin{aligned}
S_1(l,\tau) &= \mathbb{P}\big[\mathbf{S}_\tau(\hat{\mathcal{R}}) \le l \mid Q=1\big] = \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=1}^{l} p_1(k_1,\dots,k_m,\tau) \\
&\le \sum_{j=0}^{l} \binom{\tau}{j} \pi_1^j (1-\pi_1)^{\tau-j}, \\
S_0(l,\tau) &= \mathbb{P}\big[\mathbf{S}_\tau(\hat{\mathcal{R}}) \le l \mid Q=0\big] = \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=1}^{l} p_0(k_1,\dots,k_m,\tau) \\
&\ge \sum_{j=0}^{l} \binom{\tau}{j} \pi_0^j (1-\pi_0)^{\tau-j},
\end{aligned} \tag{A-18}$$

where the inequalities are strict for some l in $\{0,\dots,\tau\}$.

PROOF OF STEP 1: We prove the claim for $Q=1$ (the proof for $Q=0$ is identical). The proof idea is as follows. Given $Q=1$, consider the random process $\mathbf{S}_\tau(\hat{\mathcal{R}})$. Lemma A-1 proves that the distribution of $\mathbf{S}_\tau(\hat{\mathcal{R}})$ satisfies a recursive relationship. We consider another random process, whose distribution satisfies a similar recursion with the difference that $\pi_1(\hat{\mathcal{R}}, q_\tau)$ is replaced by π_1, its minimum over all q. We prove that the distribution of this process at τ is the same as the distribution of Binomial(π_1,τ). Moreover, because this process is defined recursively using the minimum of $\pi_1(\hat{\mathcal{R}}, q)$ over q, it follows that $\mathbf{S}_\tau$ with $Q=1$ first-order stochastically dominates Binomial(π_1,τ).

We next present the formal proof. First, note that from the definition in (A-18), we have

$$\begin{aligned}
&S_1(l,\tau) - S_1(l,\tau-1) \\
&\quad= \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=1}^{l} p_1(k_1,\dots,k_m,\tau) - \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=1}^{l} p_1(k_1,\dots,k_m,\tau-1) \\
&\quad= \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=1}^{l} \sum_{j=1}^{m} p_1(k_1,\dots,k_{j-1},k_j-1,k_{j+1},\dots,k_m,\tau-1) \\
&\qquad \times \pi_1\big(j, q(k_1,\dots,k_{j-1},k_j-1,k_{j+1},\dots,k_m,\tau-1)\big) \\
&\qquad - \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=1}^{l} p_1(k_1,\dots,k_m,\tau-1) \\
&\quad= \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=1}^{l} p_1(k_1,\dots,k_m,\tau-1) \sum_{j=1}^{m} \pi_1\big(j, q(k_1,\dots,k_m,\tau-1)\big) \\
&\qquad - \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=l} \sum_{i\in\hat{\mathcal{R}}} p_1(k_1,\dots,k_m,\tau-1)\pi_1\big(i, q(k_1,\dots,k_m,\tau-1)\big)
\end{aligned}$$

$$
\begin{aligned}
&-\sum_{k_{[m]\setminus\hat{\mathcal{R}}}}\sum_{k_{\hat{\mathcal{R}}}=1}^{l} p_1(k_1,\dots,k_m,\tau-1) \\
&= -\sum_{k_{[m]\setminus\hat{\mathcal{R}}}}\sum_{k_{\hat{\mathcal{R}}}=l} p_1(k_1,\dots,k_m,\tau-1)\pi_1\big(\hat{\mathcal{R}},q(k_1,\dots,k_m,\tau-1)\big),
\end{aligned}
\tag{A-19}
$$

where the second equality follows from Lemma A-1 and the last equality follows from the fact that $\sum_{i=1}^m \pi_1(i,q)=1$. We define the sequence $\{\tilde{S}_1(l,\tau)\}_{l,\tau}$ recursively as

$$
\tilde{S}_1(l,\tau)=\tilde{S}_1(l-1,\tau-1)\pi_1+\tilde{S}_1(l,\tau-1)(1-\pi_1),\quad 0\le l\le\tau,\tau\ge 0. \tag{A-20}
$$

We next prove that $\tilde{S}_1(l,\tau)\ge S_1(l,\tau)$ (with one strict inequality) by induction on τ. For $\tau=1$, we have $\tilde{S}_1(0,1)=1-\pi_1>1-\pi_1(i,\frac{1}{2})=S_1(0,1)$ and $\tilde{S}_1(1,1)=1=S_1(1,1)$. We now suppose that the induction hypothesis holds for $\tau-1$ and prove it for τ. We have

$$
\begin{aligned}
\tilde{S}_1(l,\tau) &= \tilde{S}_1(l-1,\tau-1)\pi_1+\tilde{S}_1(l,\tau-1)(1-\pi_1) \\
&\ge S_1(l-1,\tau-1)\pi_1+S_1(l,\tau-1)(1-\pi_1) \\
&\ge S_1(l,\tau-1)-\pi_1\sum_{k_{[m]\setminus\hat{\mathcal{R}}}}\sum_{K_{\hat{\mathcal{R}}}=l} p_1(k_1,\dots,k_m,\tau-1) \\
&\ge S_1(l,\tau-1)-\sum_{k_{[m]\setminus\hat{\mathcal{R}}}}\sum_{K_{\hat{\mathcal{R}}}=l} p_1(k_1,\dots,k_m,\tau-1)\pi_1\big(\hat{\mathcal{R}},q(k_1,\dots,k_m,\tau-1)\big) \\
&= S_1(l,\tau),
\end{aligned}
\tag{A-21}
$$

where the first inequality follows from the induction hypothesis, the second inequality follows from the definition in (A-18), the third inequality follows from using $\pi_1<\min_q \pi_1(\hat{\mathcal{R}},q)$ and the last equality follows from using (A-19). Also, note that by induction hypothesis for some l, $\tilde{S}(l,\tau-1)>S(l,\tau-1)$ and, therefore, $\tilde{S}(l,\tau)>S(l,\tau)$ is strict. We next prove, again by induction on τ, that the recursive definition of $\{\tilde{S}_1(l,\tau)\}$ given in (A-20) leads to $\tilde{S}_1(l,\tau)=\sum_{j=0}^{l}\binom{\tau}{j}\pi_1^j(1-\pi_1)^{\tau-j}$.

