

NBER WORKING PAPER SERIES

FAST AND SLOW LEARNING FROM REVIEWS

Daron Acemoglu
Ali Makhdoumi
Azarakhsh Malekian
Asuman Ozdaglar

Working Paper 24046
<http://www.nber.org/papers/w24046>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2017

We thank participants at several seminars and conferences for useful suggestions and comments. We gratefully acknowledge financial support from the Toulouse Network with Information Technology and Army Research Office, ARO MURI W911NF-12-1-0509. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Fast and Slow Learning From Reviews

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar

NBER Working Paper No. 24046

November 2017

JEL No. C72,D83,L15

ABSTRACT

This paper develops a model of Bayesian learning from online reviews, and investigates the conditions for asymptotic learning of 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. A sequence of potential customers decide whether to join the platform. After joining and observing the ratings of the product, and conditional on her ex ante valuation, a customer decides whether to purchase or not. If she purchases, the true quality of the product, her ex ante valuation, an ex post idiosyncratic preference term and the price of the product determine her overall satisfaction. Given the rating system of the platform, she decides to leave a review as a function of her overall satisfaction. 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 that they have available at the time of their purchase. We provide conditions for asymptotic learning under both full history and summary statistics, and show how the selection effect becomes more difficult to correct for with summary statistics. Conditional on asymptotic learning, the speed (rate) of learning is always exponential and is governed by similar forces under both types of rating systems, though the exact rates differ. Using this characterization, we provide the rate of learning under several different types of rating systems. We show that providing more information does not always lead to faster learning, but strictly finer rating systems always do. We also illustrate how different rating systems, with the same distribution of preferences, can lead to very fast or very slow speeds of learning.

Daron Acemoglu
Department of Economics, E52-446
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and CIFAR
and also NBER
daron@mit.edu

Ali Makhdoumi
Electrical Engineering and Computer Science
Massachusetts Institute of Technology
77 Massachusetts Ave, 32-D640
Cambridge, MA 02139
makhdoum@mit.edu

Azarakhsh Malekian
Rotman School of Management
University of Toronto
and LIDS, MIT
azarakhsh.malekian@rotman.utoronto.ca

Asuman Ozdaglar
Dept of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
77 Massachusetts Ave, E40-130
Cambridge, MA 02139
asuman@mit.edu

1 Introduction

The fraction of consumer purchases performed online has reached 8.9%, and now accounts for \$100 billion. Amazon, alone, now has over 70 million prime customers.¹ As important as the lower costs of online transactions are the potential benefits of online information sharing. For instance, many platforms such as Amazon, eBay or Airbnb encourage users to provide reviews and present to potential new customers/users summaries of reviews by past users. It would not be an exaggeration to say that the success of many of those platforms depends in large part on the informativeness of their reviews, since otherwise users would not know which sellers to trust and may often be unable to gather enough information about some of the product choices available to them. Despite the centrality of such online reviews, there is relatively little work investigating the specific challenges such systems face and their efficacy in aggregating dispersed information of diverse users.

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* systems (where the reviews of all past users, and their exact sequence, are presented to current potential users), and more realistically, *summary statistics* systems (where some summary statistics of the reviews of past users are shown 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 online platform and decide whether to enter the platform. Upon entry, they observe their taste parameter, which determines how likely they are to like the product in question, obtain information about the product according to the rating system, and decide whether to purchase it or not. Following purchase, they experience their utility from the product, which also depends on the realization of an additional ex post idiosyncratic preference parameter, and may decide to leave a review. Whether they leave a review and which score they choose depends on their ex post utility relative to some pre-specified thresholds.² Using this model, we investigate how well (and how rapidly) the information of users is aggregated by various rating systems.

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 users know part of their preferences before purchase, those making a purchase are “selected” according to their taste parameters. Moreover, this selection becomes more pronounced when the information about the product is not very favorable — e.g., only those very biased towards a particular type of book or service would consider buying it if past purchasers have very negative assessments of its quality. The selection effect is different from the difficulty faced by

¹US Census Bureau News, U.S. Department of Commerce.

²In much of our analysis, we take these thresholds as given, and return to a discussion of how they are determined at the end.

models of observational learning, such as Bikhchandani et al. [1992], Welch [1992] and Banerjee [1992], which we discuss in greater detail below. 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 clarify the conditions under which asymptotic learning — which enables users to obtain very accurate estimates of the underlying quality from sufficiently many reviews — occurs, and highlight the role of Bayesian updating in undoing the implications of the selection effect. In particular, with naive agents who do not recognize the presence of the selection effect, we show that asymptotic learning typically fails.

Second, in addition to conditions for asymptotic learning, we investigate *the speed of learning*. In particular, both in the full history and summary statistics cases, we show that learning is exponentially fast and provide a tight characterization of the speed (or the rate) of learning as a function of the Kullback-Leibler (KL) divergence between the probability distribution of reviews conditional on high vs. low quality.³ In addition, because of the aforementioned selection effect, these probability distributions depend on the “public belief”, which summarizes the assessment of past users concerning the quality of the product. 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 problem depends on the exact information structures.

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 rely on potential users having this information in real time. In fact, we show that the revenues of the platform are higher under a rating system with higher learning speed, which confirms that the platform’s incentives are in fact aligned with accelerating learning.

Third, we also provide a number of results on what types of rating systems lead to faster aggregation of information. We show that, in general, providing more information to users does not guarantee faster learning. In fact, some summary statistics may lead to more rapid learning than allowing users to see the full history of past reviews. Nevertheless, we also show that “strict refinements” of the review system (which increase the granularity of the information provided by the rating system) lead to faster learning. Examples of such strict refinements include a change in the rating system so that different types of reviews that were aggregated previously are now disaggregated, and replacing averages of past reviews with more detailed information about fractions

³The 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, which we overcome by using a different line of proof and show that the speed of learning is still related to KL divergence.

of past customers leaving different reviews. We also illustrate that, for a given set of underlying preference and information parameters, different rating systems can lead to very slow or very fast learning.

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 very different and novel approach.⁴ Asymptotic learning takes place under very mild assumptions with full history, and requires somewhat more restrictive, but still reasonable, assumptions when the rating system provides summary statistics. This difference is because having access to the full history of reviews enables users to undo the implications of the selection effect mentioned above, while the selection effect becomes more challenging when users have access only to summary statistics. Perhaps surprisingly, conditional on asymptotic learning, the speed of learning is given by a similar logic and similar expressions in the two cases, even though the exact speed of learning could differ quite significantly in these scenarios.

Our work is related to several literatures. First, as already mentioned, we build on, but are very distinct from, the Bayesian observational learning literature pioneered by Bikhchandani et al. [1992], Welch [1992] and Banerjee [1992], where each agent sees the actions of some or all past agents. Smith and Sørensen [2000] provide a comprehensive analysis of such models when individuals observe all past actions (see also Banerjee and Fudenberg [2004]), while Acemoglu et al. [2011], Lobel and Sadler [2015a], Lobel and Sadler [2015b], Mossel et al. [2014] and Mossel et al. [2015], among others, study Bayesian observational learning when agents observe a subset of past actions determined according to a stochastic network.⁵ Our main innovation relative to this literature is the selection effect (which emerges as the main barrier to learning in our model), and our characterization of the speed of learning.

We are not aware of any comprehensive study of the speed of learning in Bayesian models of observational learning. The partial exceptions are Acemoglu et al. [2009], which studies the speed of learning in the baseline observational learning model when each agent observes the previous action and when each agent observes one randomly drawn action from the past; Hann-Caruthers et al. [2017], which compares the speed of learning in the baseline observational learning model when each agent observes the previous action versus the case in which they also observe past signals; Harel et al. [2014], which studies the speed of learning in a setting where finitely many agents repeatedly observe each other's actions; and Vives [1993], Vives [1995], and Amador and Weill [2012], which focus on the speed of learning in rational expectations equilibria, where agents learn from prices.

Our paper is also related to a few works on review and rating systems. Most relevant is Ifrach et al. [2014], which studies a setting similar to the full history version of our model with pricing,

⁴Namely, we construct two distributions, one majorized by the distribution of our summary statistics conditional on high quality, and one majorizing the distribution of our summary statistics conditional on low quality, and show that these two distributions are asymptotically separated.

⁵See also Tay et al. [2008] and Drakopoulos et al. [2013] for distributed detection models, where some of the same issues as in the baseline observational learning models may prevent learning.

and investigates the implications of the pricing strategy of the seller on learning.⁶ Their model provides guidelines for pricing in this setting, but it neither features the selection effect nor considers summary statistics nor characterizes the speed of learning. Less directly related but still relevant are: Che and Horner [2015], which studies the optimal review system to encourage experimentation by early users; Horner and Lambert [2016], which investigates the trade-off between the informational role of reviews and their impact on the seller’s effort on quality; and Garg and Johari [2017], which studies the implications of pairwise comparisons on online reputation building.⁷

The rest of the paper is organized as follows. In Section 2, we introduce our model. In Section 3 we formulate both customers’ and platform’s problems and define an equilibrium. Sections 4 and 5 provide conditions for asymptotic learning and characterize the speed of learning under full history and summary statistics, respectively. Section 6 presents several results on the effects of the design of rating systems on the speed of learning. Section 7 concludes, while we present the proofs of all of our results in the Appendix.

2 Environment

We consider a platform that is selling a product to a set of customers/users. The true quality of the product is unknown to both the customers and the platform. The platform has a rating system, which collects reviews from previous customers and provides a rating of the product (which could be a summary statistic of these reviews or their entire history) for future potential customers.⁸ New customers decide whether to join the platform. Upon joining, they observe their ex ante valuation and the information from the rating system of the platform, and decide whether to purchase the product. After the purchase decision, their utility from this product, which depends on true quality and an additional ex post idiosyncratic preference term, is realized, and they decide whether and what review to leave.

2.1 Customer Utility and Decisions

We assume that the true quality of the product is binary, corresponding to low or high quality, and denote it by $Q^* \in \{0, 1\}$. A sequence of potential customers decides whether to join the platform to collect additional information about the product and potentially to purchase it. In particular, one customer arrives at each time and is denoted by the time of her arrival, $t \in \mathbb{N}$. Customer t faces an entry cost drawn independently from a continuous distribution F_c . After observing this

⁶Crapis et al. [2016], Vaccari et al. [2016], and Besbes and Scarsini [2015] study various non-Bayesian models of learning from reviews.

⁷There is also a small empirical literature on reviews, including Blake et al. [2016], Talwar et al. [2007], Li and Hitt [2008], Lafky [2014], Mayzlin et al. [2014], and Chua and Banerjee [2016]. Though the motivations and the exact strategies of users when they leave reviews are complex, the evidence is broadly consistent with the notion that reviews are informative about the underlying quality of the product or provider and also reflect various idiosyncratic factors.

⁸Throughout, we refer to the scores or other information left by customers as “review,” and to the aggregate of these reviews provided by the platform as “rating”.

cost (and with the full knowledge of the rating system being used by the platform and the price of the product), the customer decides $j_t \in \{0, 1\}$, designating whether she has joined the platform.

The utility of customer t from purchasing the product is

$$\theta_t + \zeta_t + Q - p,$$

where p is the price of the product; θ_t is an ex ante idiosyncratic preference (valuation) term, drawn independently for each customer from a continuous distribution F_θ ; and ζ_t is an ex post idiosyncratic preference term, drawn also independently from a different continuous distribution F_ζ . Note that this is the “gross” utility of the customer, since we have not subtracted the cost of joining the platform c_t , which is already sunk at this point (the “net” utility can be obtained by subtracting c_t from this expression). The “gross” utility from not purchasing the good is normalized to zero.⁹

After entry into the platform ($j_t = 1$), the customer observes her ex ante valuation θ_t and the rating of the product provided by the platform based on past reviews. After receiving this information, she decides whether to purchase the product, which is denoted by $b_t \in \{0, 1\}$.

If she decides to purchase (i.e., $b_t = 1$), she experiences her full utility (or equivalently, she observes the true quality of the product and her ex post preference term ζ_t). At this point, the consumer decides whether to leave a review of the product and what review to leave. We assume that the rating system of the platform allows one of $-K_1, \dots, 0, \dots, K_2$ reviews, with 0 interpreted as leaving no review. For most of the paper, we assume that all customers have thresholds denoted by $\lambda_{-K_1} \leq \dots \leq \lambda_{-1} \leq \lambda_1 \leq \dots \leq \lambda_{K_2}$, and their reviews will be determined by the location of their utility relative to these thresholds. In particular, customer t chooses review r_t such that

$$r_t = \begin{cases} -K_1, & \theta_t + \zeta_t + Q - p \leq \lambda_{-K_1} \\ i, & \lambda_{i-1} \leq \theta_t + \zeta_t + Q - p \leq \lambda_i, \quad -K_1 < i < K_2, \\ K_2, & \theta_t + \zeta_t + Q - p \geq \lambda_{K_2}. \end{cases} \quad (1)$$

In what follows, we will often refer to the most favorable review K_2 , as “like,” and sometimes refer to the least favorable review $-K_1$, as “dislike”.

We also note that the idiosyncratic preference terms, θ_t and ζ_t , are customer t 's private information.

Throughout our analysis, we assume that the thresholds for reviews, $\lambda_{-K_1}, \dots, \lambda_{K_2}$ are fixed, same across customers, and publicly known. But this assumption is only for notational convenience. We could alternatively assume that these thresholds are given by $\lambda_{-K_1} + \nu_{-K_1}^t, \dots, \lambda_1 + \nu_1^t, \dots, \lambda_{K_2} + \nu_{K_2}^t$, where the ν terms are iid draws across individuals and thresholds, and are not observed by other agents. It is straightforward to see that in this case, the analysis applies without

⁹The important restriction implied by this form of customer utility is that there is a common “quality” that all customers care about. The additive structure also implies that they all care about it with similar intensity; this feature can be relaxed without changing our main results.

any modifications with the ex post idiosyncratic preference term ζ_t , being replaced by the sum of ζ_t and $-\nu$ terms. We return to a discussion of endogenous thresholds at the end of the paper.

We adopt the following “richness” assumption which guarantees that all reviews have a non-zero probability.

Assumption 1 (Richness). *The random variables, c , θ and ζ , have continuous cumulative density functions and the support of ζ is sufficiently wide to ensure that all possible reviews have positive probability; i.e.,*

$$\bar{\zeta} + \underline{\theta} - p > \lambda_{K_2} \text{ and } \underline{\zeta} + \bar{\theta} + 1 - p < \lambda_{-K_1},$$

where the support of θ is denoted by $[\underline{\theta}, \bar{\theta}]$ and the support of ζ is denoted by $[\underline{\zeta}, \bar{\zeta}]$.

2.2 The Rating System

The platform observes purchase decision b_t as well as review decision r_t . We let $h_t \in \{N\} \cup \{-K_1, \dots, K_2\}$, where $h_t = N$ designates “no purchase” by customer t , i.e., $b_t = 0$. We call h_t the action at time t . The history available to platform at time t is $H_t = \{h_1, \dots, h_{t-1}\}$ and by convention $H_1 = \emptyset$. The platform has a rating system denoted by Ω , which at time t maps the history of reviews by customers into a rating—which could be the same as the history itself or as summary statistic of the reviews. We denote the rating available for customer t by Ω_t . Examples of Ω_t include: (i) full history $\Omega_t = H_t$, and (ii) fraction of “likes”, i.e., fraction of K_2 reviews among all customers, and (iii) fraction of “likes” among “likes” and “dislikes”, i.e., fraction of K_2 reviews among $-K_1$ and K_2 reviews.

3 Customers’ and Platform’s Problems

We next introduce the optimization problems of customers and the platform. Our key assumptions are that both sets of agents maximize their utility and are Bayesian. This essentially amounts to looking for a *perfect Bayesian equilibrium* of the dynamic incomplete information game between the platform and the customers: all agents maximize their expected utility, and their expectations are given by Bayes’ rule whenever possible (i.e., whenever the event in question has nonzero probability). Though the extent of strategic interactions in our setup is limited, customers still need to reason through the review decisions of previous users in order to form their beliefs about the underlying quality of the product.

3.1 Customers’ Problem

We break the customers’ problem into an entry, a purchase, and a review decision as illustrated in Figure 1.

Entry Decision: before joining the platform, a customer only knows the price p , the cost c_t , and the rating system Ω (not the realization of Ω_t). In particular, she does not know the number of people

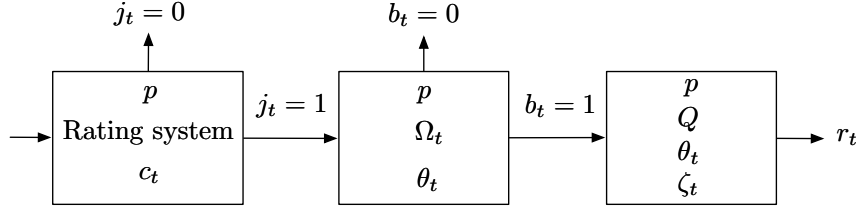


Figure 1: Entry, purchase and review decisions: the entry decision, j_t , is based on price, the rating system, and the cost of entry according to Eq. (2). The purchase decision, b_t , is based on price, the realization of the rating of the product, and the private ex ante valuation according to Eq. (3). The review decision, r_t , is based on price, the true quality of the product, the ex ante and ex post valuations according to Eq. (1).

who have joined the platform before her and has an improper uniform prior on this number.¹⁰ Therefore, in a perfect Bayesian equilibrium customer t 's entry decision is given by

$$j_t = \arg \max_{j \in \{0,1\}} \mathbf{1}\{j = 1\} (\mathbb{E}_{t, \Omega_t, Q, \theta_t, \zeta_t} [\mathbf{1}\{b_t = 1\} (\theta_t + \zeta_t + Q - p)] - c_t), \quad (2)$$

where $\mathbb{E}_{t, \Omega_t, Q, \theta_t, \zeta_t}$ is the Bayesian expectation over all random variables.

Purchase Decision: after joining the platform, customer t observes the realized rating Ω_t . Based on Ω_t , she forms a belief regarding the quality of the product as

$$q_t \triangleq \mathbb{P}[Q = 1 | \Omega_t].$$

When customer t joins the platform, she also observes her ex ante valuation θ_t . Therefore, in a perfect Bayesian equilibrium, customer t 's purchase decision is given by

$$b_t = \arg \max_{b \in \{0,1\}} \mathbf{1}\{b = 1\} (\mathbb{E}_{Q, \zeta_t} [\theta_t + \zeta_t + Q - p | \Omega_t]) = \arg \max_{b \in \{0,1\}} \mathbf{1}\{b = 1\} (\theta_t + \mathbb{E}[\zeta] + q_t - p). \quad (3)$$

This maximization problem makes the *selection effect*, described in the Introduction, apparent. If q_t is very low, only agents with high θ_t will choose $b_t = 1$.

Review Decision: after a customer purchases the product, she observes the true quality $Q^* \in \{0, 1\}$ as well as her ex post valuation from the product ζ_t based on which she leaves a review as given in Eq. (1).

3.2 The Platform's Problem

The platform does not know the true quality of the product and we also assume that it does not control the price (which is set by a seller). Its only decision is the rating system, Ω . Its revenues

¹⁰We can relax this assumption, which plays no role in our analysis of learning and the speed of learning (at that point the customer does know her place in the sequence). This assumption is only used in our analysis of the platform's maximization problem in subsection 6.6.

come from advertisements or other benefits that it derives from potential customers joining the platform. We assume, in particular, that it receives a fixed revenue for each customer that decides to join. Therefore, the platform’s problem is to maximize the expected number of customers who join the platform.

In Section 6, we show that the platform’s revenue is maximized by choosing a rating system with the highest speed of learning. Intuitively, this is because a rating system that ensures faster learning provides greater expected utility to potential customers, encouraging them to join the platform. Before establishing this result, we show that under mild assumptions, the relevant set of rating systems guarantee asymptotic learning and we characterize their speed of learning. We separately study learning dynamics for a rating system that provides the full history of reviews and rating systems that provide summary statistics of past reviews.

4 Full History

With full history, the rating system provides the entire history of previous actions, i.e, $\Omega_t = H_t$.¹¹ We show that under Assumption 1, as long as potential customers do not stop purchasing the product, this type of rating system guarantees asymptotic learning (almost sure convergence of beliefs to the true quality of the product). We then characterize the speed (rate) of learning.

We refer to the probability that the underlying quality is high given the history of reviews up to time t as *public belief* at time t , and denote it by q_t .¹² Clearly,

$$q_t = \mathbb{P}[Q = 1 | H_t].$$

Customer t , who receives this information from the rating system, can compute this public belief by Bayes’ rule. In fact, given full history, each customer can compute (and thus knows) the public belief at any time $s \leq t$.

We denote by l_t the likelihood ratio of this belief, i.e.,

$$l_t \triangleq \frac{q_t}{1 - q_t} = \frac{\mathbb{P}[Q = 1 | H_t]}{\mathbb{P}[Q = 0 | H_t]}.$$

Therefore, the likelihood ratio at time t is

$$l_t = \frac{q_t}{1 - q_t} = \frac{\mathbb{P}[Q = 1 | H_t]}{\mathbb{P}[Q = 0 | H_t]} = \frac{\mathbb{P}[H_t | Q = 1]}{\mathbb{P}[H_t | Q = 0]} = \prod_{s=1}^{t-1} \frac{\mathbb{P}[h_s | Q = 1, H_s]}{\mathbb{P}[h_s | Q = 0, H_s]}.$$

Note that l_t is a sufficient statistic of H_t for (estimating) Q .

¹¹In particular, we are assuming that the history also includes the “no purchases” and “no reviews”. This assumption is adopted to simplify the notation in our baseline analysis. We show at the end of the section that all of our results translate to the case in which the history only includes information about those who have left reviews.

¹²We follow the observational learning literature in referring to q_t as the “public belief”. This emphasizes that under full history, all customers $s \geq t$ can accurately infer this belief.

4.1 Learning Dynamics

We next show how the public belief evolves for customer $t + 1$. Customer $t + 1$ observes h_t (and not θ_t and ζ_t) based on which she updates her belief as

$$l_{t+1} = l_t \times \frac{\mathbb{P}[h_t|Q = 1, l_t]}{\mathbb{P}[h_t|Q = 0, l_t]},$$

where using Eq. (1) and Eq. (3) we have

$$\mathbb{P}[h_t|Q, l_t] = \begin{cases} \mathbb{P}[q_t + \theta_t + \mathbb{E}[\zeta] - p < 0], & h_t = N \\ \mathbb{P}[q_t + \theta_t + \mathbb{E}[\zeta] - p \geq 0, \theta_t + \zeta_t + Q - p \leq \lambda_{-K_1}], & h_t = -K_1 \\ \mathbb{P}[q_t + \theta_t + \mathbb{E}[\zeta] - p \geq 0, \lambda_{i-1} \leq \theta_t + \zeta_t + Q - p \leq \lambda_i], & -K_1 < i < K_2, h_t = i, \\ \mathbb{P}[q_t + \theta_t + \mathbb{E}[\zeta] - p \geq 0, \theta_t + \zeta_t + Q - p \geq \lambda_{K_2}], & h_t = K_2. \end{cases} \quad (4)$$

Therefore, the dynamics of public belief is captured by the following random process

$$l_{t+1} = l_t \times \frac{\mathbb{P}[h|Q = 1, l_t]}{\mathbb{P}[h|Q = 0, l_t]}, \quad \text{w.p. } \mathbb{P}[h|Q = Q^*, l_t], h \in \{N\} \cup \{-K_1, \dots, K_2\},$$

where Q^* is the true quality of the product.

4.2 Asymptotic Learning

We next show necessary and sufficient conditions for learning the true quality. We recall that $\bar{\theta}$ is the upper support of the distribution of θ .

Theorem 1. *Suppose Assumption 1 holds.*

- (a) *If $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$, then there is asymptotic learning, i.e., $q_t \rightarrow Q^*$ almost surely.*
- (b) *If $\bar{\theta} + \mathbb{E}[\zeta] - p < 0$, then there is a positive probability with which purchases stop and asymptotic learning does not occur.*

The proofs of all our results are provided in the Appendix.

The theorem shows that the condition $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$ is sufficient for asymptotic learning. When this condition does not hold, with positive probability all potential customers stop buying the product and asymptotic learning fails. This can be seen by noting that at the time of the purchase, the most positive assessment of expected quality will be from customer with the highest ex ante valuation $\bar{\theta}$, and is $\bar{\theta} + \mathbb{E}[\zeta] + q - p$, 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 thus beliefs remain stuck at q . Conversely, however, when condition $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$ holds, even for very pessimistic beliefs about the underlying quality,

some customers purchase the product and this generates sufficient information for asymptotic learning.

Since our interest is in the speed of learning, in what follows, we impose:

Assumption 2. $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$.

It is also instructive to revisit the difficulty faced by the agents in their updating decisions. As explained in the Introduction, this difficulty has its roots in the *selection effect* — the fact that the unobserved ex ante valuations of those who have previously purchased the good are correlated with their public belief. One way of seeing the role of the selection effect is to consider a “naive” learning rule, which simply considers the averages of past reviews, without attempting to correct for the fact that the set of customers leaving reviews was determined by the information available to them. Formally, we define a \bar{q} -naive learning rule as an updating rule that aggregates past reviews holding the public belief fixed at some level \bar{q} . The naivety of this rule comes from the fact that it does not take into account how the ex ante valuations of past customers change with the public belief at the time. The evolution of beliefs under a \bar{q} -naive learning rule is given by

$$\frac{q_t}{1 - q_t} = \prod_{s=1}^{t-1} \frac{\mathbb{P}[h_s | q = \bar{q}, Q = 1]}{\mathbb{P}[h_s | q = \bar{q}, Q = 0]}.$$

The next proposition shows that asymptotic learning fails with naive learning rules.

Proposition 1. *Suppose Assumptions 1 and 2 hold. Provided that*

$$\mathbb{E}_{h \sim \mathbb{P}[\cdot | q=0, Q=0]} \left[\frac{\mathbb{P}[h | q = \bar{q}, Q = 0]}{\mathbb{P}[h | q = \bar{q}, Q = 1]} \right] < 1, \quad (5)$$

there is no asymptotic learning with a \bar{q} -naive learning rule when $Q^ = 0$, and provided that*

$$\mathbb{E}_{h \sim \mathbb{P}[\cdot | q=1, Q=1]} \left[\frac{\mathbb{P}[h | q = \bar{q}, Q = 1]}{\mathbb{P}[h | q = \bar{q}, Q = 0]} \right] < 1, \quad (6)$$

there is no asymptotic learning with a \bar{q} -naive learning rule when $Q^ = 1$.*

This proposition thus shows that under conditions (5) and (6), naive customers, who do not recognize the selection effect, will not learn the underlying quality. To understand these conditions, let us consider the case where $Q^* = 1$ so that the relevant condition is (6). The term in square brackets is the likelihood ratio of history h under $Q = 1$ relative to $Q = 0$ holding the public belief at $q = \bar{q}$. The expectation of this likelihood ratio formed under correct public beliefs (that is, under $q = 1$) is always greater than 1, and this ensures asymptotic learning with Bayesian agents. However, when $q = \bar{q} < 1$, this expectation can be less than 1. To see when this will be the case, consider a rating system with only two review options, “like” and “dislike”. Then condition (6) is equivalent to $\mathbb{P}[h = K_2 | q = 1, Q = 1] < \mathbb{P}[h = K_2 | q = \bar{q}, Q = 0]$. This condition is satisfied if there is a sufficiently strong selection effect — if the entry of customers with relatively unfavor-

able ex ante types when $q = 1$ makes a “like” (K_2) review less likely under $Q = 1$ than it would be with $q = \bar{q}$ and $Q = 0$.

In contrast to this result, recall that Theorem 1 has shown that Bayesian updating always leads to asymptotic learning with full history. This is because with access to the full history of reviews, Bayesian agents can form a correct assessment of the public beliefs of all past users and thus undo the selection effect. For example, in the case with the two review options, the relevant comparison for Bayesian agents is $\mathbb{P}[h = K_2 \mid q = 1, Q = 1]$ vs. $\mathbb{P}[h = K_2 \mid q = 1, Q = 0]$, and because the former is always greater, asymptotic learning is guaranteed. Put differently, Bayesian agents successfully filter out the correlation between reviews at time t and the distribution of ex ante valuations of potential customers who would have purchased at time t , and this removes the selection effect and ensures asymptotic learning. We will see in the next section that, without access to the full history of reviews, dealing with the selection effect becomes more challenging, though Bayesian customers will still be able to learn the underlying quality under some additional, relatively weak conditions.

We should also note at this point that the intuition for asymptotic learning in Theorem 1 is very different than in the baseline model of observational learning (in particular, Smith and Sørensen [2000] and Acemoglu et al. [2011]). In these prior works learning occurs if the signals are unbounded, meaning that there is a positive probability of a very informative signal. This ensures that “herding,” where all agents follow the action suggested by the public belief, disregarding their own information, is not possible. In contrast, there is no possibility of herding here, because agents leave reviews after experiencing the true quality. Instead, as we have just discussed, the present challenge is to filter out the selection effect, which Bayesian agents can always do when they have access to full history.

4.3 Speed of Learning

We next characterize the speed of learning under the full history rating system, i.e., $\Omega_t = H_t$ for all t . For this purpose, we introduce the Kullback-Leibler (KL) divergence between two distributions.

Definition 1 (KL divergence). Suppose μ and ν are two distributions on a measurable space \mathcal{X} . Kullback-Leibler divergence is defined as

$$D(\mu||\nu) \triangleq \begin{cases} \mathbb{E}_\mu \left[\log \left(\frac{d\mu}{d\nu} \right) \right], & \mu \ll \nu \\ \infty, & \text{o.w.,} \end{cases}$$

where $\mu \ll \nu$ means μ is absolutely continuous with respect to ν .¹³ In particular, if μ and ν are two

¹³A measure μ is absolutely continuous with respect to another measure ν if $\mu(E) = 0$ for every set with $\nu(E) = 0$. Also, throughout, \log refers to logarithm with base e .

distributions on a finite set \mathcal{X} , KL divergence becomes

$$D(\mu||\nu) \triangleq \sum_{x \in \mathcal{X}} \mu(x) \log \left(\frac{\mu(x)}{\nu(x)} \right).$$

The next theorem characterizes the rate of learning under full history.

Theorem 2. *Suppose Assumptions 1 and 2 hold. Then learning is exponentially fast, i.e., almost surely q_t converges exponentially fast to Q^* . Moreover, for $Q^* = 0$, we almost surely have*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log q_t = -D(\mathbb{P}[\cdot|q=0, Q=0] || \mathbb{P}[\cdot|q=0, Q=1]),$$

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

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log(1 - q_t) = -D(\mathbb{P}[\cdot|q=1, Q=1] || \mathbb{P}[\cdot|q=1, Q=0]).$$

In what follows we use rate of learning and speed of learning interchangeably to refer to the exponent with which belief converges to the true quality.

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., $h \in \{N\} \cup \{-K_1, \dots, K_2\}$) when the underlying quality is Q^* and the public belief is asymptotically equal to $q = Q^*$ and the probably distribution of observing the same sequence when the underlying quality is $1 - Q^*$. There are three components to the intuition for this result. First, the fact that the learning is exponentially fast follows from the ability of customers to overcome the selection effect, which then enables them to combine the information of the t signals (for customer $t + 1$) they are receiving from the reviews of past customers.

Second, the exact speed of learning is given by KL divergence. This is also intuitive. 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 (e.g., Cover and Thomas [2012, Theorem 11.8.3]). Technically, in our case the sequence of reviews does not correspond to an independently-drawn sample because of the selection effect — the current belief affects the distribution of types that will purchase and thus the probability distribution of reviews. Nevertheless, as $q \rightarrow Q^*$ 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. More specifically, note that when $Q^* = 0$,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log l_t = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{h \in \{N\} \cup \{-K_1, \dots, K_2\}} \sum_{s=1}^t \mathbf{1}\{h_s = h\} \log \left(\frac{\mathbb{P}[h | q_s, Q = 1]}{\mathbb{P}[h | q_s, Q = 0]} \right).$$

If we could apply the strong law of large numbers with $q_s = 0$, this would imply

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \log l_t &= \sum_{h \in \{N\} \cup \{-K_1, \dots, K_2\}} \mathbb{P}[h \mid q = 0, Q = 0] \log \left(\frac{\mathbb{P}[h \mid q = 0, Q = 1]}{\mathbb{P}[h \mid q = 0, Q = 0]} \right) \\ &= -D(\mathbb{P}[\cdot \mid q = 0, Q = 0] \parallel \mathbb{P}[\cdot \mid q = 0, Q = 1]). \end{aligned}$$

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.

Finally, it is also worth noting that both probability distributions in the KL divergence condition on $q = Q^*$. This is because, under full history, each customer correctly reasons about the public belief of previous users, which is converging to Q^* . It is this feature of the full history rating system that enables the effective filtering of selection effect, and will contrast with our analysis of the case of summary statistics, which we present in the next section.

4.4 When Customers Observe a Subset of Previous Actions

We end this section by showing that the results presented so far directly generalize to the case in which the history available to potential customers only includes a subset of actions. For instance, the history available to potential customers may only include reviews (and no information about those who have not purchased the product or have not left a review). Formally, we consider a subset of all possible actions $\{N\} \cup \{K_1, \dots, K_2\}$ denoted by T , and denote a partition of T by T_1, \dots, T_m , i.e., $T = \cup_{i=1}^m T_i$ and $T_i \cap T_j = \emptyset$, $i \neq j \in [m]$. We also use the notation τ to denote the subsequence of time t at which users observe an action in T (e.g., if customers observe past reviews, τ corresponds to the number of reviews so far). In other words, the history observed by the customer acting after $\tau - 1$ is $\{\tilde{h}_1, \dots, \tilde{h}_{\tau-1}\}$, where $\tilde{h}_s = i$ if the s -th action in T belongs to T_i . Customers do not observe any other event, and have uniform priors over the times at which there have been unobservable actions. The next proposition generalizes Theorem 2 to this case.

Proposition 2. *Suppose that Assumptions 1 and 2 hold, and that customers only observe actions in T and have uniform priors over the times at which actions that are not in T have taken place. Then, there is again asymptotic learning. Moreover, for $Q^* = 0$, we almost surely have*

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log q_\tau = -D(\mathbb{P}[\cdot \mid h \in T, q = 0, Q = 0] \parallel \mathbb{P}[\cdot \mid h \in T, q = 0, Q = 1]),$$

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

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log(1 - q_\tau) = -D(\mathbb{P}[\cdot \mid h \in T, q = 1, Q = 1] \parallel \mathbb{P}[\cdot \mid h \in T, q = 1, Q = 0]).$$

To understand this result, consider a customer at time t that observes a history with τ actions in T . This customer does not know the number of past users who have joined the platform between

the most recent action in T and her time. However, she knows exactly the belief of any user who has taken an action in T because she observes the same history as this user, and she can filter out the selection effect and aggregate information of all past customers who have taken an action in T , which is sufficient to ensure asymptotic learning. The speed of learning is modified in view of the fact that the information of customers who are not taking an action in T is not being transmitted, but is still given by a similar KL divergence term.

5 Summary Statistics

In this section, we focus on more realistic rating systems where the platform provides summary statistics of reviews by past customers. We then establish asymptotic learning in this case under an additional assumption (which is needed because filtering out the selection effect is more difficult without access to full history), and subsequently characterize the speed of learning in this case.