This evidently holds for $\tau=1$ as we have $\tilde{S}_1(0,1)=1-\pi_1$ and $\tilde{S}_1(1,1)=1$. We now suppose it holds for $\tau-1$ and prove it for τ. We have

$$
\begin{aligned}
\tilde{S}_1(l,\tau) &= \tilde{S}_1(l-1,\tau-1)\pi_1+\tilde{S}_1(l,\tau-1)(1-\pi_1) \\
&= \pi_1\sum_{j=0}^{l-1}\binom{\tau-1}{j}\pi_1^j(1-\pi_1)^{\tau-1-j}+(1-\pi_1)\sum_{j=0}^{l}\binom{\tau-1}{j}\pi_1^j(1-\pi_1)^{\tau-1-j} \\
&= \sum_{j=0}^{l}\pi_1^j(1-\pi_1)^{\tau-j}\left(\binom{\tau-1}{j}+\binom{\tau-1}{j-1}\right)=\sum_{j=0}^{l}\pi_1^j(1-\pi_1)^{\tau-j}\binom{\tau}{j}.
\end{aligned}
\tag{A-22}
$$

Combining (A-21) and (A-22) completes the proof of Step 1.

Step 2: Using definitions of π_1 and π_2 in (A-17), we have $\pi_1 - \pi_0 = \frac{\Delta_3}{2}$. We let $\Delta_4 = \frac{\Delta_3}{2}$, which is strictly positive. For any $\epsilon > 0$, we have

$$\mathbb{P}[q_\tau > \epsilon \mid Q = 0] \le \frac{1}{\epsilon} e^{-\tau \frac{\Delta_4^2}{2}}, \quad \forall \tau \ge 1, \quad \text{and} \quad \mathbb{P}[1 - q_\tau > \epsilon \mid Q = 1] \le \frac{1}{\epsilon} e^{-\tau \frac{\Delta_4^2}{2}}, \quad \forall \tau \ge 1.$$

PROOF OF STEP 2: Using Step 1, for $\gamma = \frac{\pi_1 + \pi_0}{2}$, we have $\pi_1 - \frac{\Delta_4}{2} = \pi_0 + \frac{\Delta_4}{2} = \gamma$. This implies

$$S_1(\gamma\tau, \tau) \le \sum_{j=0}^{\gamma\tau} \binom{\tau}{j} \pi_1^j (1-\pi_1)^{\tau-j} = \mathbb{P}\left[\text{Binomial}(\pi_1, \tau) \le \left(\pi_1 - \frac{\Delta_4}{2}\right)\tau\right] \le e^{-\tau \frac{\Delta_4^2}{2}}, \quad \text{(A-23)}$$

where we used Chernoff–Hoeffding's inequality in establishing the last inequality. We also have

$$\begin{aligned} S_0(\gamma\tau, \tau) &\ge \sum_{j=0}^{\gamma\tau} \binom{\tau}{j} \pi_0^j (1-\pi_0)^{\tau-j} = \mathbb{P}\left[\text{Binomial}(\pi_0, \tau) \le \left(\pi_0 + \frac{\Delta_4}{2}\right)\tau\right] \\ &\ge 1 - e^{-\tau \frac{\Delta_4^2}{2}}. \end{aligned} \quad \text{(A-24)}$$

We next prove the claim for $Q = 0$ (the proof for $Q = 1$ is analogous). We have

$$\begin{aligned} \mathbb{P}[q_\tau > \epsilon \mid Q = 0] &= \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} p_0(k_1, \dots, k_m, \tau) \mathbf{1}\{q(k_1, \dots, k_m, \tau) > \epsilon\} \\ &= \sum_{l=0}^{\tau\gamma} \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=l} p_0(k_1, \dots, k_m, \tau) \mathbf{1}\left\{\frac{q(k_1, \dots, k_m, \tau)}{1 - q(k_1, \dots, k_m, \tau)} > \frac{\epsilon}{1-\epsilon}\right\} \\ &\quad + \sum_{l=\tau\gamma}^{\tau} \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=l} p_0(k_1, \dots, k_m, \tau) \mathbf{1}\left\{\frac{q(k_1, \dots, k_m, \tau)}{1 - q(k_1, \dots, k_m, \tau)} > \frac{\epsilon}{1-\epsilon}\right\} \\ &\overset{\text{(a)}}{\le} \sum_{l=0}^{\tau\gamma} \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=l} p_0(k_1, \dots, k_m, \tau) \mathbf{1}\left\{\frac{q(k_1, \dots, k_m, \tau)}{1 - q(k_1, \dots, k_m, \tau)} > \frac{\epsilon}{1-\epsilon}\right\} \\ &\quad + \big(1 - S_0(\gamma\tau, \tau)\big) \\ &\overset{\text{(b)}}{\le} \sum_{l=0}^{\tau\gamma} \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=l} p_0(k_1, \dots, k_m, \tau) \mathbf{1}\left\{\frac{p_1(k_1, \dots, k_m, \tau)}{p_0(k_1, \dots, k_m, \tau)} > \frac{\epsilon}{1-\epsilon}\right\} + e^{-\tau \frac{\Delta_4^2}{2}} \\ &\overset{\text{(c)}}{\le} \sum_{l=0}^{\tau\gamma} \sum_{k_{[m]\setminus\hat{\mathcal{R}}}} \sum_{k_{\hat{\mathcal{R}}}=l} \frac{1-\epsilon}{\epsilon} p_1(k_1, \dots, k_m, \tau) + e^{-\tau \frac{\Delta_4^2}{2}} \le \frac{1-\epsilon}{\epsilon} e^{-\tau \frac{\Delta_4^2}{2}} + e^{-\tau \frac{\Delta_4^2}{2}} \\ &= \frac{1}{\epsilon} e^{-\tau \frac{\Delta_4^2}{2}}, \end{aligned} \quad \text{(A-25)}$$

where (a) simply follows from the fact that indicator function is less than or equal to one, (b) follows from (A-24), and (c) follows from (A-23). This completes the proof of Step 2.

Step 3: We have $q_\tau \to Q$ almost surely.

PROOF OF STEP 3: We present the proof for $Q = 0$ (the proof for $Q = 1$ is identical). The idea is to use the Borell–Cantelli lemma together with the exponential tail bound obtained in Step 2 (see Etemadi (1981) and Korchevsky and Petrov (2010) for similar arguments). For any $m, \tau \in \mathbb{N}$, we let $\mathcal{A}_\tau^{(\frac{1}{m})} = \{q_\tau \geq \frac{1}{m}\}$. From Step 2, for any $m \in \mathbb{N}$ we have $\sum_{\tau=1}^{\infty} \mathbb{P}[\mathcal{A}_\tau^{(\frac{1}{m})}] \leq m \sum_{\tau=1}^{\infty} e^{-\tau \frac{\Delta_4^2}{2}} < \infty$. Therefore, the Borell–Cantelli lemma implies $\mathbb{P}[\text{i.o. } A_\tau^{(1/m)}] = \mathbb{P}[\bigcap_{\tau\in\mathbb{N}} \bigcup_{k\geq\tau} A_k^{(1/m)}] = 0$. Therefore,

$$\mathbb{P}\Big[\Big(\lim_{\tau\to\infty} q_\tau = 0\Big)^C\Big] = \mathbb{P}\Big[\bigcup_{m\in\mathbb{N}} \bigcap_{\tau\in\mathbb{N}} \bigcup_{k\geq\tau} A_k^{(1/m)}\Big] \leq \sum_{m=1}^{\infty} \mathbb{P}\big[\text{i.o. } A_\tau^{(1/m)}\big] = 0,$$

where we used a standard union bound for the inequality. This completes the proof of Part 1.