5.1 Learning Dynamics

We now focus on rating systems in which there exists a vector of summary statistics \mathbf{S} , providing various summaries of past reviews. As in our analysis in subsection 4.4, with summary statistics customers will not see the full history, and we again find it convenient to index them not by the time of their arrival, but by the number of actions that have taken place so far (i.e., actions that are in the set T defined in subsection 4.4). In our leading case, this will be the number of reviews left so far, denoted by τ . We also assume that customers know their τ (and not their exact place in the original sequence, and as in subsection 4.4, have uniform priors over their exact place in the sequence). Therefore, for consumer indexed by τ , the rating system takes the form $\Omega_\tau = \mathbf{S}_\tau$, and reports information from the first $\tau - 1$ reviews.

To define the summary statistics, we consider a partitioning of all review options $\{-K_1, \dots, K_2\} \setminus \{0\}$ (recall that 0 represents “no review”) denoted by $T = \cup_{i=1}^m T_i$. A summary statistic is a vector $\mathbf{S}_\tau \in \mathbb{R}^m$ with its i -th entry representing fraction of those reviews in T_i that has occurred among all reviews, i.e.,¹⁴

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

The following are examples of rating systems with summary statistics:

- The fractions of each one of $K_1 + K_2$ reviews are reported. In this case, $\mathbf{S}_\tau \in \mathbb{R}^m$ where $m = K_1 + K_2$, and $T = \{T_1, \dots, T_m\}$ is $T_1 = \{-K_1\}, \dots, T_{K_1} = \{-1\}, T_{K_1+1} = \{1\}, \dots$, and $T_{K_1+K_2} = \{K_2\}$.
- “Likes” among all reviews meaning that the rating system reports only the fraction of reviews that give the highest score, $h = K_2$, out of the available $K_1 + K_2$ options. This rating

¹⁴With this definition, for any $\tau \geq 0$, we have $\sum_{i=1}^m \mathbf{S}_\tau(i) = 1$. Therefore, one of the entries of vector \mathbf{S}_τ is redundant. This is chosen to simplify the notation used in the analysis and does not change any of our results.

system is represented by $\mathbf{S}_\tau \in \mathbb{R}^2$ where

$$\mathbf{S}_\tau(1) = \frac{1}{\tau} \sum_{s=1}^{\tau} \mathbf{1}\{h_s = K_2\}, \quad \text{and} \quad \mathbf{S}_\tau(2) = 1 - \mathbf{S}_\tau(1). \quad (7)$$

We will also discuss other, albeit perhaps less realistic, rating systems which provide information about fractions of reviews among all purchases (including the customers who have purchased the good in question, but have left no review in the denominator) and among all customers who have joined the platform. For example, when we focus on all purchases, with a slight abuse of notation, we will use τ to denote the number of purchases, and we consider partitions of $\{-K_1, \dots, K_2\}$ (rather than partitions of $\{-K_1, \dots, K_2\} \setminus \{0\}$). A particularly simple version of such a rating system would be one that reports “likes” among all purchases which can also be represented by Eq. (7), but with τ now corresponding to the number of purchases.

Similar to our analysis of the full history case, we denote the expectation of the underlying quality Q , conditional on the information available from the rating system \mathbf{S}_τ , by q_τ . In other words, $q_\tau = \mathbb{P}[Q = 1 \mid \mathbf{S}_\tau]$. The likelihood ratio implied by this belief is then

$$l_\tau = \frac{q_\tau}{1 - q_\tau} = \frac{\mathbb{P}[Q = 1 \mid \mathbf{S}_\tau]}{\mathbb{P}[Q = 0 \mid \mathbf{S}_\tau]} = \frac{\mathbb{P}[\mathbf{S}_\tau \mid Q = 1]}{\mathbb{P}[\mathbf{S}_\tau \mid Q = 0]}. \quad (8)$$

Crucially, however, unlike the full history case, q_τ is no longer a public belief — future customers who do not have access to \mathbf{S}_τ neither observe nor can compute q_τ . For this reason, the likelihood ratio l_τ is no longer a martingale, and we need to develop a different approach for analyzing the dynamics of beliefs and their asymptotic properties.¹⁵

With this objective, let us write the governing stochastic process for \mathbf{S}_τ given the true quality Q^* as

$$\mathbf{S}_{\tau+1} = \frac{\tau}{\tau+1} \mathbf{S}_\tau + \frac{1}{\tau+1} \mathbf{Y}_{\tau+1}, \quad \forall \tau \geq 0, \quad (9)$$

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

$$\mathbf{Y}_{\tau+1} = \mathbf{e}_i, \quad \text{w.p. } \mathbb{P}[h \in T_i \mid h \in T, q_\tau, Q^*],$$

where $\mathbf{e}_i \in \mathbb{R}^m$ is the i -th canonical basis (i.e., $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$, only with its i -th entry being equal to 1).

We should note that the same challenge that was discussed in the previous section is present here, i.e., in finding posterior probabilities in Eq. (8), Bayesian customers must take into account that there have been some unobserved actions that do not belong to T .

¹⁵In Lemma 1 in the Appendix, we show that, given uniform priors on unobserved actions, Eq. (8) holds exactly when users apply Bayes’ rule fully recognizing that there may have been other customers whose actions are not recorded in the summary statistics (e.g., because they have not purchased or have not left reviews).

5.2 Asymptotic Learning

Let us use $G_1(i, q)$ (respectively, $G_0(i, q)$) to denote the probability of $\mathbf{Y}_\tau = \mathbf{e}_i$ given belief q and $Q = 1$ (respectively, $Q = 0$). Therefore,

$$\begin{aligned} G_1(i, q) &= \mathbb{P}[h \in T_i \mid h \in T, q, Q = 1], \quad i \in [m], \\ G_0(i, q) &= \mathbb{P}[h \in T_i \mid h \in T, q, Q = 0], \quad i \in [m], \end{aligned} \tag{10}$$

where $[m]$ denotes the set $\{1, \dots, m\}$. Since T_1, \dots, T_m form a partition of T , for any q , we have $\sum_{i=1}^m G_1(i, q) = \sum_{i=1}^m G_0(i, q) = 1$.

As explained in the previous section, the selection effect becomes more difficult to handle without full history. Recall that with full history, a customer knows the public belief at each point in the past, and thus can work out the average ex ante valuation of those who purchased at that point, enabling her to correct for the selection effect. This is no longer possible with a rating system that provides summary statistics, because a customer does not know the belief of previous customers. Nevertheless, we will show that asymptotic learning still obtains provided that the following additional assumption is imposed.

Assumption 3 (Separation). A rating system (T_1, \dots, T_m) has the separation property if there exists $i \in [m]$ such that ranges of $G_1(i, \cdot)$ and $G_0(i, \cdot)$ are separated, i.e., either

$$\min_{q \in [0,1]} G_1(i, q) > \max_{q \in [0,1]} G_0(i, q), \text{ or } \min_{q \in [0,1]} G_0(i, q) > \max_{q \in [0,1]} G_1(i, q).$$

The separation property ensures that, whatever the value of the belief q , the asymptotic distribution of at least one dimension of the summary statistics has support when the underlying quality is high that is everywhere above (or below) the support when the underlying quality is low. This separation property enables customers to correctly identify the underlying quality of the product despite the selection effect.

Theorem 3. *Suppose Assumptions 1, 2, and 3 hold. Then, there is asymptotic learning, i.e., $q_\tau \rightarrow Q^*$ almost surely.*

Intuitively, given a summary statistic \mathbf{S}_τ , customers form their beliefs based on the likelihood of \mathbf{S}_τ given $Q = 0$ and $Q = 1$. If the probability distributions of summary statistics given $Q = 0$ and $Q = 1$ both assign positive probabilities to the same events, then conditional on the realization of these events (e.g., certain fractions of reviews), it would be impossible for customers to learn the underlying state almost surely. Conversely, consider the case in which these two probability distributions are separated. Specifically, without loss of generality, suppose that for some $i \in [m]$, $G_1(i, \cdot)$ is above $G_0(i, \cdot)$. As τ grows, the marginal distributions of $\mathbf{S}_\tau(i)$ conditional on $Q = 0$ and $Q = 1$ overlap with smaller and smaller probabilities, and depending on the value of $\mathbf{S}_\tau(i)$ as $\tau \rightarrow \infty$, customers can learn the underlying state almost surely.

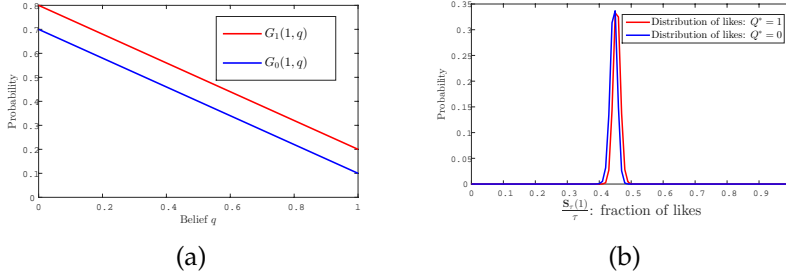


Figure 2: (a) $G_1(1, q)$ and $G_0(1, q)$ as a function of $q \in [0, 1]$ (b) distribution of fraction of “likes” for $Q = 1$ and $Q = 0$ for $\tau = 100$.

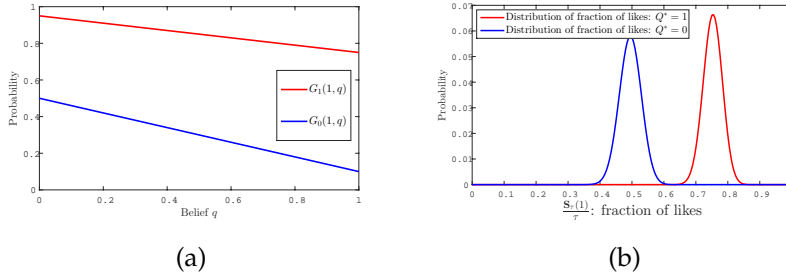


Figure 3: (a) $G_1(1, q)$ and $G_0(1, q)$ as a function of $q \in [0, 1]$ (b) distribution of fraction of “likes” for $Q = 1$ and $Q = 0$ for $\tau = 100$.

We will next show via an example that if the “separation assumption” does not hold, then asymptotic learning fails.

Example 1. We consider a rating system that reports the fraction of “likes” among all reviews, i.e., $T = \{-K_1, \dots, K_2\} \setminus \{0\}$, $T_1 = \{K_2\}$, and $T_2 = T \setminus T_1$. The probability distributions of this summary statistics conditional on the underlying state Q and belief q , $G_1(1, q)$ and $G_0(1, q)$ are shown in Figure 2a, which illustrates that these distributions are not separated (and neither are $G_1(2, q) = 1 - G_1(1, q)$ and $G_0(2, q) = 1 - G_0(1, q)$). In the Appendix, we show that asymptotic learning does not occur in this example: given the lack of separation, the likelihood ratio of S_τ given belief q and $Q = 0$ vs. $Q = 1$ remains bounded away from 0 or ∞ as shown in Figure 2b. In contrast, $G_1(1, q)$ and $G_0(1, q)$ in Figure 3a are separated and as illustrated in Figure 3b, the distribution of fraction of “likes” among all reviews given $Q = 0$ and $Q = 1$ are also separated, and consequently, there is asymptotic learning both when $Q = 0$ and when $Q = 1$.

5.3 Speed of Learning

We next characterize the asymptotic speed of learning under rating systems with summary statistics. As we have seen, ensuring asymptotic learning requires additional conditions under summary statistics. However, the next theorem shows that conditional on asymptotic learning, the speed of learning is governed by a related KL divergence.

Theorem 4. For a given rating system (T_1, \dots, T_m) , suppose Assumptions 1, 2, and 3 hold. Then, for $Q^* = 0$, we almost surely have

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log q_\tau = -D(\mathbb{P}[\cdot \mid h \in T, q = 0, Q = 0] \parallel \mathbb{P}[\cdot \mid h \in T, q = 1, Q = 1])$$

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

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log(1 - q_\tau) = -D(\mathbb{P}[\cdot \mid h \in T, q = 1, Q = 1] \parallel \mathbb{P}[\cdot \mid h \in T, q = 0, Q = 0]).$$

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. While under full history, the probability distribution under the alternative hypothesis $Q = 1 - Q^*$, was still conditioned on $q = Q^*$, it is now conditioned on $q = 1 - Q^*$. This again reflects the role of the selection effect. Because under full history, customers know the public belief at each point, they can fully correct for the selection effect, and this enables them to condition on the correct public belief. Here, this is not possible. In particular, the implicit assumption has to be that if $Q = 1 - Q^*$, then the belief of all other customers, which is not observed, would have also converged to $q = 1 - Q^*$. This difference explains the exact form of the KL divergence in the previous theorem.

6 Comparison of Rating Systems

In this section, we compare the speed of learning under full history and summary statistics; we show how varying certain properties of the rating system affects the speed of learning; we provide examples of fast and slow learning; and finally, we show how the platform has a natural incentive for choosing rating systems that ensure fast learning.

6.1 Full History versus Summary Statistics

In this subsection, we show that the speed of learning could be greater under rating systems with either full history or summary statistics. For simplicity, we focus on a one-dimensional summary statistics, i.e., those that report the fraction of one type of review among either all reviews or all customers. Whether asymptotic learning under full history is faster will turn out to depend on whether the probability of positive reviews under a rating system with summary statistics exhibits *negative selection* or *positive selection*. Negative selection is what we have focused on so far. It corresponds to the case where as the public belief becomes more favorable to $Q = 1$, the ex ante valuations of customers become more “negatively selected” (more likely to be negative). Positive selection, conversely, corresponds to the case where as the public belief becomes more favorable to $Q = 1$, the ex ante valuations of customers become more “positively selected”. Though negative selection is the more natural case (because with more favorable reviews about the quality of the product, customers with lower ex ante valuations also purchase it), we illustrate that positive

selection can emerge in some rating systems. The next two corollaries show that with negative selection, rating systems with summary statistics lead to slower learning than full history, but with positive selection, they may entail faster learning. For simplicity, in these two corollaries, we focus on rating systems with only three review options: “dislike” ($h = -K_1$), “no review” ($h = 0$), and “like” ($h = K_2$), and compare the speed of learning under full history and summary statistics when customers have access to these three review options.

Corollary 1 (negative selection). *Suppose Assumptions 1, 2, and 3 hold and $Q^* = 0$ (the case where $Q^* = 1$ is analogous). Consider a rating system that reports the fraction of “likes among reviews”, i.e., $\mathbf{S}_\tau(1) = \frac{1}{\tau} \sum_{s=1}^t \mathbf{1}\{h_s = K_2\}$ with τ corresponding to the number of reviews. Then*

$$\mathbb{P}[h = K_2 \mid h \in \{-K_1, K_2\}, q = 1, Q = 1] < \mathbb{P}[h = K_2 \mid h \in \{-K_1, K_2\}, q = 0, Q = 1], \quad (11)$$

and the speed of learning under this rating system with summary statistics is slower than under full history.

The rating systems considered in Corollary 1 exhibit negative selection precisely because $\mathbb{P}[h = K_2 \mid h \in \{-K_1, K_2\}, q, Q]$ is decreasing in q (as well as being increasing in Q). This reflects the fact that as q increases, the marginal type that purchases has lower θ , and is thus less likely to gain a very high overall satisfaction leading to a positive (“like”) review. We next discuss why learning is slower in this case with summary statistics than under full history. Suppose $Q^* = 0$, and recall that with summary statistics the inference problem of customers is to distinguish the probability distribution of reviews conditional on $q = 1$ and $Q = 1$ (the “alternative hypothesis”) from the probability distribution of reviews conditional on $q = 0$ and $Q = 0$ (the “null hypothesis”). In contrast, with full history they must distinguish the probability distributions conditional on $q = 0$ and $Q = 1$ and conditional on $q = 0$ and $Q = 0$ (recall Theorem 2). But then inequality (11), which follows from the fact that with negative selection $\mathbb{P}[h = K_2 \mid h \in \{-K_1, K_2\}, q, Q]$ is decreasing in q , implies that this inference is more difficult with summary statistics than with full history. This comparison is at the root of the faster rate of learning under full history.

The next corollary, however, shows that the opposite result is possible if we consider a rating system with “likes among all (potential) customers”, which induces positive selection.

Corollary 2 (positive selection). *Suppose Assumptions 1, 2, and 3 hold and $Q^* = 0$ (the case where $Q^* = 1$ is analogous). Consider a rating system that reports the fraction of “likes among all customers”, i.e., $\mathbf{S}_\tau(1) = \frac{1}{\tau} \sum_{s=1}^t \mathbf{1}\{h_s = K_2\}$ with τ now corresponding to the number of customers who joined the platform. Then*

$$\mathbb{P}[h = K_2 \mid h \in \{-K_1, K_2\}, q = 1, Q = 1] > \mathbb{P}[h = K_2 \mid h \in \{-K_1, K_2\}, q = 0, Q = 1], \quad (12)$$

and the speed of learning under this rating system with summary statistics is faster than under full history.

The intuition for this corollary is similar to that of Corollary 1, with the only difference that now there is positive selection. This is because as q increases, there are more purchases, and

more purchases always increase the number of more “likes” (or more generally, of more positive reviews). But then positive selection implies that the gap between the probability distributions of reviews conditional on $q = 0$ and $Q = 0$ and conditional on $q = 0$ and $Q = 1$ is less than the gap between those conditional on $q = 0$ and $Q = 0$ and conditional on $q = 1$ and $Q = 1$. This means that learning is slower under full history. The simplest intuition is that positive selection heightens the difference between the probability distributions implied by the null and alternative hypotheses.¹⁶

6.2 The Effects of More 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. Given two rating systems Ω and Ω' with the same review options, we say that Ω' is more refined than Ω if the latter reports ratings in a partition T that is (strictly) coarser than the former’s partition T' . For example, this will be the case when Ω reports the fraction of customers that have left one of the most positive two review options, while Ω' separately reports the fractions of customers that have left each one of those two review options. The next proposition is the main result of this subsection.