Part 2: Here, we first establish that, with negative selection and $m = 2$ reviews, weak separation is sufficient for complete learning to happen. We use the same line of argument as part 1 of the theorem to bound $\mathbb{P}[q_\tau > \epsilon \mid Q = 0]$ similar to inequality (A-25). For a given ϵ, we let $\gamma = \frac{\pi_0(2,0)+\pi_0(2,\epsilon)}{2}$. We can write

$$\begin{aligned}
\mathbb{P}[q_\tau > \epsilon \mid Q = 0] &= \sum_{k=0}^{\tau} p_0(\tau - k, k, \tau)\mathbf{1}\big\{q(\tau - k, k, \tau) > \epsilon\big\} \\
&\overset{(a)}{\leq} \sum_{k=0}^{\gamma\tau} p_1(\tau - k, k, \tau)\frac{1-\epsilon}{\epsilon} + \sum_{k=\gamma\tau}^{\tau} p_0(\tau - k, k, \tau)\mathbf{1}\big\{q(\tau - k, k, \tau) > \epsilon\big\} \\
&\overset{(b)}{\leq} \frac{1-\epsilon}{\epsilon}\mathbb{P}\big[\text{Binomial}\big(\pi_0(2, 0), \tau\big) \leq \gamma\tau\big] \\
&\quad + \sum_{k=\gamma\tau}^{\tau} p_0(\tau - k, k, \tau)\mathbf{1}\big\{q(\tau - k, k, \tau) > \epsilon\big\} \\
&\overset{(c)}{\leq} \frac{1-\epsilon}{\epsilon}\mathbb{P}\big[\text{Binomial}\big(\pi_0(2, 0), \tau\big) \leq \gamma\tau\big] \\
&\quad + \mathbb{P}\big[\text{Binomial}\big(\pi_0(2, \epsilon), \tau\big) \geq \gamma\tau\big] \\
&\overset{(d)}{\leq} \frac{1}{\epsilon} e^{-\tau \frac{(\pi_0(2,0)-\pi_0(2,\epsilon))^2}{2}},
\end{aligned}$$

where (a) follows a similar argument to that of (A-25), (b) follows from the fact that with negative selection $\min_q \pi_1(2, q) = \pi_1(2, 1) = \pi_0(2, 0)$, (c) follows from the fact that for $q_\tau > \epsilon$ the negative selection implies $\pi_0(2, q_\tau) \leq \pi_0(2, \epsilon)$, and (d) follows from Chernoff–Hoeffding's inequality. We use a similar argument to that of part 1 to complete the proof. In particular, for any $m, \tau \in \mathbb{N}$, we let $\mathcal{A}_\tau^{(\frac{1}{m})} = \{q_\tau \geq \frac{1}{m}\}$. Using the above inequality, for any $m \in \mathbb{N}$ we have $\sum_{\tau=1}^{\infty} \mathbb{P}[\mathcal{A}_\tau^{(\frac{1}{m})}] \leq m \sum_{\tau=1}^{\infty} e^{-\tau \frac{(\pi_0(2,0)-\pi_0(2,1/m))^2}{2}} < \infty$, and similar to the proof of Part 1, the Borell–Cantelli lemma implies $\mathbb{P}[(\lim_{\tau\to\infty} q_\tau = 0)^C] = 0$.

We next prove that the weak separation condition is also necessary for complete learning by using the following claim.

CLAIM 1: *If a rating system with $m = 2$ has negative selection, then for any τ, we have*

$$\left(\mathbf{S}_\tau(\{2\}) \mid Q = 1\right) \succ \left(\mathbf{S}_\tau(\{2\}) \mid Q = 0\right),$$

where recall that $\mathbf{S}_\tau(\{2\})$ is the number of top reviews at time τ.

PROOF OF CLAIM 1: Let

$$\pi_1(2, q) = \pi(i = 2; F_{\theta,\zeta}, Q = 1, q), \qquad \pi_0(2, q) = \pi(i = 2; F_{\theta,\zeta}, Q = 0, q),$$

where the subscript of $\pi_Q(\mathcal{S}, q)$ for $Q \in \{0, 1\}$ indicates true quality, on which this expression conditions. We also define

$$S_1(l, \tau) = \mathbb{P}\left[\mathbf{S}_\tau(\{2\}) \le l \mid Q = 1\right], \qquad S_0(l, \tau) = \mathbb{P}\left[\mathbf{S}_\tau(\{2\}) \le l \mid Q = 0\right].$$

We next prove that, for all l, we have $S_1(l, \tau) \le S_0(l, \tau)$ and the inequality is strict for some l. The proof is by induction on τ. The base of induction (for $\tau = 1$) holds because $S_1(0, 1) = 1 - \pi_1(2, q = \frac{1}{2}) < 1 - \pi_0(2, q = \frac{1}{2}) = S_0(0, 1)$ and $S_1(1, 1) = S_0(1, 1) = 1$. This inequality holds since, for the same belief, the probability of a top review with quality $Q = 1$ is strictly greater than the same probability with quality $Q = 0$ (this follows from Assumption 1 and a similar argument to (A-2) in the proof of Theorem 1). We next show that if this inequality holds for $\tau - 1$, then it holds for τ as well. We write