Theorem 5. *Consider a rating system with either full history or with summary statistics, and suppose that Assumptions 1 and 2 hold, and in the case of summary statistics, suppose also that Assumption 3 holds. Then the speed of learning is always faster under a more refined rating system.*

Intuitively, a more refined rating system always gives more information, and this makes it easier for customers to distinguish between the probability distribution of reviews induced under the true state of nature and the alternative. Importantly, however, this result does not imply that changing the rating system by increasing the number of review options always leads to faster learning. This is because, when the platform changes the review options in the rating system, the thresholds of users might also change (for this reason, the comparison between a more and less refined rating system in Theorem 5 kept the number of review options the same, but modified the reporting strategy of the platform). The next example shows that if increasing the number of review options changes the thresholds, it can lead to slower learning.

Example 2. Consider a rating system with full history. Assume ζ is uniform over $[-2, 2]$, θ is uniform over $[-1, 1]$, and $p = 0$. We consider the following two rating systems:

1. there are two review options “like” and “dislike”, with thresholds $\lambda_{K_2} = \lambda_{K_1} = 0$;
2. there are three review options “like”, “no review”, and “dislike”, with thresholds $\lambda_{K_1} = -1/2$ and $\lambda_{K_2} = 1/2$.

¹⁶It is also straightforward to show that the set of parameters that ensures that the conditions of both Corollaries 1 and 2 is nonempty. For example, we show in Appendix that they are satisfied when we take both ζ and θ to have uniform distributions with mean zero and large enough supports.

We can then use Theorem 1 to characterize the speed of learning in these two cases. We see that even though the second rating system has more review options, it leads to slower learning. This result does not contradict Theorem 5 because the partition of the second rating system, T' , is not strictly finer than that of the first rating system's partition T .

6.3 Average Summary Statistics

Several online platforms report an average score of past reviews rather than depicting detailed fractions of reviews that fall in different categories. In this subsection, we show that this type of average summary statistics leads to slower learning than their detailed counterpart. More formally, consider a rating system with $K \geq 1$ review options:

$$h_\tau = \begin{cases} 1, & \theta_\tau + \mathbb{E}[\zeta_\tau] + q_\tau - p \geq 0, \theta_\tau + \zeta_\tau + Q - p \leq \lambda_1 \\ i, & \theta_\tau + \mathbb{E}[\zeta_\tau] + q_\tau - p \geq 0, \lambda_{i-1} \leq \theta_\tau + \zeta_\tau + Q - p \leq \lambda_i, 2 \leq i \leq K-1 \\ K, & \theta_\tau + \mathbb{E}[\zeta_\tau] + q_\tau - p \geq 0, \theta_\tau + \zeta_\tau + Q - p \geq \lambda_{K-1}. \end{cases}$$

A rating system is said to have an average summary statistics if it reports the number

$$S_\tau = \frac{1}{\tau} \sum_{s=1}^{\tau} h_s,$$

where τ is the number of reviews. Similar to Eq. (10), for $Q \in \{0, 1\}$, let us define

$$G_Q(i, q) = \begin{cases} \mathbb{P}[\theta_\tau + \zeta_\tau + Q - p \leq \lambda_1 \mid \theta + \mathbb{E}[\zeta] + q - p \geq 0], & i = 1, \\ \mathbb{P}[\lambda_{i-1} \leq \theta_\tau + \zeta_\tau + Q - p \leq \lambda_i \mid \theta + \mathbb{E}[\zeta] + q - p \geq 0], & 2 \leq i \leq K-1, \\ \mathbb{P}[\theta_\tau + \zeta_\tau + Q - p \geq \lambda_{K-1} \mid \theta + \mathbb{E}[\zeta] + q - p \geq 0], & i = K. \end{cases}$$

Again, similar to Eq. (9), given true quality $Q^* \in \{0, 1\}$, the governing stochastic process for S_τ becomes

$$S_{\tau+1} = \frac{\tau}{\tau+1} S_\tau + \frac{1}{\tau+1} Y_{\tau+1}, \quad \forall \tau \geq 0,$$

where

$$Y_{\tau+1} = i, \quad \text{w.p. } G_{Q^*}(i, q_\tau), i \in [K].$$

To distinguish the rating systems we have analyzed so far (which report fractions of users that have left various different reviews) from those with average statistics, we refer to them as “rating systems with vector summary statistics”.

Theorem 6. *Suppose Assumptions 1 and 2 hold, and for all $i = 2, \dots, K$, $\sum_{j=i}^K G_1(j, q)$ and $\sum_{j=i}^K G_0(j, q)$ are separated, i.e.,*

$$\min_q \sum_{j=i}^K G_1(j, q) > \max_q \sum_{j=i}^K G_0(j, q), \quad i = 2, \dots, K. \quad (13)$$

Then asymptotic learning occurs under average summary statistics, i.e., $q_\tau \rightarrow Q^*$ almost surely, but the speed of learning is slower under average statistics than with vector summary statistics.

Theorem 6 establishes that the speed of learning for rating systems with average summary statistics is always slower than the comparable rating system with vector summary statistics. The intuition is similar to Theorem 5. Note also that we did not impose Assumption 3, since condition (13), which ensures separation with average summary statistics, is sufficient for Assumption 3 (to see this we can take $i = K$ in condition (13)). This point is illustrated in greater detail in the next example, which shows that asymptotic learning might fail under average summary statistics, while it applies with a rating system with summary statistics that reports fractions of different reviews.

Example 3. We consider a rating system with three review options: “like” (3), “neutral” (2), and “dislike” (1), where the probability of “like” and “dislike” as a function of belief are shown in Figure 4.

1. *Average summary statistic:* the probability of average of reviews at $\tau = 100$ is plotted in Figure 5a. The probability for $Q = 0$ and $Q = 1$ have overlap, showing that with positive probability, learning does not occur. Note that the probabilities plotted in Figure 5b do not satisfy the assumptions of Theorem 6 because for $i = 2$, we have

$$\min_q G_1(2, q) + G_1(3, q) = 1 - \max_q G_1(1, q) < 1 - \min_q G_0(1, q) = \max_q G_0(2, q) + G_0(3, q).$$

2. *Vector summary statistics:* the probability of “like” at $\tau = 100$ is plotted in Figure 5b. The probability for $Q = 0$ and $Q = 1$ are separated, showing that by considering only the fraction of “likes”, we can identify whether $Q = 0$ or $Q = 1$, hence learning occurs.

This implies that for the same distributions of preference parameters (i.e., θ and ζ), asymptotic learning fails for the rating system with average summary statistic, while there continues to be asymptotic learning for the rating system with the same review options but with vector summary statistics.

6.4 Effects of Targeted Information

In addition to the overall ratings of a product, platforms such as Amazon offer information about the reviews of groups of customers with certain characteristics. For instance, in reviews of a book at the intersection of climate science and economics, Amazon separately shows reviews among customers who are interested in economics as well as reviews among customers who are interested in climate science. We refer to this type of additional information as “targeted information”.

Our next result shows that targeted information always increases the speed of learning.

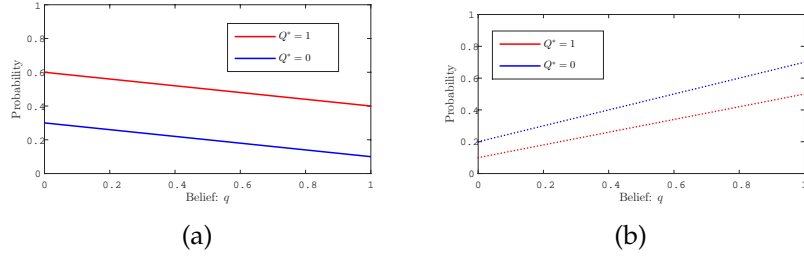


Figure 4: (a) Probability of “like” given belief for $Q^* = 0$ and $Q^* = 1$ which are separated. (b) Probability of “dislike” given belief for $Q^* = 0$ and $Q^* = 1$ which are not separated.

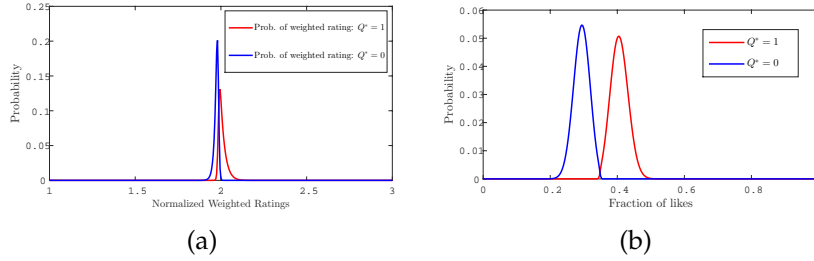


Figure 5: (a) Probability distribution of normalized average ratings for $Q^* = 0$ and $Q^* = 1$ which are not separated (for $\tau = 100$). (b) Probability distribution of fraction of “likes” for $Q^* = 0$ and $Q^* = 1$ which are separated (for $\tau = 100$).

Theorem 7. Consider a rating system with either full history or with summary statistics, and suppose that Assumptions 1 and 2 hold, and in the case of summary statistics, suppose also that Assumption 3 holds. Then the speed of learning is always faster with additional targeted information.

The intuition for faster learning with targeted information is related to the selection effect. Recall that the selection effect makes it more difficult for customers to learn from observed reviews. Targeted information increases customers’ ability to distinguish the true quality from its opposite, because with additional information they can more successfully filter out the selection effect. Note, however, that for this result of faster learning, the targeted information has to be provided in addition to the baseline. If the platform reports only information about the reviews of the subgroup of customers (and suppresses any information about reviews from other subgroups), this can slow down learning relative to the baseline without targeted information.

6.5 Fast and Slow Learning from Reviews

We have seen so far that the speed of rating system for both summary statistics and full history is given by the KL divergence between the probability distribution of reviews under the true quality and its opposite. This implies, for example, that for a rating system with summary statistics, learning is fast when the ratio of probability of a review given $(q = Q^*, Q = Q^*)$ and $(q = 1 - Q^*, Q = 1 - Q^*)$ is high. We next illustrate how the speed of learning depends on the parameters

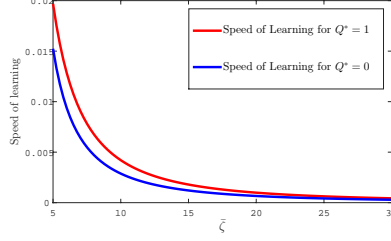


Figure 6: The speed of learning as a function of the support of ζ for Example 4 with $p = 0$, $\bar{\theta} = -\underline{\theta} = 2$, and $\lambda_{K_2} = -\lambda_{-K_1} = 1$.

of the rating system and the distribution of preferences.

In the first example, we show that increasing the support of ex post preference parameter, ζ , can significantly decrease the speed of learning.

Example 4. Suppose θ and ζ have symmetric uniform distributions with diffuse supports such that

$$p \in (-\bar{\theta} + 1, \bar{\theta}), \quad \bar{\zeta} \geq \max\{\lambda_{K_2} + 1, -\lambda_{-K_1} - p + \bar{\theta} + 1\}.$$

From Theorem 2, the speed of learning with full history can be computed as shown in Figure 6. We can see that the speed of learning for both $Q^* = 0$ and $Q^* = 1$ significantly decrease as the support of ζ widens (i.e., as $\bar{\zeta}$ increases). The characterization of the speed of learning in Theorem 2 provides the intuition for this result: when ζ has a diffuse support, the likelihood of each review under both $(q = Q^*, Q = Q^*)$ or $(q = Q^*, Q = 1 - Q^*)$ is similar, and thus the likelihood ratio of each review is close to 1, leading to a low KL divergence. The differences in the speed of learning can be quite substantial. For example, Figure 6 shows that when $\bar{\zeta} = 5$, the speed of learning for $Q^* = 1$ is 0.02, which means that the “half life” of convergence to the true state of nature is 50 periods (meaning that starting from any belief half of the distance to $q = 1$ will be close in 50 periods). In contrast, when $\bar{\zeta}$ is 20, the speed of learning changes to 0.001, and the half life of learning the underlying state is 1000.

The next example shows that a refinement of the review options in the rating system can significantly increase the speed of learning.

Example 5. Suppose that θ has a uniform distribution over $[-2, 2]$, $p = 0$, and ζ has the following symmetric bimodal distribution

$$f_{\zeta}(z) = \begin{cases} \frac{1}{24}, & z \leq -b - \frac{1}{4} \\ \frac{1}{24} + 4(z + b + \frac{1}{4}), & -b - \frac{1}{4} < z \leq -b, \\ \frac{1}{24} + 4(-z - b + \frac{1}{4}), & -b < z \leq -b + \frac{1}{4}, \\ \frac{1}{24}, & -b + \frac{1}{4} < z \leq 0, \end{cases}$$

over the support $[-6, 6]$. This distribution implies that the likelihood of a realization of ζ is very

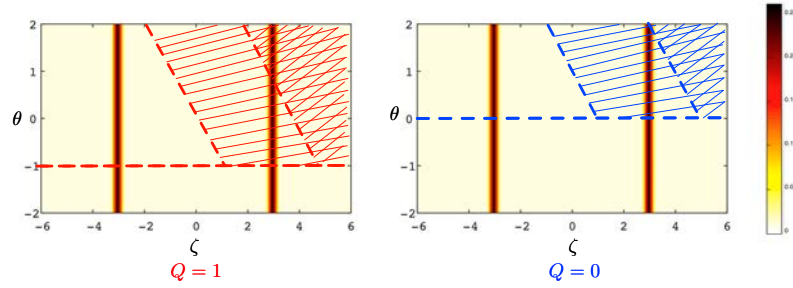


Figure 7: The joint distribution of (θ, ζ) and the probabilities of the review options “like” and “love” for two cases under $Q = 0$ and $Q = 1$ for $b = 3$ for Example 5. The double shaded areas correspond to the “love” option, and the shaded areas correspond to the “like” option.

high around b and $-b$. We next compute the speed of learning of rating system with full history and the following review options:

- **Case 1:** “dislike”, “no review”, and “like” and the thresholds are 1 and -1 .
- **Case 2:** “dislike”, “no review”, “like”, and “love” and the thresholds are 1, -1 , and 5.

The joint distribution of (θ, ζ) and the probabilities of observing the “like” and “love” reviews under $Q = 0$ and $Q = 1$ are depicted in Figure 7. The speed of learning for these two cases as a function of parameter b are, in turn, plotted in Figure 8. In Figure 8, when $b = 5$, the speed of learning without the “love” option is approximately 0.001, which means that the half time over learning the truth would be 1000 periods. In contrast, with the “love” option present, the speed of learning changes to 0.2, equivalent to the half time for learning the truth be 5. The speed of learning of the first case is slow because for each review, the ratio of the probability of observing that review when $Q = 0$ to the probability of observing it when $Q = 1$ is close to 1. In fact, as b increases, this ratio becomes closer to 1 and the speed of learning declines further (towards zero). The speed of learning in the presence of the “love” option is shown in Figure 8, and is significantly higher. This is because the probability of observing the “love” review is very close to zero when $Q = 0$, and is positive when $Q = 1$. This leads to a large value of KL divergence and thus to relatively fast learning.¹⁷

6.6 The Choice of Rating System

In this subsection, we show that the speed of learning is highly relevant for platforms. In fact, platforms themselves have an incentive to choose rating systems with faster learning speed. Recall, in

¹⁷Theorem 5 implies that for a given distribution for parameters θ and ζ , the rating system with the highest speed of learning will be the one with the most refined rating system, which means reporting the exact utility of previous customers (i.e., $\theta_t + \zeta_t + Q - p$). For instance, when $Q^* = 0$ and under full history, the highest rate of learning is $D(\mathbb{P}[\theta + \zeta - p \mid \theta + \mathbb{E}[\zeta] - p \geq 0] \parallel \mathbb{P}[\theta + \zeta - p + 1 \mid \theta + \mathbb{E}[\zeta] - p \geq 0])$, which is 0.55 in Example 5.

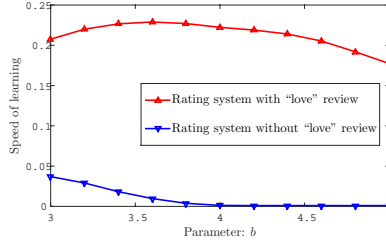


Figure 8: The speed of learning as a function of parameter b for Example 5. As b increases, the speed of learning of the rating system without the “love” option becomes very small, while the speed of learning of the rating system with the “love” option remains large.

particular, that in our model the platform’s revenues come from advertisements or other benefits that it derives from potential customers joining the platform. For example, the platform may be receiving a fixed revenue for each customer that joins. Consequently, we can write the platform’s problem as maximizing the expected number of customers that join the platform. We next show that this problem is equivalent to choosing a rating system with a higher speed of learning. We say rating system Ω_t^I has a higher learning speed than rating system Ω_t^{II} , if

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \log q_t^I &\leq \lim_{t \rightarrow \infty} \frac{1}{t} \log q_t^{II}, \quad \text{given } Q^* = 0, \\ \lim_{t \rightarrow \infty} \frac{1}{t} \log (1 - q_t^I) &\leq \lim_{t \rightarrow \infty} \frac{1}{t} \log (1 - q_t^{II}), \quad \text{given } Q^* = 1, \end{aligned}$$

where $q_t^I = \mathbb{P}[Q = 1 \mid \Omega_t^I]$ and $q_t^{II} = \mathbb{P}[Q = 1 \mid \Omega_t^{II}]$.

Theorem 8. *Given two rating systems I and II , the one with higher learning speed has a higher revenue for the platform.*

The main result in this theorem is intuitive: with a rating system that guarantees a faster learning speed, potential customers know that once they join the platform, they are less likely to make both type I and type II errors in their purchase decisions. This increases their ex ante utility from joining, and thus increases the expected revenues of the platform. To see why the probabilities of type I and type II errors go down, let us consider separately the case of low and high quality. Suppose $Q^* = 0$. Then the problem for potential customers is to incorrectly buy the product thinking that it is high-quality. In particular, if $\theta + \zeta - p \in (-1, 0)$, then with probability $\mathbb{P}_{t, \Omega_t} [q_t \geq p - \theta - \zeta \mid Q = 0]$, they purchase the product and suffer a loss of $\theta + \zeta - p$. Yet, this probability of incorrectly purchasing the product becomes smaller when q_t is closer to $Q^* = 0$. Conversely, suppose that $Q^* = 1$. Then the problem for potential customers is to fail the purchase of the product when they should be buying. In particular, if $\theta + \zeta - p \in (-1, 0)$, then with probability $\mathbb{P}_{t, \Omega_t} [q_t \leq p - \theta - \zeta \mid Q = 1]$ customers do not purchase the product and forgo a potential utility gain of $1 + \theta + \zeta - p$. The probability of such a mistake becomes smaller when q_t is closer to $Q^* = 1$. Thus in both cases a faster learning speed implies smaller expected probabilities of mistakes, encouraging customers to join the platform.

7 Concluding Remarks

As the amount of goods and services sold online continues to increase rapidly, 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 have also become critical. Despite their importance, properties and efficacy of online rating systems have attracted only limited attention in the recent literature.

In this paper, we developed a model of Bayesian learning from online reviews, and investigated the conditions for asymptotic learning of 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. Prominent examples of such rating systems are those that provide average scores or the distribution of numerical ratings left by past customers.

In our model, a sequence of potential customers decide whether to join a platform and contemplate the purchase of a product with unknown quality, which is assumed to be binary, corresponding to either high or low quality. After joining and observing the ratings of the product, customers observe their *ex ante* valuation for the product, and then decide whether or not to purchase it. If they purchase, the true quality of the product, their *ex ante* valuation for the product (which are heterogeneous), an *ex post* idiosyncratic preference term and the price of the product determine their overall satisfaction. Finally, given the rating system of the platform, a customer decides to leave a review as a function of her overall satisfaction. In most of our analysis, we took the thresholds that determine which type of review (from the set of available reviews) a customer will leave to be exogenous.

In our formal analysis, we distinguished 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. We argued that in both cases, learning dynamics are complicated by a *selection effect* — the types of customers who purchase the good and thus their overall satisfaction and reviews depend on the information that they have available at the time of their purchase. For example, only those who have a relatively high *ex ante* valuation for the product would purchase it after seeing fairly negative reviews of its underlying quality. But this implies that this selected group of purchasers are much more likely to leave positive reviews than a random customer. Unless corrected, this selection effect would imply that the information of past customers about the quality of the product cannot be aggregated accurately.

Our formal analysis provides conditions for asymptotic learning under both full history and summary statistics. These conditions encapsulate how customers deal with the selection problem under different rating systems. Under full history, beliefs about the underlying quality of the product follow a martingale, meaning in particular that each customer can accurately estimate the beliefs at the time that another past customer has made her purchase. As a result, it is feasible to correct for the selection effect, which enables asymptotic learning to take place under relatively mild conditions. Conditional on asymptotic learning, we characterized the speed (rate) of learning, which is crucial for the success of online rating systems — it is not practically relevant to

have a rating system that aggregates past information but does so extremely slowly. We show that under full history, learning is exponentially fast, and its exact rate is given by the KL divergence between the probability distribution of reviews under the true quality and its opposite. The emergence of the KL divergence as the measure of distance between these two probability distributions is intuitive, because asymptotically, the problem of estimating the quality of the product resembles a binary hypothesis testing problem, though with sample observations that are correlated with each other because of the selection effect.

We also show how the selection effect becomes more difficult to filter out with summary statistics, because now customers do not know what type of information past users had when they made their decisions. Nevertheless, we provide conditions for asymptotic learning in this case, and demonstrate that conditional on asymptotic learning, the rate of learning is still exponential and its exact rate, though different than the full history case, is still given by a KL divergence term.

The last part of the paper studied how different properties of the rating system influence the rate of learning. We showed that providing more information does not always lead to faster learning. In particular, despite its greater success in filtering out the selection effect, having access to full history can lead to slower learning in cases where the selection effect, paradoxically, heightens the gap between the probability distributions of reviews that will arise under the true quality and its opposite. Nevertheless, we also established several comparative statics that increase the speed of learning from reasonable modifications of a rating system. First, allowing for strictly finer ratings leads to faster learning, provided that this does not shift the thresholds of customers (which would otherwise make the two information structures non-nested). Second, we showed that moving from a rating system that provides averages of previous reviews to one that provides fractions of users that have left different reviews always increases the rate of learning. Third, we demonstrated that providing more information about the characteristics of reviewers (for example, supplementing existing summary statistics with summary statistics among users with similar profiles to the current customer) also always increases the rate of learning. We also showed that different rating systems can lead to very fast or very slow speeds of learning.

In all these cases, it is also worth noting that the platform's incentives are often aligned with increasing the speed of learning of customers. We demonstrated this by showing that, in our baseline model, the revenues of the platform are higher under a rating system that leads to a faster rate of learning.

We view our paper as a first step in a comprehensive theoretical analysis of learning from online rating systems. Several simplifying assumptions enabled us to obtain sharp results, but need to be relaxed in future work to gain a more holistic understanding of the performance of such systems. We briefly discuss some of the important ones here.

First, our analysis focused on Bayesian learning. We believe that the intuition that comes from Bayesian learning is often relevant even if agents are not fully Bayesian. Nevertheless, it is important to also study how the presence of (some) non-Bayesian agents modifies these conclusions. The selection effect implies that non-Bayesians will not be able to aggregate information from past

customers accurately. But in our baseline model, the presence of some non-Bayesians does not change the ability of Bayesian agents to learn.

Second, and more fundamentally, throughout we took the thresholds agents use for mapping from their overall satisfaction into reviews as fixed and exogenous. As commented before, making these thresholds random (with independently distributed increments) has no effect on our analysis. But it would be more challenging to endogenize these thresholds, which would necessitate a systematic theory of why customers leave reviews and how these motives change as the choices of the platform, including the rating system, change. One possibility would be an expressive motive (e.g., they may wish to complain if their overall satisfaction is low). With an expressive motive, endogenizing these thresholds is not difficult, though it is unclear whether as the rating system is modified, these expressive motives change as well. The alternative is a strategic motive, where customers leave reviews in order to influence the behavior of future customers. This is an exciting and challenging area for research, in part because with a strategic motive, the way in which customers leave reviews will change as the rating system is modified. This of course complicates the comparisons of the speed of learning across different rating systems.

Third, one could allow for a richer preference structure. This could include moving away from a setting in which there is a common quality. For example, a high-quality book about global warming is likely to yield high utility for many customers, but to may also lead to low satisfaction to those who do not believe in man-made climate change). It could also move in the direction of imperfect information about satisfaction, which would make users still rely on the information they have received from the rating system in choosing their reviews, thus introducing issues related to herding as studied in the previous literature.

Finally, an exciting research area is to close the gap between the theoretical models of learning from online reviews and the burgeoning empirical literature on behavior in online markets.

Appendix: Proofs

Proof of Theorem 1

At time t , the likelihood ratio is a random variable defined as

$$Z(\cdot | l_t) \triangleq \frac{\mathbb{P}[\cdot | Q = 1, l_t]}{\mathbb{P}[\cdot | Q = 0, l_t]},$$

where $Z(\cdot | l_t) = Z(h | l_t)$ with probability $\mathbb{P}[h | Q = Q^*, l_t]$. We will use this random variable in the proof of this theorem as well as the proof of Theorem 2.

Part (a): If $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$ holds, then customers almost surely learn the true quality.

Without loss of generality, we assume $Q^* = 0$ and then show $q_t \rightarrow 0$ almost surely. The proof for $Q^* = 1$ is similar and can be obtained by analyzing $\frac{\mathbb{P}[Q=0|\Omega_t]}{\mathbb{P}[Q=1|\Omega_t]}$. We first show 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 $Z(\cdot | l_t)$ are all mean 1 (conditioned on the history). This is because we have

$$\mathbb{E}[Z(\cdot | l_{t+1}) | Z(\cdot | l_1), \dots, Z(\cdot | l_t)] = \sum_h \mathbb{P}[h | Q = 0, l_t] \frac{\mathbb{P}[h | Q = 1, l_t]}{\mathbb{P}[h | Q = 0, l_t]} = \sum_h \mathbb{P}[h | Q = 1, l_t] = 1.$$

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 $l_t \rightarrow l_\infty$ almost surely. We next show that the limiting random variable l_∞ is 0 almost surely.

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

$$\begin{aligned} Z(h = K_2 | l_t) &= \frac{\mathbb{P}[h = K_2 | Q = 1, l_t]}{\mathbb{P}[h = K_2 | Q = 0, l_t]} = \frac{\mathbb{P}[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta + 1 - p \geq \lambda_{K_2}]}{\mathbb{P}[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{K_2}]} \\ &= 1 + \frac{\mathbb{P}[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{K_2} \in [-1, 0)]}{\mathbb{P}[q_t + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{K_2}]} \\ &\geq 1 + \frac{\mathbb{P}[\theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{K_2} \in [-1, 0)]}{\mathbb{P}[1 + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{K_2}]}, \end{aligned}$$

where the inequality follows by substituting $q_t = 0$ in numerator and $q_t = 1$ in the denominator. Using condition $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$ of part (a) and Assumption 1, we have

$$\epsilon \triangleq \frac{\mathbb{P}[\theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p - \lambda_{K_2} \in [-1, 0)]}{\mathbb{P}[1 + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{K_2}]} > 0.$$

Therefore, irrespective of the belief (or corresponding likelihood ratio l), whenever $h = K_2$, the random variable $Z(\cdot | l)$ is strictly larger than 1, i.e.,

$$\min_l Z(h = K_2 | l) - 1 \geq \epsilon > 0. \quad (14)$$

Again, using Assumption 1 and $\bar{\theta} + \mathbb{E}[\zeta] - p \geq 0$, for any belief probability of $h = K_2$ is positive, i.e.,

$$\eta \triangleq \min_{q \in [0, 1]} \mathbb{P}[q + \theta + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p \geq \lambda_{K_2}] > 0.$$

With these definitions for ϵ and η , for all t and l_t , we have

$$\mathbb{P}[|Z(\cdot | l_t) - 1| \geq \epsilon | l_t] = \mathbb{P}[h = K_2 | l_t] \geq \eta.$$

We next show that $l_\infty = 0$ with probability 1. Using Eq. (14), for an arbitrary $\delta > 0$, we can write

$$\begin{aligned} \mathbb{P}[|l_{t+1} - l_t| \geq \delta \epsilon] &= \mathbb{E}[\mathbf{1}\{|l_t(Z(\cdot | l_t) - 1)| \geq \delta \epsilon\}] \geq \mathbb{E}[\mathbf{1}\{l_t \geq \delta\} \mathbf{1}\{|Z(\cdot | l_t) - 1| \geq \epsilon\}] \\ &= \mathbb{E}_{l_t}[\mathbf{1}\{l_t \geq \delta\} \mathbb{E}[\mathbf{1}\{|Z(\cdot | l_t) - 1| \geq \epsilon\} | l_t]] \\ &= \mathbb{E}_{l_t}[\mathbf{1}\{l_t \geq \delta\} \mathbb{P}[|Z(\cdot | l_t) - 1| \geq \epsilon | l_t]] \geq \eta \mathbb{E}_{l_t}[\mathbf{1}\{l_t \geq \delta\}] = \eta \mathbb{P}[l_t > \delta]. \end{aligned} \quad (15)$$

Since $l_t \rightarrow l_\infty$ almost surely, we have $\mathbb{P}[|l_{t+1} - l_t| \geq \delta \epsilon] \rightarrow 0$ which along with Eq. (15) leads to

$\mathbb{P}[l_t > \delta] \rightarrow 0$. This shows that $l_t \rightarrow 0$ in probability, which together with $l_t \rightarrow l_\infty$ almost surely, establishes that $\mathbb{P}[l_\infty = 0] = 1$.¹⁸

Finally, note that since $l_t = \frac{q_t}{1-q_t}$, from $l_t \rightarrow 0$, we have $q_t \rightarrow 0$ almost surely.

Part (b): If $\bar{\theta} + \mathbb{E}[\zeta] - p < 0$, then with positive probability customers do not learn the true quality. We break the proof into two steps. In the first step we show that with a positive probability in finite time q_t becomes very small. In the second step we show that once this happens no purchase takes place and learning stops.

Step 1: There exists a finite t such that with positive probability we have $q_t < \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}$.

First note that from Assumption 1 and using a similar argument to that of Eq. (14), we have

$$\max_{q \in [0,1]} \frac{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta + 1 - p \leq \lambda_{-K_1}]}{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-K_1}]} < 1.$$

Therefore, we obtain

$$\lim_{t \rightarrow \infty} \left(\max_{q \geq \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}} \left(\frac{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta + 1 - p \leq \lambda_{-K_1}]}{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-K_1}]} \right) \right)^t = 0.$$

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

$$\left(\max_{\min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}} \left(\frac{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta + 1 - p \leq \lambda_{-K_1}]}{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-K_1}]} \right) \right)^t < \frac{\min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}}{1 - \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}}.$$

We also let

$$\rho \triangleq \left(\min_{q \geq \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}} \left(\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \leq \lambda_{-K_1}] \right) \right)^{T_0} > 0.$$

We next show that with probability at least ρ there exists a time $t \in [1, T_0]$ such that

$$q_t \leq \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}.$$

We let \mathcal{E} be the event that in the interval $[1, T_0]$, the belief goes below $\min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}$ and \mathcal{E}_1 be the event that in the interval $[1, T_0 - 1]$, the belief goes below $\min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}$. We can write

$$\begin{aligned} \mathbb{P}[\mathcal{E}] &= \mathbb{P}[\mathcal{E} \mid \mathcal{E}_1] \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E} \mid \mathcal{E}_1^c] \mathbb{P}[\mathcal{E}_1^c] = \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_1^c] \times \\ \mathbb{P} \left[q_{T_0} \leq \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\} \mid q_t \geq \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}, 1 \leq t \leq T_0 - 1 \right] &\geq \mathbb{P}[\mathcal{E}_1] + \rho \mathbb{P}[\mathcal{E}_1^c] \geq \rho. \end{aligned}$$

¹⁸To see this, first note that from $l_t \rightarrow l_\infty$ almost surely, we have $l_t \rightarrow l_\infty$ in probability and the result follows by noting that for a sequence of random variables $\{X_n\}$, if $X_n \rightarrow X$ and $X_n \rightarrow Y$ in probability, then $X = Y$ almost surely.

Therefore, with probability $\rho > 0$, there exists $t^* \in [1, T_0]$ such that $q_{t^*} < \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\}$.

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

We show this claim by induction on t . It evidently holds for $t = t^*$. Since

$$q_{t^*} - p + \bar{\theta} + \mathbb{E}[\zeta] \leq \min \left\{ \frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2}, 1 \right\} - p + \bar{\theta} + \mathbb{E}[\zeta] = -\frac{p - \bar{\theta} - \mathbb{E}[\zeta]}{2} < 0,$$

purchase does not happen at time t^* . Now that purchase does not happen at time t^* , the belief at time $t^* + 1$ is the same as q_{t^*} because

$$\frac{\mathbb{P}[h = N \mid Q = 1, l_{t^*}]}{\mathbb{P}[h = N \mid Q = 0, l_{t^*}]} = \frac{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q_{t^*} - p < 0]}{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q_{t^*} - p < 0]} = 1.$$

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 $Q^* = 0$ (and similarly from $Q^* = 1$) because using Assumption 1 if purchase occurs, then probability of any review is non-zero. Hence, the likelihood ratio at each time is multiplied by a non-zero number (bounded number) and hence cannot become 0 in finite time.