$$\begin{aligned}
S_1(l, \tau) &= S_1(l, \tau - 1) - p_1(\tau - 1 - l, l, \tau - 1)\pi_1\big(2, q(\tau - 1 - l, l, \tau - 1)\big) \\
&\overset{\text{(a)}}{=} S_1(l, \tau - 1)\big(1 - \pi_1\big(2, q(\tau - 1 - l, l, \tau - 1)\big)\big) \\
&\quad + S_1(l - 1, \tau - 1)\pi_1\big(2, q(\tau - 1 - l, l, \tau - 1)\big) \\
&\overset{\text{(b)}}{\le} S_0(l, \tau - 1)\big(1 - \pi_1\big(2, q(\tau - 1 - l, l, \tau - 1)\big)\big) \\
&\quad + S_0(l - 1, \tau - 1)\pi_1\big(2, q(\tau - 1 - l, l, \tau - 1)\big) \\
&\overset{\text{(c)}}{=} S_0(l, \tau - 1) - p_0(\tau - 1 - l, l, \tau - 1)\pi_1\big(2, q(\tau - 1 - l, l, \tau - 1)\big) \\
&\overset{\text{(d)}}{\le} S_0(l, \tau - 1) - p_0(\tau - 1 - l, l, \tau - 1)\pi_0\big(2, q(\tau - 1 - l, l, \tau - 1)\big) = S_0(l, \tau),
\end{aligned}$$

where (a) follows from $p_1(\tau - 1 - l, l, \tau - 1) = S_1(l, \tau - 1) - S_1(l - 1, \tau - 1)$, (b) follows from the induction hypothesis, (c) follows from a similar argument to that of (a), and (d) follows from the fact that $\pi_1(2, q) \ge \pi_0(2, q)$ for all $q \in [0, 1]$.

Also, note that by induction hypothesis, there exists $l \in \{0, \dots, \tau\}$ for which we have $S_1(l, \tau - 1) < S_0(l, \tau - 1)$. This establishes that the inequality for τ is strict for some l.

We now complete the proof of Part 2. Since there is complete learning, when $Q = 0$ we have $\lim_{\tau\to\infty} \frac{\mathbf{S}_\tau(\{2\})}{\tau} = \pi_0(2, 0)$ almost surely and when $Q = 1$ we have $\lim_{\tau\to\infty} \frac{\mathbf{S}_\tau(\{2\})}{\tau} = \pi_1(2, 1)$ almost surely. From Claim 1, $\mathbf{S}_\tau(\{2\})$ when $Q = 1$ first-order stochastically dominates $\mathbf{S}_\tau(\{2\})$ when $Q = 0$, establishing that $\pi_1(2, 1) \ge \pi_0(2, 0)$. We next prove that this inequality implies weak separation with set $\mathcal{S} = \{2\}$. We can write

$$\min_{q\in[0,1]} \pi_1(2, q) \overset{\text{(a)}}{=} \pi_1(2, 1) \ge \pi_0(2, 0) \overset{\text{(b)}}{=} \max_{q\in[0,1]} \pi_0(2, q),$$

where (a) and (b) follow from negative selection, completing the proof of Part 2 of the theorem. *Q.E.D.*

PROOF OF THEOREM 4: We prove the theorem for $Q=0$ (the proof for $Q=1$ is identical). First, recall that under Assumptions 1 and 2 and the strict separation condition, Theorem 3 establishes that we have $q_\tau \to Q$ almost surely. Moreover, a similar argument to the proof of Theorem 1 implies that under Assumptions 1 and 2 for any belief $q \in [0,1]$ and any quality $Q \in \{0,1\}$ all actions have nonzero probability and the probability of each action is continuous in the belief. As in the proof of Theorem 3, we let $\pi_Q(i,q) = \pi(i; F_{\theta,\zeta}, Q, q)$ for $Q \in \{0,1\}$ and $i \in [m]$.

When $Q=0$, complete learning implies $q_\tau \to 0$ a.s. and thus $\frac{S_\tau(i)}{\tau} \to \pi_0(i,0)$ a.s. for all $i \in [m]$. Analogously, when $Q=1$, we have that $q_\tau \to 1$ almost surely, and thus $\frac{S_\tau(i)}{\tau} \to \pi_1(i,1)$ almost surely for all $i \in [m]$. Therefore, for any ϵ there exists N such that for $\tau \geq N$ we have

$$q_\tau \leq \epsilon \quad \text{and} \quad \left|\frac{S_\tau(i)}{\tau} - \pi_0(i,0)\right| \leq \epsilon, \quad \text{for all } i \in [m] \text{ given } Q=0,$$
$$q_\tau \geq 1-\epsilon \quad \text{and} \quad \left|\frac{S_\tau(i)}{\tau} - \pi_1(i,1)\right| \leq \epsilon, \quad \text{for all } i \in [m] \text{ given } Q=1.$$

We let

$$\begin{aligned}\underline{\pi}_0(i,0) &= \min_{q \leq \epsilon} \mathbb{P}[a \in T_i \mid a \in T, q, Q=0], \\ \bar{\pi}_1(i,1) &= \max_{q \geq 1-\epsilon} \mathbb{P}[a \in T_i \mid a \in T, q, Q=1].\end{aligned} \tag{A-26}$$

With these definitions, for any time $\tau \geq N$ we let $N_\tau(i)$ denote the number of reviews i among τ reviews for $i \in [m]$. We let $\mathcal{H}_\tau$ be the set of possible histories that are consistent with $N_\tau(1), \ldots, N_\tau(m)$. We obtain

$$\begin{aligned}\log \frac{q_\tau}{1-q_\tau} &= \log\left(\frac{\sum_{h_\tau \in \mathcal{H}_\tau} \mathbb{P}[h_\tau \mid Q=1]}{\sum_{h_\tau \in \mathcal{H}_\tau} \mathbb{P}[h_\tau \mid Q=0]}\right) \\ &= \log\left(\frac{\sum_{h_\tau \in \mathcal{H}_\tau} \mathbb{P}[h_N \mid Q=1]\mathbb{P}[h_{N+1:\tau} \mid h_N, Q=1]}{\sum_{h_\tau \in \mathcal{H}_\tau} \mathbb{P}[h_N \mid Q=0]\mathbb{P}[h_{N+1:\tau} \mid h_N, Q=0]}\right) \\ &\overset{(a)}{\leq} \log\left(\frac{\left(\max_{h_N} \mathbb{P}[h_N \mid Q=1]\right) \sum_{h_\tau \in \mathcal{H}_\tau} \mathbb{P}[h_{N+1:\tau} \mid h_N, Q=1]}{\left(\min_{h_N} \mathbb{P}[h_N \mid Q=0]\right) \sum_{h_\tau \in \mathcal{H}_\tau} \mathbb{P}[h_{N+1:\tau} \mid h_N, Q=0]}\right)\end{aligned}$$

$$\stackrel{\text{(b)}}{\leq} \log\left(\frac{\left(\max\limits_{h_N}\mathbb{P}[h_N \mid Q=1]\right)|\mathcal{H}_\tau|\prod_{i=1}^{m}\left(\bar{\pi}_1(i,1)\right)^{\tilde{N}_\tau(i)}}{\left(\min\limits_{h_N}\mathbb{P}[h_N \mid Q=0]\right)|\mathcal{H}_\tau|\prod_{i=1}^{m}\left(\underline{\pi}_0(i,0)\right)^{\hat{N}_\tau(i)}}\right)$$