Proof of Proposition 1

We show the first part of proposition as the proof of the second part follows from an identical argument. We first show that Eq. (5) leads to

$$\mathbb{E}_{h \sim \mathbb{P}[\cdot \mid q=0, Q=0]} \left[\log \left(\frac{\mathbb{P}[h \mid q = \bar{q}, Q = 1]}{\mathbb{P}[h \mid q = \bar{q}, Q = 0]} \right) \right] > 0. \quad (16)$$

This is because using inequality $\log x \leq x - 1$ yields

$$\begin{aligned} \mathbb{E}_{h \sim \mathbb{P}[\cdot \mid q=0, Q=0]} \left[\log \left(\frac{\mathbb{P}[h \mid q = \bar{q}, Q = 1]}{\mathbb{P}[h \mid q = \bar{q}, Q = 0]} \right) \right] &= \mathbb{E}_{h \sim \mathbb{P}[\cdot \mid q=0, Q=0]} \left[-\log \left(\frac{\mathbb{P}[h \mid q = \bar{q}, Q = 0]}{\mathbb{P}[h \mid q = \bar{q}, Q = 1]} \right) \right] \\ &\geq \mathbb{E}_{h \sim \mathbb{P}[\cdot \mid q=0, Q=0]} \left[1 - \frac{\mathbb{P}[h \mid q = \bar{q}, Q = 0]}{\mathbb{P}[h \mid q = \bar{q}, Q = 1]} \right] > 0, \end{aligned}$$

where we used Eq. (5) in the last inequality. Using Eq. (16) and the fact that probability of actions are continuous (Assumption 1), we can find $\epsilon > 0$ such that for all $q_h \leq \epsilon$, $h \in \{N\} \cup \{-K_1, \dots, K_2\}$,

$$\sum_{h \in \{-K_1, \dots, K_2\} \cup \{N\}} \mathbb{P}[h \mid q_h, Q = 0] \log \left(\frac{\mathbb{P}[h \mid q = \bar{q}, Q = 1]}{\mathbb{P}[h \mid q = \bar{q}, Q = 0]} \right) > 0.$$

If $q_t \rightarrow 0$ almost surely, then there exists N such that for all $t \geq N$ we have $q_t \leq \epsilon$. Since

$$\frac{q_t}{1 - q_t} = \frac{q_N}{1 - q_N} \prod_{s=N+1}^t \frac{\mathbb{P}[h_s \mid q = \bar{q}, Q = 1]}{\mathbb{P}[h_s \mid q = \bar{q}, Q = 0]}, \quad (17)$$

using strong law of large numbers, we obtain

$$\begin{aligned} & \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{h \in \{-K_1, \dots, K_2\} \cup \{N\}} \log \left(\left(\frac{\mathbb{P}[h | q = \bar{q}, Q = 1]}{\mathbb{P}[h | q = \bar{q}, Q = 0]} \right)^{\sum_{s=N}^t \mathbf{1}\{h_s=h\}} \right) \\ & \geq \sum_{h \in \{-K_1, \dots, K_2\} \cup \{N\}} \left(\min_{q \leq \epsilon} \mathbb{P}[h | q, Q = 0] \right) \log \left(\frac{\mathbb{P}[h | q = \bar{q}, Q = 1]}{\mathbb{P}[h | q = \bar{q}, Q = 0]} \right). \end{aligned}$$

Therefore, taking limit of both sides of Eq. (17) as $t \rightarrow \infty$ we obtain

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \left(\frac{q_t}{1 - q_t} \right) \geq \sum_{h \in \{-K_1, \dots, K_2\} \cup \{N\}} \left(\min_{q \leq \epsilon} \mathbb{P}[h | q, Q = 0] \right) \log \left(\frac{\mathbb{P}[h | q = \bar{q}, Q = 1]}{\mathbb{P}[h | q = \bar{q}, Q = 0]} \right) > 0.$$

This is a contradiction because the left hand side is negative ($q_t \rightarrow 0$) and the right hand side is positive.

Proof of Theorem 2

We prove the theorem for $Q^* = 0$ as the proof for $Q^* = 1$ can be obtained by a similar argument. We re-index the set of actions and denote it by $\mathcal{A} = \{1, \dots, |\mathcal{A}|\}$. Based on the sequence l_t , we define a coupled new sequence \bar{l}_t that is larger than l_t and updates with iid increments. Since $q_t \rightarrow 0$ almost surely (equivalently $l_t \rightarrow 0$ almost surely), for any ϵ we can choose N such that $l_t \leq \epsilon$ for $t \geq N$. Using Assumption 1, for any $a \in \mathcal{A}$, letting $\epsilon_a = \arg \max_{l \in [0, \epsilon]} Z(a | l)$ leads to

$$Z(a | l_t) \leq Z(a | l = \epsilon_a), \quad \forall a \in \mathcal{A}, t \geq N. \quad (18)$$

Without loss of generality we suppose

$$0 \leq Z(a = 1 | l = \epsilon_1) \leq \dots \leq Z(a = |\mathcal{A}| | l = \epsilon_{|\mathcal{A}|}). \quad (19)$$

Using Assumption 1 once again, we can choose $p_1, \dots, p_{|\mathcal{A}|}$ such that

$$\max_{a \in \mathcal{A}} |p_a - \mathbb{P}[a | l = 0, Q = 0]| \leq \epsilon, \quad \sum_{a \in \mathcal{A}} p_a \log(Z(a | l = \epsilon_a)) < 0, \quad (20)$$

and

$$\sum_{a=1}^m \mathbb{P}[a | l_t, Q = 0] \geq \sum_{a=1}^m p_a, \quad \forall m \in \mathcal{A}, \quad \text{for } t \geq N. \quad (21)$$

We next define the sequence \bar{l}_t for $t \geq N$.

- For $t = N$: we let $\bar{l}_N = l_N$.

- For $t > N$: we will define $\{\bar{l}_t\}$ which couples with $\{l_t\}$ through a sequence of uniform iid random variables $\{\sigma_t\}$. Recall that $l_{t+1} = l_t Z(a | l_t)$ with probability $\mathbb{P}[a | l_t, Q = 0]$. We let

$$\begin{aligned} l_{t+1} &= l_t Z(a | l_t), & \text{if } \sigma_t \in \left(\sum_{i=1}^{a-1} \mathbb{P}[i | l_t, Q = 0], \sum_{i=1}^a \mathbb{P}[i | l_t, Q = 0] \right] \\ \bar{l}_{t+1} &= \bar{l}_t Z(\bar{a} | l = \epsilon_a), & \text{if } \sigma_t \in \left(\sum_{i=1}^{\bar{a}-1} p_i, \sum_{i=1}^{\bar{a}} p_i \right]. \end{aligned}$$

Note that by definition, the sequence $\{l_t\}$ defined above evolves as the likelihood ratio of the public belief. This is because we have

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

Using strong law of large numbers (Durrett [2010, Theorem 2.3.5]) because $\{\sigma_t\}$ is an iid sequence, we almost surely have

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

Using Eq. (20), this establishes that \bar{l}_t converges linearly to zero with learning speed $-\sum_{a \in \mathcal{A}} p_a \log Z(a | l = \epsilon_a)$. We next show that \bar{l}_t is larger than l_t . We show this by induction on t . It evidently holds for $t = N$. Suppose at time t , we have $\sigma_t \in \left(\sum_{i=1}^{a-1} \mathbb{P}[i | l_t, Q = 0], \sum_{i=1}^a \mathbb{P}[i | l_t, Q = 0] \right]$, i.e., $a_t = a$. Using Eq. (21), this leads to $\bar{a}_t \geq a$. Therefore, we have

$$\begin{aligned} \bar{l}_{t+1} &= Z(\bar{a}_t | l = \epsilon_{\bar{a}_t}) \bar{l}_t \geq Z(\bar{a}_t | l = \epsilon_{\bar{a}_t}) l_t && \text{By induction hypothesis} \\ &\geq Z(a | l = \epsilon_a) l_t && \text{By } \bar{a}_t \geq a \text{ and Eq. (19)} \\ &\geq Z(a | l_t) l_t && \text{By Eq. (18)} \\ &= l_{t+1}. && \text{By definition of } \{l_t\}. \end{aligned}$$

This shows that almost surely

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log l_t \leq \sum_{a \in \mathcal{A}} p_a \log(Z(a | l = \epsilon_a)).$$

Since this inequality holds for all ϵ , letting $\epsilon \rightarrow 0$ (and consequently $\epsilon_a \rightarrow 0$ and $p_a \rightarrow \mathbb{P}[a | l = 0, Q = 0]$)

for all $a \in \mathcal{A}$) leads to

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \log l_t &\leq \lim_{\epsilon \rightarrow 0} \sum_{a \in \mathcal{A}} p_a \log (Z(a | l = \epsilon_a)) = \sum_{a \in \mathcal{A}} \mathbb{P}[a | l = 0, Q = 0] \log (Z(a | l = 0)) \\ &= -D(\mathbb{P}[\cdot | q = 0, Q = 0] \| \mathbb{P}[\cdot | q = 0, Q = 1]). \end{aligned} \quad (22)$$

Similarly, by reversing the inequalities in Eq. (18) and Eq. (21), we can construct another sequence l_t with independent increments which is smaller than l_t , leading to

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log l_t \geq -D(\mathbb{P}[\cdot | l = 0, Q = 0] \| \mathbb{P}[\cdot | l = 0, Q = 1]). \quad (23)$$

Putting Eq. (22) and Eq. (23) together leads to

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log l_t = -D(\mathbb{P}[\cdot | l = 0, Q = 0] \| \mathbb{P}[\cdot | l = 0, Q = 1]).$$

Finally, we have $\lim_{t \rightarrow \infty} \frac{1}{t} \log q_t = \lim_{t \rightarrow \infty} \frac{1}{t} \log l_t$, which completes the proof.

Proof of Proposition 2

We present the proof for $Q^* = 0$ (the proof for $Q^* = 1$ follows from the same argument). Recall that τ is the sequence of time at which there was an action in T and $\tilde{H}_\tau = \{\tilde{h}_1, \dots, \tilde{h}_{\tau-1}\}$ denotes the available history after $\tau - 1$ actions in T . Similar to the proof of Theorem 1, we define random variable $\tilde{Z}(\cdot | l_\tau) \triangleq \frac{\mathbb{P}[\cdot | \tilde{h} \in T, Q=1, l_\tau]}{\mathbb{P}[\cdot | \tilde{h} \in T, Q=0, l_\tau]}$. We first show that $l_{\tau+1} = l_\tau \tilde{Z}(\cdot | l_\tau)$, and then we establish l_τ is a martingale. Note that between τ -th action in T and $\tau + 1$ -th action in T , each customer has observed the same history, and therefore has the same belief q_τ . Hence, each has taken an action \tilde{h} that does not belong to T , independently with probability $\mathbb{P}[\tilde{h} \notin T | q_\tau, Q]$. Moreover, customers do not know how many others have joined the platform before them, and instead, by assumption, have a uniform prior over this number. We let $H_{\tau:\tau+1}$ denote the history of actions in between τ -th and $\tau + 1$ -th actions in T . We next present a recursive characterization of the likelihood ratio corresponding to belief $\{q_\tau\}_{\tau=0}^\infty$, exploiting the uniform (improper) prior assumption. We have

$$\begin{aligned} l_{\tau+1} &= \frac{\mathbb{P}[\tilde{H}_\tau, \tilde{h}_\tau | Q = 1]}{\mathbb{P}[\tilde{H}_\tau, \tilde{h}_\tau | Q = 0]} = \frac{\sum_{H_{\tau:\tau+1}} \mathbb{P}[\tilde{H}_\tau, H_{\tau:\tau+1}, \tilde{h}_\tau | Q = 1]}{\sum_{H_{\tau:\tau+1}} \mathbb{P}[\tilde{H}_\tau, H_{\tau:\tau+1}, \tilde{h}_\tau | Q = 0]} \\ &= \frac{\sum_{H_{\tau:\tau+1}} \mathbb{P}[\tilde{H}_\tau | Q = 1] \mathbb{P}[H_{\tau:\tau+1} | \tilde{H}_\tau, Q = 1] \mathbb{P}[\tilde{h}_\tau | \tilde{H}_\tau, H_{\tau:\tau+1}, Q = 1]}{\sum_{H_{\tau:\tau+1}} \mathbb{P}[\tilde{H}_\tau | Q = 0] \mathbb{P}[H_{\tau:\tau+1} | \tilde{H}_\tau, Q = 0] \mathbb{P}[\tilde{h}_\tau | \tilde{H}_\tau, H_{\tau:\tau+1}, Q = 0]} \\ &= l_\tau \times \frac{\sum_{H_{\tau:\tau+1}} \mathbb{P}[H_{\tau:\tau+1} | \tilde{H}_\tau, Q = 1] \mathbb{P}[\tilde{h}_\tau | \tilde{H}_\tau, H_{\tau:\tau+1}, Q = 1]}{\sum_{H_{\tau:\tau+1}} \mathbb{P}[H_{\tau:\tau+1} | \tilde{H}_\tau, Q = 0] \mathbb{P}[\tilde{h}_\tau | \tilde{H}_\tau, H_{\tau:\tau+1}, Q = 0]}. \end{aligned} \quad (24)$$

We next consider $\mathbb{P}[H_{\tau:\tau+1} | \tilde{H}_\tau, Q]$. The randomness in $H_{\tau:\tau+1}$ comes from two sources: (i) the number of customers who joined the platform between τ -th and $\tau + 1$ -th actions in T , and (ii) the

actions they took, which depend on their θ and ζ . We let U denote the first random variable. By assumption, this random variable is a uniform (improper) over all integers. We also note that all customers in this time interval have observed the same history (i.e., \tilde{H}_τ), formed the same belief (i.e., q_τ), and took an action independently with probability $\mathbb{P}[h \notin T \mid q_\tau, Q]$. Therefore, we have

$$\mathbb{P}[H_{\tau:\tau+1} \mid \tilde{H}_\tau, Q] = \sum_{u=0}^{\infty} \mathbb{P}[H_{\tau:\tau+1} \mid U = u, \tilde{H}_\tau, Q] = \sum_{u=0}^{\infty} \mathbb{P}[h \notin T \mid \tilde{H}_\tau, Q]^u. \quad (25)$$

Invoking Eq. (25) in Eq. (24) leads to

$$\begin{aligned} l_{\tau+1} &= l_\tau \frac{\sum_{u=0}^{\infty} \mathbb{P}[h \notin T \mid q_\tau, Q = 1]^u \mathbb{P}[\tilde{h}_\tau \mid q_\tau, Q = 1]}{\sum_{u=0}^{\infty} \mathbb{P}[h \notin T \mid q_\tau, Q = 0]^u \mathbb{P}[\tilde{h}_\tau \mid q_\tau, Q = 0]} = l_\tau \frac{\frac{\mathbb{P}[\tilde{h}_\tau \mid q_\tau, Q = 1]}{1 - \mathbb{P}[h \notin T \mid q_\tau, Q = 1]}}{\frac{\mathbb{P}[\tilde{h}_\tau \mid q_\tau, Q = 0]}{1 - \mathbb{P}[h \notin T \mid q_\tau, Q = 0]}} \\ &= l_\tau \frac{\mathbb{P}[\tilde{h}_\tau \mid \tilde{h}_\tau \in T, q_\tau, Q = 1]}{\mathbb{P}[\tilde{h}_\tau \mid \tilde{h}_\tau \in T, q_\tau, Q = 0]} = l_\tau \tilde{Z}(\tilde{h}_\tau \mid l_\tau). \end{aligned}$$

Also note that, given $Q^* = 0$, $\tilde{Z}(\cdot \mid l_\tau)$ has mean 1, which implies that $\{l_\tau\}_{\tau=1}^{\infty}$ forms a nonnegative martingale. Then, from the martingale convergence theorem, we can conclude that $l_\tau \rightarrow l_\infty$ almost surely. The rest of the proof is identical to that of Theorem 1 and Theorem 2.

Proof of Theorem 3

A summary statistic \mathbf{S}_τ is a vector $(i_1/\tau, \dots, i_m/\tau)$ such that $\sum_{j=1}^m i_j = \tau$ and $i_j \geq 0, j \in [m]$. For any summary statistics, which equivalently can be expressed as (i_1, \dots, i_m) at time τ , we let

$$p_1(i_1, \dots, i_m, \tau) \triangleq \mathbb{P}[\mathbf{S}_\tau = (i_1, \dots, i_m) \mid Q = 1], \quad p_0(i_1, \dots, i_m, \tau) \triangleq \mathbb{P}[\mathbf{S}_\tau = (i_1, \dots, i_m) \mid Q = 0].$$

We first show a recursive relation that determines the evolution of the values of $p_1(i_1, \dots, i_m, \tau)$ and $p_0(i_1, \dots, i_m, \tau)$. Note that crucially, the evolution of sequences $\{p_1(\cdot, \tau)\}$ and $\{p_0(\cdot, \tau)\}$ are not independent of each other as both of them depend on $q(\cdot, \tau)$ which in turn depends on both of the sequences. The next lemma shows that despite this dependence and the potential updating that Bayesian individuals may engage in view of the fact that there may have been customers who have taken actions that do not belong to T , these probabilities satisfy an intuitive recursion under our assumption of uniform priors over the number of past people who joined the platform.