$$= \log\left(\frac{\left(\max\limits_{h_N}\mathbb{P}[h_N \mid Q=1]\right)\prod_{i=1}^{m}\left(\bar{\pi}_1(i,1)\right)^{\tilde{N}_\tau(i)}}{\left(\min\limits_{h_N}\mathbb{P}[h_N \mid Q=0]\right)\prod_{i=1}^{m}\left(\underline{\pi}_0(i,0)\right)^{\hat{N}_\tau(i)}}\right), \tag{A-27}$$

where $\tilde{N}_\tau(i)$ is the difference between $N_\tau(i)$ and the maximum number of i's over all histories up to time N that are consistent with set $\mathcal{H}_\tau$ and $\hat{N}_\tau(i)$ is the difference between $N_\tau(i)$ and the minimum number of i's over all histories up to time N that are consistent with set $\mathcal{H}_\tau$. In (A-27), (a) holds because we replaced the terms $\mathbb{P}[h_N \mid Q = 1]$ in the numerator by their maximum and the terms $\mathbb{P}[h_N \mid Q = 0]$ in the denominator by their minimum and (b) follows by using (A-26). Using Assumptions 1 and 2, all actions have a positive probability and, therefore, we have $0 < \bar{\pi}_1(i, 1) \leq 1$ and $0 < \underline{\pi}_0(i, 0) \leq 1$, which leads to

$$\frac{\left(\bar{\pi}_1(i,1)\right)^{\tilde{N}_\tau(i)}}{\left(\underline{\pi}_0(i,0)\right)^{\hat{N}_\tau(i)}} \leq \left(\frac{\bar{\pi}_1(i,1)}{\underline{\pi}_0(i,0)}\right)^{N_\tau(i)}\left(\frac{1}{\bar{\pi}_1(i,1)\underline{\pi}_0(i,0)}\right)^{N}, \quad i \in [m]. \tag{A-28}$$

Combining (A-27) and (A-28), we obtain

$$\log\frac{q_\tau}{1-q_\tau} \leq \log\left(\frac{\max\limits_{h_N}\mathbb{P}[h_N \mid Q=1]}{\min\limits_{h_N}\mathbb{P}[h_N \mid Q=0]}\prod_{i=1}^{m}\left(\frac{\bar{\pi}_1(i,1)}{\underline{\pi}_0(i,0)}\right)^{N_\tau(i)}\left(\frac{1}{\bar{\pi}_1(i,1)\underline{\pi}_0(i,0)}\right)^{N}\right).$$

Since $Q = 0$, we almost surely have $\frac{N_\tau(i)}{\tau} \to \pi_0(i, 0)$, which yields

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\frac{q_\tau}{1-q_\tau} \leq \lim_{\tau\to\infty}\frac{1}{\tau}\left(\log\left(\frac{\max\limits_{h_N}\mathbb{P}[h_N \mid Q=1]}{\min\limits_{h_N}\mathbb{P}[h_N \mid Q=0]}\right) + \log\left(\left(\frac{1}{\bar{\pi}_1(i,1)\underline{\pi}_0(i,0)}\right)^{N}\right)\right)$$
$$+ \lim_{\tau\to\infty}\frac{1}{\tau}\log\left(\prod_{i=1}^{m}\left(\frac{\bar{\pi}_1(i,1)}{\underline{\pi}_0(i,0)}\right)^{N_\tau(i)}\right)$$
$$= \lim_{\tau\to\infty}\sum_{i=1}^{m}\frac{N_\tau(i)}{\tau}\log\left(\frac{\bar{\pi}_1(i,1)}{\underline{\pi}_0(i,0)}\right) = \sum_{i=1}^{m}\pi_0(i,0)\log\left(\frac{\bar{\pi}_1(i,1)}{\underline{\pi}_0(i,0)}\right).$$

Since this inequality holds for any ϵ, by letting $\epsilon \to 0$ we have $\underline{\pi}_0(i, 0) \to \pi_0(i, 0)$ and $\bar{\pi}_1(i, 1) \to \pi_1(i, 1)$, which leads to

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\frac{q_\tau}{1-q_\tau} \leq \sum_{i=1}^{m}\pi_0(i,0)\log\left(\frac{\pi_1(i,1)}{\pi_0(i,0)}\right) = -\sum_{i=1}^{m}\pi_0(i,0)\log\left(\frac{\pi_0(i,0)}{\pi_1(i,1)}\right). \tag{A-29}$$

Similarly, define $\bar{\pi}_0(i,0) = \max_{q\le\epsilon}\mathbb{P}_{\theta,\zeta}[a \in T_i \mid a \in T, q, Q = 0]$ and $\underline{\pi}_1(i,1) = \min_{q\ge 1-\epsilon}\mathbb{P}_{\theta,\zeta}[a \in T_i \mid a \in T, q, Q = 1]$, and thus

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\frac{q_\tau}{1-q_\tau} \ge \lim_{\tau\to\infty}\sum_{i=1}^{m}\frac{N_\tau(i)}{\tau}\log\left(\frac{\underline{\pi}_1(i,1)}{\bar{\pi}_0(i,0)}\right) = \sum_{i=1}^{m}\pi_0(i,0)\log\left(\frac{\underline{\pi}_1(i,1)}{\bar{\pi}_0(i,0)}\right).$$

Again, by letting $\epsilon \to 0$ we have $\bar{\pi}_0(i,0) \to \pi_0(i,0)$ and $\underline{\pi}_1(i,1) \to \pi_1(i,1)$, which yields

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\frac{q_\tau}{1-q_\tau} \ge -\sum_{i=1}^{m}\pi_0(i,0)\log\left(\frac{\pi_0(i,0)}{\pi_1(i,1)}\right). \tag{A-30}$$

Therefore, combining (A-29) and (A-30), we obtain

$$\lim_{\tau\to\infty}\frac{1}{\tau}\log\frac{q_\tau}{1-q_\tau} = -D\big(\boldsymbol{\pi}(F_{\theta,\zeta}, Q=0, q=0, T) \parallel \boldsymbol{\pi}(F_{\theta,\zeta}, Q=1, q=1, T)\big),$$

which completes the proof. *Q.E.D.*

A.2. *Microfoundations for Equation (3)*

A.2.1. *Expressive Utility*

Suppose first that customers have an *expressive overall utility*, meaning that their review decisions reflect their wish to express their satisfaction/dissatisfaction. More specifically, users have overall utility defined over the realized material utility (u_t) and the set of review options: $V : \mathbb{R} \times \mathcal{R} \to \mathbb{R}$. Given the overall expressive utility of users, we next formally define the strategy profile of users and a Bayes–Nash equilibrium in this setting.

Review and purchase decision strategy profile: For any customer t, we let $R_t : \Omega_t \times \mathbb{R} \to \mathcal{R}$ denote the *review decision strategy* of customer t, which is a mapping from the information provided from the rating system to tth customer and her material utility u_t to a review decision in set $\mathcal{R}$. We refer to the collection of review decision strategies $\mathbf{R} = \{R_t\}_{t=1}^{\infty}$ as the *review decision strategy profile*. Given a review decision strategy profile, customer t makes a purchase decision, which is a mapping from Ω_t, θ_t, ζ_t, and the price p to her material utility and is equal to $u_t = \theta_t + \zeta_t + Q - p$, if she purchases the product and equal to zero, otherwise. We represent the purchase decision of customer t by $B_t : \Omega_t \times \mathbf{R} \to \{0,1\}$, which maps from the information provided by the rating system at time t into a purchase decision. We also refer to the collection of purchase decision strategies $\mathbf{B} = \{B_t\}_{t=1}^{\infty}$ as the *purchase decision strategy profile*.

Bayes–Nash equilibrium: A review decision strategy profile $\mathbf{R}^e = \{R_t^e\}_{t=1}^{\infty}$ and a purchase decision strategy profile $\mathbf{B}^e = \{B_t^e\}_{t=1}^{\infty}$ constitute a Bayes–Nash equilibrium if no customer has a profitable deviation from her purchase and review decision strategies. This means, in particular, that each customer uses a belief that is formed by using Bayes' rule and taking the review and purchase decision strategy profile of others as given. In particular, at time t the customer observes θ_t and Ω_t and therefore in a Bayes–Nash equilibrium, her purchase and review decision become

$$\operatorname*{arg\,max}_{b\in\{0,1\}, r\in\mathcal{R}} \mathbb{E}\big[V(u_t, r) \mid \Omega_t, \theta_t\big].$$

We now impose two assumptions relevant for this Appendix.

ASSUMPTION A-1: *The overall utility function* $V:\mathbb{R}\times\mathcal{R}\to\mathbb{R}$ *is such that for any* $q\in[0,1]$ *and* θ

$$\mathbb{E}_{\zeta,Q}\Big[\max_{r\in\mathcal{R}} V(\theta+\zeta-p+Q,r)\Big]\geq 0 \quad \textit{if and only if} \quad \mathbb{E}_{\zeta,Q}[\theta+\zeta-p+Q]\geq 0,$$

where $Q\in\{0,1\}$ *is a binary random variable such that* $\mathbb{P}[Q=1]=1-\mathbb{P}[Q=0]=q$.

This assumption decomposes the purchase and review decisions. In particular, given this assumption, the purchase decision of customer t becomes

$$b_t=1 \quad \text{if and only if} \quad \theta_t+\mathbb{E}[\zeta_t+Q-p\mid\Omega_t]\geq 0.$$

We can write the equilibrium purchase decision as a function of Ω_t and the reviews:

$$B_t(\Omega_t,\mathbf{R})=1 \quad \text{if and only if} \quad \theta_t+\mathbb{E}[\zeta]+q_t(\Omega_t,\mathbf{R})-p\geq 0,$$

where $q_t:\Omega_t\times\mathbf{R}\to[0,1]$ is a mapping form the information provided from the rating system to tth customer and the review decision strategy profile of all customers to a belief about the true quality. The review decision in a Bayes–Nash equilibrium is given by

$$R_t^e(\Omega_t,u_t)\in\arg\max_{r\in\mathcal{R}} V(u_t,r), \quad \text{for all } t\geq 1, u_t\in\mathbb{R},\Omega_t$$

ASSUMPTION A-2: *The overall utility function* $V:\mathbb{R}\times\mathcal{R}\to\mathbb{R}$ *satisfies*:
1. *Increasing differences*, *that is*,

$$V(u',r')-V(u',r)\geq V(u,r')-V(u,r), \quad \forall u'\geq u, r'\geq r.$$

2. *Boundary conditions*:

$$\lim_{u\to\infty}\Big(V(u,\overline{K})-\max_{r\in\mathcal{R}\setminus\{\overline{K}\}} V(u,r)\Big)>0 \quad \textit{and} \quad \lim_{u\to-\infty}\Big(V(u,-\underline{K})-\max_{r\in\mathcal{R}\setminus\{-\underline{K}\}} V(u,r)\Big)>0.$$

This assumption ensures that customers prefer to express a more positive opinion when their own experience was more positive. The increasing differences property in Part 1 of Assumption A-2 ensures that customers derive greater overall utility from a more positive review when their own material payoff is greater. Part 2, on the other hand, imposes simple boundary conditions that guarantee that customers will leave both the most favorable and least favorable reviews for some realizations of their material utility. Given this assumption, the next lemma characterizes users' review decisions.

LEMMA A-2: *Suppose Assumptions* A-1 *and* A-2 *hold*. *Then there exist thresholds* $\lambda_{-\underline{K}}\leq\ldots\lambda_{-1}\leq\lambda_1\leq\ldots\lambda_{\overline{K}}\in\mathbb{R}$ *such that utility-maximizing review decisions are*

$$r_t=\begin{cases}-\underline{K} & \textit{if } u_t<\lambda_{-\underline{K}},\\ i & \textit{if } \lambda_{i-1}\leq u_t<\lambda_i, -\underline{K}<i<0,\\ 0 & \textit{if } \lambda_{-1}\leq u_t<\lambda_1,\\ i & \textit{if } \lambda_i\leq u_t<\lambda_{i+1}, 0<i<\overline{K},\\ \overline{K} & \textit{if } u_t\geq\lambda_{\overline{K}}.\end{cases}$$

A.2.2. *Consequentialist Utility*

The threshold review decisions in equation (3) and our results also hold in a model in which individuals leave reviews to influence others' decisions, which we refer to as *consequentialist utility*. Here, we outline the model and the main arguments.