Lemma 1. *The sequences $\{p_1(i_1, \dots, i_m, \tau)\}$ and $\{p_0(i_1, \dots, i_m, \tau)\}$ satisfy*

$$\begin{aligned}
& p_1(i_1, \dots, i_m, \tau) \\
&= \sum_{j=1}^m p_1(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m, \tau - 1) G_1(j, q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m, \tau - 1)), \\
& p_0(i_1, \dots, i_m, \tau) \\
&= \sum_{j=1}^m p_0(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m, \tau - 1) G_0(j, q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m, \tau - 1)), \\
& q(i_1, \dots, i_m, \tau) = \frac{p_1(i_1, \dots, i_m, \tau)}{p_1(i_1, \dots, i_m, \tau) + p_0(i_1, \dots, i_m, \tau)}. \tag{26}
\end{aligned}$$

Proof. For $\tau = 1$, the belief without any observation is $1/2$. Therefore, we have

$$p_1(\mathbf{e}_i, 1) = G_1\left(i, \frac{1}{2}\right), \quad p_0(\mathbf{e}_i, 1) = G_0\left(i, \frac{1}{2}\right), \quad i \in [m].$$

Similar to the proof of Proposition 2, we let random variable $H_{\tau-1:\tau}$ denote the history of actions in between $\tau - 1$ -th and τ -th actions in T . Given customers have uniform prior on the number of people who joined the platform denoted by U , and the fact that all customers in this interval observe the same history and hence form the same belief, we can write

$$\mathbb{P}[H_{\tau-1:\tau} | \mathbf{S}_{\tau-1}, Q] = \sum_{u=0}^{\infty} \mathbb{P}[H_{\tau-1:\tau} | U = u, \mathbf{S}_{\tau-1}, Q] = \sum_{u=0}^{\infty} \mathbb{P}[h \notin T | \mathbf{S}_{\tau-1}, Q]^u. \tag{27}$$

Using Eq. (27), we next show the update rule for $\tau \geq 2$. We have

$$\begin{aligned}
& p_1(i_1, \dots, i_m, \tau) = \mathbb{P}[\mathbf{S}_\tau(1) = i_1, \dots, \mathbf{S}_\tau(m) = i_m | Q = 1] \\
&= \sum_{j=1}^m \sum_{H_{\tau-1:\tau}} \mathbb{P}[\mathbf{S}_{\tau-1} = (i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m) | Q = 1] \mathbb{P}[H_{\tau-1:\tau} | Q = 1, \mathbf{S}_{\tau-1} = (i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m)] \\
&\times \mathbb{P}[h_\tau \in T_j | H_{\tau-1:\tau}, Q = 1, \mathbf{S}_{\tau-1} = (i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m)] \\
&= \sum_{j=1}^m \mathbb{P}[\mathbf{S}_{\tau-1} = (i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m) | Q = 1] \\
&\times \sum_{u=0}^{\infty} \mathbb{P}[h \notin T | q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m), Q = 1]^u \mathbb{P}[h \in T_j | q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m), Q = 1] \\
&= \sum_{j=1}^m \mathbb{P}[\mathbf{S}_{\tau-1} = (i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m) | Q = 1] \times \frac{\mathbb{P}[h \in T_j | q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m), Q = 1]}{1 - \mathbb{P}[h \notin T | q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m), Q = 1]} \\
&= \sum_{j=1}^m \mathbb{P}[\mathbf{S}_{\tau-1} = (i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m) | Q = 1] \times \mathbb{P}[h \in T_j | h \in T, q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m), Q = 1] \\
&= \sum_{j=1}^m p_1(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m, \tau - 1) \times G_1(j, q(i_1, \dots, i_{j-1}, i_j - 1, i_{j+1}, \dots, i_m, \tau - 1)).
\end{aligned}$$

We can write a similar recursion for $p_0(i_1, \dots, i_m, \tau)$. The proof concludes by using Bayes' rule to

find $q(i_1, \dots, i_m, \tau)$ as in Eq. (26). ■

We break the rest of the proof into three steps. Using Assumption 3, there exists $i \in [m]$ such that the range of $G_1(i, \cdot)$ and $G_0(i, \cdot)$ are separated. Without loss of generality, we suppose the range of $G_1(i, \cdot)$ is above the range of $G_0(i, \cdot)$, i.e., $G_1 = \min_q G_1(i, q) > \max_q G_0(i, q) = G_0$.

Step 1: The summary statistic i with $Q = 1$ majorizes a binomial random variable and the summary statistic i with $Q = 0$ is majorized by a binomial random variable. Formally, we have:¹⁹

$$\mathbf{S}_\tau(i) \succeq \text{Binomial}(G_1, \tau), \quad Q = 1, \quad \mathbf{S}_\tau(i) \preceq \text{Binomial}(G_0, \tau), \quad Q = 0.$$

Equivalently, we have

$$\begin{aligned} S_1(l, \tau) &\triangleq \mathbb{P}[\mathbf{S}_\tau(i) \leq l \mid Q = 1] = \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l p_1(k_1, \dots, k_m, \tau) \leq \sum_{j=0}^l \binom{\tau}{j} G_1^j (1 - G_1)^{\tau-j}, \\ S_0(l, \tau) &\triangleq \mathbb{P}[\mathbf{S}_\tau(i) \leq l \mid Q = 0] = \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l p_0(k_1, \dots, k_m, \tau) \geq \sum_{j=0}^l \binom{\tau}{j} G_0^j (1 - G_0)^{\tau-j}. \end{aligned} \quad (28)$$

Proof of Step 1: we prove the claim for $Q = 1$ as the proof for $Q = 0$ is identical. The proof idea is as follows. Given $Q = 1$, consider the random process $\mathbf{S}_\tau(i)$. Lemma 1 shows that the distribution of $\mathbf{S}_\tau(i)$ satisfies a recursive relation. We consider another random process, whose distribution satisfies a similar recursion with the difference that $G_1(i, q_\tau)$ is replaced by its minimum over all q , i.e., G_1 . We show that the distribution of this process at τ is the same as the distribution of $\text{Binomial}(G_1, \tau)$. Moreover, because this process is recursively defined using the minimum of $G_1(i, q)$ over q , it follows that \mathbf{S}_τ with $Q = 1$ majorizes $\text{Binomial}(G_1, \tau)$. We next show the formal proof. First, note that from definition in Eq. (28) we have

$$\begin{aligned} S_1(l, \tau) - S_1(l, \tau - 1) &= \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l p_1(k_1, \dots, k_m, \tau) - \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l p_1(k_1, \dots, k_m, \tau - 1) \\ &= \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l \sum_{j=1}^m p_1(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m, \tau - 1) G_1(j, q(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_m, \tau - 1)) \\ &\quad - \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l p_1(k_1, \dots, k_m, \tau - 1) \\ &= \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l p_1(k_1, \dots, k_m, \tau - 1) \sum_{j=1}^m G_1(j, q(k_1, \dots, k_m, \tau - 1)) \\ &\quad - \sum_{k_j, j \in [m] \setminus \{i\}} p_1(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau - 1) G_1(i, q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau - 1)) \\ &\quad - \sum_{k_j, j \in [m] \setminus \{i\}} \sum_{k_i=1}^l p_1(k_1, \dots, k_m, \tau - 1) \\ &= - \sum_{k_j, j \in [m] \setminus \{i\}} p_1(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau - 1) G_1(i, q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau - 1)), \end{aligned} \quad (29)$$

¹⁹We say random variable X majorizes random variable Y and write $X \succeq Y$ if X first-order stochastically dominates Y , i.e., for any a we have $\mathbb{P}[X \geq a] \geq \mathbb{P}[Y \geq a]$.

where the second equality follows from Lemma 1 and the last equality follows from $\sum_{i=1}^m G_1(i, q) = 1$. We define the sequence $\{\tilde{S}_1(l, \tau)\}_{l, \tau}$ recursively as

$$\tilde{S}_1(l, \tau) = \tilde{S}_1(l-1, \tau-1)G_1 + \tilde{S}_1(l, \tau-1)(1-G_1), \quad 0 \leq l \leq \tau, \tau \geq 0. \quad (30)$$

We will next show by induction on τ that $\tilde{S}_1(l, \tau) \geq S_1(l, \tau)$. For $\tau = 1$, we have

$$\tilde{S}_1(0, 1) = 1 - G_1 \geq 1 - G_1 \left(i, \frac{1}{2}\right) = S_1(0, 1), \quad \tilde{S}_1(1, 1) = 1 = S_1(1, 1).$$

We now suppose that the induction hypothesis holds for $\tau - 1$ and prove it for τ . We have

$$\begin{aligned} \tilde{S}_1(l, \tau) &= \tilde{S}_1(l-1, \tau-1)G_1 + \tilde{S}_1(l, \tau-1)(1-G_1) \geq S_1(l-1, \tau-1)G_1 + S_1(l, \tau-1)(1-G_1) \\ &\geq S_1(l, \tau-1) - G_1 \sum_{k_j, j \in [m] \setminus \{i\}} p_1(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau-1) \\ &\geq S_1(l, \tau-1) - \sum_{k_j, j \in [m] \setminus \{i\}} p_1(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau-1) G_1(i, q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau-1)) \\ &= S_1(l, \tau), \end{aligned} \quad (31)$$

where the first inequality follows from induction hypothesis, the second inequality follows from definition in Eq. (28), the third inequality follows from using $G_1 = \min_q G_1(i, q)$, and the last equality follows from using Eq. (29). We next show by induction on τ that the recursive definition of $\{\tilde{S}_1(l, \tau)\}$ given in Eq. (30) leads to

$$\tilde{S}_1(l, \tau) = \sum_{j=0}^l \binom{\tau}{j} G_1^j (1-G_1)^{\tau-j}.$$

This evidently holds for $\tau = 1$ as we have $\tilde{S}_1(0, 1) = 1 - G_1$ and $\tilde{S}_1(1, 1) = 1$. We now suppose it holds for $\tau - 1$ and prove it for τ . We have

$$\begin{aligned} \tilde{S}_1(l, \tau) &= \tilde{S}_1(l-1, \tau-1)G_1 + \tilde{S}_1(l, \tau-1)(1-G_1) \\ &= G_1 \sum_{j=0}^{l-1} \binom{\tau-1}{j} G_1^j (1-G_1)^{\tau-1-j} + (1-G_1) \sum_{j=0}^l \binom{\tau-1}{j} G_1^j (1-G_1)^{\tau-1-j} \\ &= \sum_{j=0}^l G_1^j (1-G_1)^{\tau-j} \left(\binom{\tau-1}{j} + \binom{\tau-1}{j-1} \right) = \sum_{j=0}^l G_1^j (1-G_1)^{\tau-j} \binom{\tau}{j}. \end{aligned} \quad (32)$$

Putting Eq. (31) and Eq. (32) together, completes the proof of Step 1.

Step 2: Letting $\Delta = G_1 - G_0$, for any $\epsilon > 0$, we have

$$\mathbb{P}[q_\tau > \epsilon \mid Q = 0] \leq \frac{1}{\epsilon} e^{-\tau \Delta^2/2}, \quad \forall \tau \geq 1, \text{ and } \mathbb{P}[1 - q_\tau > \epsilon \mid Q = 1] \leq \frac{1}{\epsilon} e^{-\tau \Delta^2/2}, \quad \forall \tau \geq 1. \quad (33)$$

Proof of Step 2: Using Step 1, for $\gamma = \frac{G_1 + G_0}{2}$, we have $G_1 - \frac{\Delta}{2} = G_0 + \frac{\Delta}{2} = \gamma$. This leads to

$$S_1(\gamma\tau, \tau) \leq \sum_{j=0}^{\gamma\tau} \binom{\tau}{j} G_1^j (1 - G_1)^{\tau-j} = \mathbb{P} \left[\text{Binomial}(G_1, \tau) \leq \left(G_1 - \frac{\Delta}{2} \right) \tau \right] \leq e^{-\tau\Delta^2/2},$$

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

$$S_0(\gamma\tau, \tau) \geq \sum_{j=0}^{\gamma\tau} \binom{\tau}{j} G_0^j (1 - G_0)^{\tau-j} = \mathbb{P} \left[\text{Binomial}(G_0, \tau) \leq \left(G_0 + \frac{\Delta}{2} \right) \tau \right] \geq 1 - e^{-\tau\Delta^2/2}.$$

We show the claim for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. We have

$$\begin{aligned} \mathbb{P}[q_\tau > \epsilon \mid Q = 0] &= \sum_{k_j, j \in [m]} p_0(k_1, \dots, k_m, \tau) \mathbf{1}\{q(k_1, \dots, k_m, \tau) > \epsilon\} \\ &= \sum_{l=0}^{\tau\gamma} \sum_{k_j, j \in [m] \setminus \{i\}} p_0(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau) \mathbf{1} \left\{ \frac{q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)}{1 - q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)} > \frac{\epsilon}{1 - \epsilon} \right\} \\ &+ \sum_{l=\tau\gamma}^{\tau} \sum_{k_j, j \in [m] \setminus \{i\}} p_0(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau) \mathbf{1} \left\{ \frac{q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)}{1 - q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)} > \frac{\epsilon}{1 - \epsilon} \right\} \\ &\leq \sum_{l=0}^{\tau\gamma} \sum_{k_j, j \in [m] \setminus \{i\}} p_0(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau) \mathbf{1} \left\{ \frac{q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)}{1 - q(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)} > \frac{\epsilon}{1 - \epsilon} \right\} \\ &+ (1 - S_0(\gamma\tau, \tau)) \quad \text{Using } \mathbf{1}\{\cdot\} \leq 1 \\ &\leq \sum_{l=0}^{\tau\gamma} \sum_{k_j, j \in [m] \setminus \{i\}} p_0(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau) \mathbf{1} \left\{ \frac{p_1(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)}{p_0(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau)} > \frac{\epsilon}{1 - \epsilon} \right\} \\ &+ e^{-\tau\Delta^2/2} \quad \text{Using Eq. (33)} \\ &\leq \sum_{l=0}^{\tau\gamma} \sum_{k_j, j \in [m] \setminus \{i\}} \frac{1 - \epsilon}{\epsilon} p_1(k_1, \dots, k_{i-1}, l, k_{i+1}, \dots, k_m, \tau) + e^{-\tau\Delta^2/2} \\ &\leq \frac{1 - \epsilon}{\epsilon} e^{-\tau\Delta^2/2} + e^{-\tau\Delta^2/2} = \frac{1}{\epsilon} e^{-\tau\Delta^2/2}. \quad \text{Using Eq. (33)} \end{aligned}$$

This completes the proof of Step 2.

Step 3: We have $q_\tau \rightarrow Q^*$ almost surely.

Proof of Step 3: We show the proof for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. The idea is to use 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

$$\mathcal{A}_\tau^{(\frac{1}{m})} \triangleq \left\{ q_\tau \geq \frac{1}{m} \right\}$$

Using Step 2, for any $m \in \mathbb{N}$ we have

$$\sum_{\tau=1}^{\infty} \mathbb{P} \left[\mathcal{A}_{\tau}^{(1/m)} \right] \leq m \sum_{\tau=1}^{\infty} e^{-\tau \Delta^2 / 2} < \infty.$$

Therefore, using Borell-Cantelli lemma leads to

$$\mathbb{P} \left[\text{i.o. } \mathcal{A}_{\tau}^{(1/m)} \right] = \mathbb{P} \left[\bigcap_{\tau \in \mathbb{N}} \bigcup_{k \geq \tau} \mathcal{A}_k^{(1/m)} \right] = 0.$$

Finally, we can write

$$\mathbb{P} \left[\left(\lim_{\tau \rightarrow \infty} q_{\tau} = 0 \right)^{\mathcal{C}} \right] = \mathbb{P} \left[\bigcup_{m \in \mathbb{N}} \bigcap_{\tau \in \mathbb{N}} \bigcup_{k \geq \tau} \mathcal{A}_k^{(1/m)} \right] \leq \sum_{m=1}^{\infty} \mathbb{P}[\text{i.o. } \mathcal{A}_{\tau}^{(1/m)}] = 0,$$

where we used union bound for the inequality. This completes the proof.

Proof of the Claim in Example 1

We show that asymptotic learning does not occur in the example. If there were asymptotic learning, then given $Q = 0$, $\frac{\mathbf{S}_{\tau}(1)}{\tau}$ would be concentrated around $G_0(1, q = 0)$ and given $Q = 1$, $\frac{\mathbf{S}_{\tau}(1)}{\tau}$ would be concentrated around $G_1(1, q = 1)$. However, as we will show next, $(\mathbf{S}_{\tau}(1) \mid Q = 1)$ majorizes $(\mathbf{S}_{\tau}(1) \mid Q = 0)$, establishing that $G_1(1, q = 1) \geq G_0(1, q = 0)$ which is contradiction. We will next establish this majorization relationship. Recall that

$$S_1(l, \tau) \triangleq \mathbb{P}[\mathbf{S}_{\tau}(1) \leq l \mid Q = 1], \quad S_0(l, \tau) \triangleq \mathbb{P}[\mathbf{S}_{\tau}(1) \leq l \mid Q = 0].$$

We will show by induction on τ that for all l , we have $S_1(l, \tau) \leq S_0(l, \tau)$. This holds for $\tau = 1$, as $S_1(0, 1) = 1 - G_1(1, q = \frac{1}{2}) \leq 1 - G_0(1, q = \frac{1}{2}) = S_0(0, 1)$ and $S_1(1, 1) = S_0(1, 1) = 1$. We next show that if this claim holds for $\tau - 1$, then it holds for τ as well. We can write

$$\begin{aligned} S_1(l, \tau) &= S_1(l, \tau - 1) - G_1(q(l, \tau - 1 - l, \tau - 1)) p_1(l, \tau - 1 - l, \tau - 1) \\ &= S_1(l, \tau - 1) (1 - G_1(q(l, \tau - 1 - l, \tau - 1))) + G_1(q(l, \tau - 1 - l, \tau - 1)) S_1(l - 1, \tau - 1) \\ &\leq S_0(l, \tau - 1) (1 - G_1(q(l, \tau - 1 - l, \tau - 1))) + G_1(q(l, \tau - 1 - l, \tau - 1)) S_0(l - 1, \tau - 1) \\ &= S_0(l, \tau - 1) - G_1(q(l, \tau - 1 - l, \tau - 1)) p_0(l, \tau - 1 - l, \tau - 1) \\ &\leq S_0(l, \tau - 1) - G_0(q(l, \tau - 1 - l, \tau - 1)) p_0(l, \tau - 1 - l, \tau - 1) = S_0(l, \tau), \end{aligned}$$

where the first inequality follows from induction hypothesis and the second inequality follows from $G_1(q) \geq G_0(q)$ for all $q \in [0, 1]$.

Proof of Theorem 4

We show the theorem for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. First note that using the assumptions we can invoke Theorem 3 to show $q_\tau \rightarrow Q^*$ almost surely.