We simplify our analysis by assuming that each user cares only about the action of the next user (rather than the full sequence of subsequent actions). Let Ω_t again denote the information provided from the rating system to tth customer (this can either be full history or a summary statistics) and q_t denote the belief of customer t. Customer t makes a purchase decision and a review decision. Similar to our baseline model, the purchase decision is $b_t \in \{0, 1\}$ where $b_t = 1$ if and only if $\theta_t + \mathbb{E}[\zeta] + q_t - p \geq 0$. A customer who has purchased the product experiences her material utility, $u_t = \theta_t + \zeta_t + Q - p$, and then on the basis of this, decides what review to leave. We formalize the idea that a user cares about the next user's purchasing decision (summarized by their purchase probability, denoted by pur_{t+1}) by positing an overall utility function $V : \mathbb{R} \times [0, 1] \to \mathbb{R}$, where the first argument is the user's material utility and the second is the purchase probability of the next user. Here, again, we impose Assumption A-1. In a Bayes–Nash equilibrium, users will correctly take into account how the next user's purchase probability depends on their reviews and try to influence this purchase probability. Review and purchase decisions, and the Bayes–Nash equilibrium are defined similarly. In particular, for any customer t, we let $R_t : \Omega_t \times \mathbb{R} \to \mathcal{R}$ denote the *review decision strategy* of customer t, which is a mapping from the information provided from the rating system to tth customer and her material utility u_t to a review decision in set $\mathcal{R}$. We refer to the collection of review decision strategies $\mathbf{R} = \{R_t\}_{t=1}^{\infty}$ as the *review decision strategy profile*. The purchase decision of customer t is $B_t : \Omega_t \times \mathbf{R} \to \{0, 1\}$, which maps from the information provided by the rating system at time t into a purchase decision. We can write this function as

$$B_t(\Omega_t, \mathbf{R}) = 1 \quad \text{if and only if} \quad \theta_t + \mathbb{E}[\zeta] + q_t(\Omega_t, \mathbf{R}) - p \geq 0, \tag{A-31}$$

where $q_t : \Omega_t \times \mathbf{R} \to [0, 1]$ is a mapping form the information provided from the rating system to tth customer and the review decision strategy profile of all customers to a belief about the true quality.[16] We also refer to the collection of purchase decision strategies $\mathbf{B} = \{B_t\}_{t=1}^{\infty}$ as the *purchase decision strategy profile*. A Bayes–Nash equilibrium (or equivalently a sequential equilibrium) is defined analogous to the expressive utility case, and we omit the details to save space.

The following assumption replaces Assumption A-2 and Assumption 1.

ASSUMPTION A-3: 1. *The random variables θ and ζ have continuous and strictly increasing cumulative distribution functions over their supports*, $[\underline{\theta}, \bar{\theta}]$ *and* $[\underline{\zeta}, \bar{\zeta}]$, *respectively*.

2. *The utility function $V : \mathbb{R} \times [0, 1] \to \mathbb{R}$ is continuous and satisfies increasing differences*, *that is*,

$$\begin{aligned} &V(u_t', \mathrm{pur}_{t+1}') - V(u_t', \mathrm{pur}_{t+1}) \\ &\quad \geq V(u_t, \mathrm{pur}_{t+1}') - V(u_t, \mathrm{pur}_{t+1}), \quad \forall u_t' \geq u_t, \mathrm{pur}_{t+1}' \geq \mathrm{pur}_{t+1}. \end{aligned}$$

3. *We impose the following boundary conditions*: *for any* $\mathrm{pur}' > \mathrm{pur}$ *and* $u \geq \bar{\theta} + \bar{\zeta} - p$, *we have* $V(u, \mathrm{pur}') - V(u, \mathrm{pur}) > 0$, *and for any* $\mathrm{pur}' > \mathrm{pur}$ *and* $u \leq \bar{\theta} + \underline{\zeta} - p + 1$, *we have* $V(u, \mathrm{pur}) - V(u, \mathrm{pur}') > 0$.

[16]In both purchase decision B_t and belief q_t, it is the review decisions strategy of previous customers (i.e., R_s for $s = 1, \dots, t$) that are relevant, but we condition on the entire $\mathbf{R}$ for notational simplicity.

Part 1 of this assumption is similar to Assumption 1. Part 2 is similar to Part 1 of Assumption A-2, but with a different interpretation. What the customer now cares about is whether or not the next customer purchases the product, and increasing differences in this context implies that the greater is her material utility, the more she would like the next customer to purchase. The complication arises from the fact that the purchase probability pur_{t+1} of the next customer is itself endogenous and depends on her belief, in turn determined by the review decision as well as the strategy shaping the review decision. Finally, Part 3 of this assumption is a combination of Part 2 of Assumption A-2 and the boundary conditions in Assumption 1 in the text. Recall that Part 2 of Assumption A-2 ensured that both review $\overline{K}$ ("like") and review $-\underline{K}$ ("dislike") are possible for some values of material utility and the boundary conditions in Assumption 1 ensured that the distributions of θ and ζ were such that both of these reviews were left with positive probability. With "consequentialist utility" we need to combine these two assumptions because the utility function depends on the purchase probability of the next customer which in turn depends on the distributions of θ and ζ. This is the feature that Part 3 of this assumption imposes.

Our main result is (proof in Appendix B.1):

THEOREM A-1: *Suppose that customers have consequentialist utility and Assumptions A-1 and A-3 hold. Then there exists a Bayes–Nash equilibrium in which review decisions take a threshold form. Suppose in addition that the strict separation condition holds. Then, in this equilibrium, Theorems 1–4 hold.*

REFERENCES

ACEMOGLU, DARON, MUNTHER DAHLEH, ILAN LOBEL, AND ASUMAN OZDAGLAR (2009): "Rate of Convergence of Learning in Social Networks," in *Proceedings of the American Control Conference*, 2825–2830. [2859]

——— (2011): "Bayesian Learning in Social Networks," *Review of Economic Studies*, 78, 1201–1236. [2859]

ACEMOGLU, DARON, ALI MAKHDOUMI, AZARAKHSH MALEKIAN, AND ASU OZDAGLAR (2022): "Supplement to 'Learning From Reviews: The Selection Effect and the Speed of Learning'," *Econometrica Supplemental Material*, 90, https://doi.org/10.3982/ECTA15847. [2860]

AMADOR, MANUEL, AND PIERRE WEILL (2012): "Learning From Private and Public Observations of Others' Actions," *Journal of Economic Theory*, 147, 910–940. [2859]

BANERJEE, ABHIJIT, AND DREW FUDENBERG (2004): "Word-of-Mouth Learning," *Games and Economic Behavior*, 46, 1–22. [2859]

BANERJEE, ABHIJIT V. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817. [2858]

BESBES, OMAR, AND MARCO SCARSINI (2018): "On Information Distortions in Online Ratings," *Operations Research*, 66, 597–610. [2860]

BIKHCHANDANI, SUSHIL, DAVID HIRSHLEIFER, AND IVO WELCH (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026. [2858]

BOHREN, J. AISLINN, AND DANIEL HAUSER (2017): "Bounded Rationality and Learning: A Framework and a Robustness Result," Working paper. [2860]

CAI, HONGBIN, YUYU CHEN, AND HANMING FANG (2009): "Observational Learning: Evidence From a Randomized Natural Field Experiment," *American Economic Review*, 99, 864–882. [2860]

CHE, YEON-KOO, AND JOHANNES HÖRNER (2018): "Recommender Systems as Mechanisms for Social Learning," *Quarterly Journal of Economics*, 133, 871–925. [2860]

CHEVALIER, JUDITH A., AND DINA MAYZLIN (2006): "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43, 345–354. [2860]

CHUA, ALTON Y. K., AND SNEHASISH BANERJEE (2016): "Helpfulness of User-Generated Reviews as a Function of Review Sentiment, Product Type and Information Quality," *Computers in Human Behavior*, 54, 547–554. [2860]

COVER, THOMAS M., AND JOY A. THOMAS (2012): *Elements of Information Theory*. John Wiley & Sons. [2858, 2866]

DASARATHA, KRISHNA, AND KEVIN HE (2019): "Aggregative Efficiency of Bayesian Learning in Networks," Working paper. [2859]
DURRETT, RICK (2010): *Probability: Theory and Examples*. Cambridge University Press. [2878]
ETEMADI, NASROLLAH (1981): "An Elementary Proof of the Strong Law of Large Numbers," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55, 119–122. [2890]
EYSTER, ERIK, AND MATTHEW RABIN (2010): "Naive Herding in Rich-Information Settings," *American Economic Journal: Microeconomics*, 2, 221–243. [2860]
FRICK, MIRA, RYOTA IIJIMA, AND YUHTA ISHII (2020): "Misinterpreting Others and the Fragility of Social Learning," *Econometrica*, 88, 2281–2328. [2860]
GARG, NIKHIL, AND RAMESH JOHARI (2017): "Pairwise Comparisons for Online Reputation Systems," Preprint. [2860]
GLOSTEN, LAWRENCE R., AND PAUL R. MILGROM (1985): "Bid, Ask and Transaction Prices in a Specialist Market With Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100. [2866]
GOLUB, BENJAMIN, AND MATTHEW O. JACKSON (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149. [2860]
GOLUB, BENJAMIN, AND EVAN SADLER (2016): "Learning in Social Networks," in *Oxford Handbook of the Economics of Networks*. [2860]
GUARINO, ANTONIO, AND PHILIPPE JEHIEL (2013): "Social Learning With Coarse Inference," *American Economic Journal: Microeconomics*, 5, 147–174. [2860]
HANN-CARUTHERS, WADE, VADIM V. MARTYNOV, AND OMER TAMUZ (2018): "The Speed of Sequential Asymptotic Learning," *Journal of Economic Theory*, 173, 383–409. [2859]
HAREL, MATAN, ELCHANAN MOSSEL, PHILIPP STRACK, AND OMER TAMUZ (2014): "When More Information Reduces the Speed of Learning," Available at SSRN 2541707. [2859]
HÖRNER, JOHANNES, AND NICOLAS S. LAMBERT (2021): "Motivational Ratings," *The Review of Economic Studies*, 88, 1892–1935. [2860]
HU, NAN, PAUL A. PAVLOU, AND JENNIFER ZHANG (2006): "Can Online Reviews Reveal a Product's True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication," in *Proceedings of the 7th ACM Conference on Electronic Commerce*, 324–330. [2860]
HU, NAN, PAUL A. PAVLOU, AND JIE JENNIFER ZHANG (2017): "On Self-Selection Biases in Online Product Reviews," *MIS Quarterly*, 41, 449–471. [2860]
HU, NAN, JIE ZHANG, AND PAUL A. PAVLOU (2009): "Overcoming the J-Shaped Distribution of Product Reviews," *Communications of the ACM*, 52, 144–147. [2860]
IFRACH, BAR, COSTIS MAGLARAS, MARCO SCARSINI, AND ANNA ZSELEVA (2019): "Bayesian Social Learning From Consumer Reviews," *Operations Research*, 67, 1209–1221. [2860]
KORCHEVSKY, VALERY, AND VALENTIN PETROV (2010): "On the Strong Law of Large Numbers for Sequences of Dependent Random Variables," *Vestnik St. Petersburg University: Mathematics*, 43, 143–147. [2890]
LI, XINXIN, AND LORIN M. HITT (2008): "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19, 456–474. [2860]
LOBEL, ILAN, AND EVAN SADLER (2015a): "Information Diffusion in Networks Through Social Learning," *Theoretical Economics*, 10, 807–851. [2859]
——— (2015b): "Preferences, Homophily, and Social Learning," *Operations Research*, 64, 564–584. [2859]
MCLENNAN, ANDREW (1984): "Price Dispersion and Incomplete Learning in the Long Run," *Journal of Economic Dynamics and Control*, 7, 331–347. [2859,2865]
MOSSEL, ELCHANAN, ALLAN SLY, AND OMER TAMUZ (2014): "Asymptotic Learning on Bayesian Social Networks," *Probability Theory and Related Fields*, 158, 127–157. [2859]
——— (2015): "Strategic Learning and the Topology of Social Networks," *Econometrica*, 83, 1755–1794. [2859]
ROSENBERG, DINAH, AND NICOLAS VIEILLE (2019): "On the Efficiency of Social Learning," *Econometrica*, 87, 2141–2168. [2859]
SMITH, LONES, AND PETER SØRENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398. [2858,2859,2865]
TALWAR, ARJUN, RADU JURCA, AND BOI FALTINGS (2007): "Understanding User Behavior in Online Feedback Reporting," in *Proceedings of the 8th ACM Conference on Electronic Commerce*, 134–142. [2860]
THOMAS EWART A. C. (1971): "Sufficient Conditions for Monotone Hazard Rate an Application to Latency-Probability Curves," *Journal of Mathematical Psychology*, 8, 303–332. [2873]
VELLODI, NIKHIL (2018): "Ratings Design and Barriers to Entry," Available at SSRN 3267061. [2860]
VIVES, XAVIER (1993): "How Fast Do Rational Agents Learn?" *Review of Economic Studies*, 60, 329–347. [2859]
——— (1995): "The Speed of Information Revelation in a Financial Market Mechanism," *Journal of Economic Theory*, 1, 178–204. [2859]

WELCH, IVO (1992): "Sequential Sales, Learning, and Cascades," *Journal of Finance*, 47, 695–732. [2858]
WOLITZKY, ALEXANDER (2018): "Learning From Others' Outcomes," *American Economic Review*, 108, 2763–2801. [2859]

Co-editor Dirk Bergemann handled this manuscript.

Manuscript received 13 November, 2017; final version accepted 25 July, 2022; available online 27 July, 2022.