Given $Q = 0$, since $q_\tau \rightarrow 0$ almost surely, we have that $\frac{S_\tau(i)}{\tau} \rightarrow G_0(i, 0)$ almost surely for all $i \in [m]$. Also, given $Q = 1$, since $q_\tau \rightarrow 1$ almost surely, we have that $\frac{S_\tau(i)}{\tau} \rightarrow G_1(i, 1)$ almost surely for all $i \in [m]$. Therefore, for any ϵ there exists large enough N such that for $\tau \geq N$ we have

$$\begin{aligned} q_\tau \leq \epsilon, \quad Q^* = 0 \quad & \left| \frac{S_\tau(i)}{\tau} - G_0(i, 0) \right| \leq \epsilon, \quad \forall i \in [m], Q^* = 0 \\ q_\tau \geq 1 - \epsilon, \quad Q^* = 1, \quad & \left| \frac{S_\tau(i)}{\tau} - G_1(i, 1) \right| \leq \epsilon, \quad \forall i \in [m], Q^* = 1. \end{aligned}$$

We let

$$\underline{G}_0(i, 0) = \min_{q \leq \epsilon} \mathbb{P}[h \in T_i \mid h \in T, q, Q = 0], \quad \bar{G}_1(i, 1) = \max_{q \geq 1 - \epsilon} \mathbb{P}[h \in T_i \mid h \in T, q, Q = 1].$$

With these definitions, suppose that at time $\tau \geq N$ we have $\mathbf{S}_\tau(i) = N_\tau(i)$ for $i \in [m]$. We let \mathcal{H}_τ be the set of possible histories that are consistent with $N_\tau(1), \dots, N_\tau(m)$. We obtain

$$\begin{aligned} \log \frac{q_\tau}{1 - q_\tau} &= \log \left(\frac{\sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_\tau \mid Q = 1]}{\sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_\tau \mid Q = 0]} \right) = \log \left(\frac{\sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_N \mid Q = 1] \mathbb{P}[H_{N+1:\tau} \mid H_N, Q = 1]}{\sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_N \mid Q = 0] \mathbb{P}[H_{N+1:\tau} \mid H_N, Q = 0]} \right) \\ &\leq \log \left(\frac{(\max_{H_N} \mathbb{P}[H_N \mid Q = 1]) \sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_{N+1:\tau} \mid H_N, Q = 1]}{(\min_{H_N} \mathbb{P}[H_N \mid Q = 0]) \sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_{N+1:\tau} \mid H_N, Q = 0]} \right) \\ &\leq \log \left(\frac{(\max_{H_N} \mathbb{P}[H_N \mid Q = 1]) |\mathcal{H}_\tau| \prod_{i=1}^m (\bar{G}_1(i, 1))^{\tilde{N}_\tau(i)}}{(\min_{H_N} \mathbb{P}[H_N \mid Q = 0]) |\mathcal{H}_\tau| \prod_{i=1}^m (\underline{G}_0(i, 0))^{\hat{N}_\tau(i)}} \right) \\ &= \log \left(\frac{(\max_{H_N} \mathbb{P}[H_N \mid Q = 1]) \prod_{i=1}^m (\bar{G}_1(i, 1))^{\tilde{N}_\tau(i)}}{(\min_{H_N} \mathbb{P}[H_N \mid Q = 0]) \prod_{i=1}^m (\underline{G}_0(i, 0))^{\hat{N}_\tau(i)}} \right), \end{aligned} \quad (34)$$

where $\tilde{N}_\tau(i)$ is $N_\tau(i)$ minus the number of i 's in $\arg \max_{H_N} \mathbb{P}[H_N \mid Q = 1]$ and $\hat{N}_\tau(i)$ is $N_\tau(i)$ minus the number of i 's in $\arg \min_{H_N} \mathbb{P}[H_N \mid Q = 0]$. Using Assumptions 1 and 2, we have $0 < \bar{G}_1(i, 1) \leq 1$ and $0 < \underline{G}_0(i, 0) \leq 1$, which leads to

$$\frac{(\bar{G}_1(i, 1))^{\tilde{N}_\tau(i)}}{(\underline{G}_0(i, 0))^{\hat{N}_\tau(i)}} \leq \left(\frac{\bar{G}_1(i, 1)}{\underline{G}_0(i, 0)} \right)^{N_\tau(i)} \left(\frac{1}{\bar{G}_1(i, 1) \underline{G}_0(i, 0)} \right)^N, \quad i \in [m]. \quad (35)$$

Putting Eq. (34) and Eq. (35) together, we obtain

$$\log \frac{q_\tau}{1 - q_\tau} \leq \log \left(\frac{\max_{H_N} \mathbb{P}[H_N \mid Q = 1]}{\min_{H_N} \mathbb{P}[H_N \mid Q = 0]} \prod_{i=1}^m \left(\frac{\bar{G}_1(i, 1)}{\underline{G}_0(i, 0)} \right)^{N_\tau(i)} \left(\frac{1}{\bar{G}_1(i, 1) \underline{G}_0(i, 0)} \right)^N \right). \quad (36)$$

Since $Q^* = 0$, we almost surely have $\frac{N_\tau(i)}{\tau} \rightarrow G_0(i, 0)$, which yields

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \frac{q_\tau}{1 - q_\tau} \leq \lim_{\tau \rightarrow \infty} \sum_{i=1}^m \frac{N_\tau(i)}{\tau} \log \left(\frac{\bar{G}_1(i, 1)}{\underline{G}_0(i, 0)} \right) = \sum_{i=1}^m G_0(i, 0) \log \left(\frac{\bar{G}_1(i, 1)}{\underline{G}_0(i, 0)} \right).$$

Since this inequality holds for any ϵ , by letting $\epsilon \rightarrow 0$ we have $\underline{G}_0(i, 0) \rightarrow G_0(i, 0)$ and $\bar{G}_1(i, 1) \rightarrow G_1(i, 1)$, which leads to

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \frac{q_\tau}{1 - q_\tau} \leq \sum_{i=1}^m G_0(i, 0) \log \left(\frac{G_1(i, 1)}{G_0(i, 0)} \right) = - \sum_{i=1}^m G_0(i, 0) \log \left(\frac{G_0(i, 0)}{G_1(i, 1)} \right). \quad (37)$$

Similarly, by letting

$$\bar{G}_0(i, 0) = \max_{q \leq \epsilon} \mathbb{P}[h \in T_i \mid h \in T, q, Q = 0], \quad \underline{G}_1(i, 1) = \min_{q \geq 1 - \epsilon} \mathbb{P}[h \in T_i \mid h \in T, q, Q = 1],$$

we obtain

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \frac{q_\tau}{1 - q_\tau} \geq \lim_{\tau \rightarrow \infty} \sum_{i=1}^m \frac{N_\tau(i)}{\tau} \log \left(\frac{\underline{G}_1(i, 1)}{\bar{G}_0(i, 0)} \right) = \sum_{i=1}^m G_0(i, 0) \log \left(\frac{\underline{G}_1(i, 1)}{\bar{G}_0(i, 0)} \right).$$

Again, by letting $\epsilon \rightarrow 0$ we have $\bar{G}_0(i, 0) \rightarrow G_0(i, 0)$ and $\underline{G}_1(i, 1) \rightarrow G_1(i, 1)$, which yields

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \frac{q_\tau}{1 - q_\tau} \geq - \sum_{i=1}^m G_0(i, 0) \log \left(\frac{G_0(i, 0)}{G_1(i, 1)} \right). \quad (38)$$

Therefore, combining Eq. (37) and Eq. (38), we obtain

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \frac{q_\tau}{1 - q_\tau} = -D(\mathbb{P}[\cdot \mid h \in T, q = 0, Q = 0] \parallel \mathbb{P}[\cdot \mid h \in T, q = 1, Q = 1]),$$

which completes the proof.

Proof of Corollary 1

We show the corollary for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. In this proof we let Like to denote $h = K_2$ and Review to denote $h \in \{K_2, -K_1\}$. We break the proof into two steps. We first show negative selection, i.e.,

$$\begin{aligned} \mathbb{P}[\text{Like} \mid \text{Review}, q = 1, Q = 0] &< \mathbb{P}[\text{Like} \mid \text{Review}, q = 0, Q = 0], \\ \mathbb{P}[\text{Like} \mid \text{Review}, q = 1, Q = 1] &< \mathbb{P}[\text{Like} \mid \text{Review}, q = 0, Q = 1]. \end{aligned}$$

We then compare the speed of the full history with summary statistics.

Step 1: We show $G_0(1, q)$ defined as

$$G_0(1, q) = \mathbb{P}[\text{Like} \mid \text{Review}, q, Q = 0],$$

is decreasing in q . An identical proof shows $G_1(1, q)$ is decreasing in q as well. We let λ_{-1} and λ_1 denote the thresholds for “like”, “no review”, and “dislike”. We have

$$\begin{aligned} G_0(1, q) &= \frac{\mathbb{P}[\theta + \zeta - p \geq \lambda_1, \theta + \mathbb{E}[\zeta] + q - p \geq 0]}{\mathbb{P}[\theta + \mathbb{E}[\zeta] + q - p \geq 0, \theta + \zeta - p \notin (\lambda_{-1}, \lambda_1)]} \\ &= \frac{\mathbb{P}[\theta + \zeta - p \geq \lambda_1, \theta + \mathbb{E}[\zeta] + q - p \geq 0]}{\mathbb{P}[\theta + \zeta - p \geq \lambda_1, \theta + \mathbb{E}[\zeta] + q - p \geq 0] + \mathbb{P}[\theta + \zeta - p \leq \lambda_{-1}, \theta + \mathbb{E}[\zeta] + q - p \geq 0]}. \end{aligned}$$

To show $G_0(1, q)$ is decreasing, it is sufficient to show

$$\frac{\mathbb{P}[\theta + \zeta - p \geq \lambda_1, \theta + \mathbb{E}[\zeta] + q - p \geq 0]}{\mathbb{P}[\theta + \zeta - p \leq \lambda_{-1}, \theta + \mathbb{E}[\zeta] + q - p \geq 0]}$$

is decreasing in q . This in turn is equal to

$$\frac{\int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (1 - F_{\zeta}(p + \lambda_1 - x)) dx}{\int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (F_{\zeta}(p + \lambda_{-1} - x)) dx},$$

where we used the independence of θ and ζ . Taking derivative of this term with respect to q yields

$$\begin{aligned} f_{\theta}(p - q - \mathbb{E}[\zeta]) \times &\left(\frac{(1 - F_{\zeta}(\lambda_1 + q + \mathbb{E}[\zeta])) \int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (F_{\zeta}(p + \lambda_{-1} - x)) dx}{\left(\int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (F_{\zeta}(p + \lambda_{-1} - x)) dx \right)^2} \right. \\ &\left. - \frac{(F_{\zeta}(\lambda_{-1} + q + \mathbb{E}[\zeta])) \int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (1 - F_{\zeta}(p + \lambda_1 - x)) dx}{\left(\int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (F_{\zeta}(p + \lambda_{-1} - x)) dx \right)^2} \right), \end{aligned}$$

which is non-positive because

$$\begin{aligned} &(1 - F_{\zeta}(\lambda_1 + q + \mathbb{E}[\zeta])) \int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (F_{\zeta}(p + \lambda_{-1} - x)) dx \\ &- (F_{\zeta}(\lambda_{-1} + q + \mathbb{E}[\zeta])) \int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (1 - F_{\zeta}(p + \lambda_1 - x)) dx \\ &\leq (1 - F_{\zeta}(\lambda_1 + q + \mathbb{E}[\zeta])) \int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (F_{\zeta}(\lambda_{-1} + q + \mathbb{E}[\zeta])) dx \\ &- (F_{\zeta}(\lambda_{-1} + q + \mathbb{E}[\zeta])) \int_{p-q-\mathbb{E}[\zeta]}^{\infty} f_{\theta}(x) (1 - F_{\zeta}(\lambda_1 + q + \mathbb{E}[\zeta])) dx = 0, \end{aligned}$$

where we used the fact that cumulative density function is non-decreasing for the inequality.

Step 2: We next compare the speed of learning with summary statistics and full history. We show

the proof for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. We let

$$\begin{aligned} G_{00} &= \mathbb{P}[\text{Like} \mid \text{Review}, q = 0, Q = 0], & G_{01} &= \mathbb{P}[\text{Like} \mid \text{Review}, q = 0, Q = 1], \\ G_{11} &= \mathbb{P}[\text{Like} \mid \text{Review}, q = 1, Q = 1]. \end{aligned}$$

Subtracting the speed of learning for full history (Theorem 2) from the speed of learning for summary statistics (Theorem 4), leads to

$$G_{00} \log \left(\frac{G_{11}}{G_{01}} \right) + G_{00} \log \left(\frac{G_{11}}{G_{01}} \right). \quad (39)$$

We next show given $G_{00} \leq G_{11} \leq G_{01}$, Eq. (39) is non-negative. Using $-\log(1+x) \geq -x$, we have

$$\begin{aligned} G_{00} \log \left(\frac{G_{11}}{G_{01}} \right) + (1 - G_{00}) \log \left(\frac{1 - G_{11}}{1 - G_{01}} \right) &\geq G_{00} \left(1 - \frac{G_{01}}{G_{11}} \right) + (1 - G_{00}) \left(1 - \frac{1 - G_{01}}{1 - G_{11}} \right) \\ &= 1 - \left(\frac{G_{00}G_{01}}{G_{11}} + \frac{(1 - G_{00})(1 - G_{01})}{1 - G_{11}} \right) = \frac{(G_{01} - G_{11})(G_{11} - G_{00})}{G_{11}(1 - G_{11})} \geq 0. \end{aligned}$$

Proof of Corollary 2

We will show the corollary for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. We let λ_{-1} and λ_1 denote the thresholds for “like”, “no review”, and “dislike”. First note that $\mathbb{P}[\text{Like} \mid q, Q]$ is increasing in both q and Q because we have

$$\mathbb{P}[\text{Like} \mid q, Q] = \mathbb{P}[\theta + q + \mathbb{E}[\zeta] - p \geq 0, \theta + \zeta - p + Q \geq \lambda_1].$$

We next show that the speed of learning with summary statistics is faster than the speed of learning with full history. We let

$$G_{00} = \mathbb{P}[\text{Like} \mid q = 0, Q = 0], \quad G_{01} = \mathbb{P}[\text{Like} \mid q = 0, Q = 1], \quad G_{11} = \mathbb{P}[\text{Like} \mid q = 1, Q = 1].$$

Subtracting the speed of learning for full history (Theorem 2) from the speed of learning for summary statistics (Theorem 4), leads to

$$G_{00} \log \left(\frac{G_{01}}{G_{11}} \right) + (1 - G_{00}) \log \left(\frac{1 - G_{01}}{1 - G_{11}} \right).$$

We next show that given $G_{00} \leq G_{01} \leq G_{11}$, this expression is non-negative. Using $-\log(1+x) \geq -x$, we obtain

$$\begin{aligned} G_{00} \log \left(\frac{G_{01}}{G_{11}} \right) + (1 - G_{00}) \log \left(\frac{1 - G_{01}}{1 - G_{11}} \right) &\geq G_{00} \left(1 - \frac{G_{11}}{G_{01}} \right) + (1 - G_{00}) \left(1 - \frac{1 - G_{11}}{1 - G_{01}} \right) \\ &= 1 - \left(\frac{G_{00}G_{11}}{G_{01}} + \frac{(1 - G_{00})(1 - G_{11})}{1 - G_{01}} \right) = \frac{(G_{11} - G_{01})(G_{01} - G_{00})}{G_{01}(1 - G_{01})} \geq 0. \end{aligned}$$

This completes the proof.

When Corollaries 1 and 2 Hold Simultaneously

Suppose $Q^* = 0$. For uniform θ and ζ with $\underline{\zeta} = -\bar{\zeta}$ (and hence $\mathbb{E}[\zeta] = 0$) and $\underline{\theta} = -\bar{\theta}$, we have

$$\begin{aligned}
\mathbb{P}[\text{Like} \mid q = 0, Q = 1] &= \mathbb{P}[\theta + \zeta + 1 - p \geq \lambda_1, \theta - p \geq 0] = \int_p^{\bar{\theta}} \frac{1}{\bar{\theta} - \underline{\theta}} \mathbb{P}[\zeta \geq \lambda_1 + p - 1 - x] dx \\
&= \frac{1}{(\bar{\zeta} - \underline{\zeta})(\bar{\theta} - \underline{\theta})} \int_p^{\bar{\theta}} (\bar{\zeta} - \lambda_1 - p + 1 + x) dx && \text{using } \bar{\zeta} \geq \max\{\lambda_1 + 1, \bar{\theta} + 1 - p - \lambda_1\} \\
&= \frac{1}{(\bar{\zeta} - \underline{\zeta})(\bar{\theta} - \underline{\theta})} \left((\bar{\theta} - p) \left(\bar{\zeta} - \lambda_1 - \frac{1}{2}p + 1 + \frac{1}{2}\bar{\theta} \right) \right) && \text{using } \bar{\theta} \geq \max\{p + \frac{1}{2}, 1 - p\}. \quad (40)
\end{aligned}$$

Similarly, we have

$$\mathbb{P}[\text{Like} \mid q = 1, Q = 0] = \frac{1}{(\bar{\zeta} - \underline{\zeta})(\bar{\theta} - \underline{\theta})} \left((\bar{\theta} - p + 1) \left(\bar{\zeta} - \lambda_1 - \frac{1}{2}p - \frac{1}{2} + \frac{1}{2}\bar{\theta} \right) \right). \quad (41)$$

Comparing Eq. (40) and Eq. (41), given $\bar{\theta} - p > \bar{\zeta} - \lambda_1 - \frac{1}{2}$, we obtain

$$\mathbb{P}[\text{Like} \mid q = 0, Q = 1] > \mathbb{P}[\text{Like} \mid q = 1, Q = 0],$$

which guarantees separation assumption. We next show the assumptions of Corollary 2 hold. Using $\bar{\theta} \geq \max\{p, 1 - p\}$ and $\bar{\zeta} \geq \max\{\lambda_1 + 1, \bar{\theta} + 1 - p - \lambda_1\}$, we can write

$$\begin{aligned}
\mathbb{P}[\text{Like} \mid \text{Review}, q = 1, Q = 1] &= \frac{\mathbb{P}[\theta + \zeta + 1 - p \geq \lambda_1, \theta - p + 1 \geq 0]}{\mathbb{P}[\theta - p + 1 \geq 0, \theta + \zeta + 1 - p \notin (\lambda_{-1}, \lambda_1)]} \\
&= \frac{1}{\bar{\zeta} - \underline{\zeta}} \left(\bar{\zeta} - \lambda_1 - \frac{p}{2} + \frac{1}{2} + \frac{1}{2}\bar{\theta} \right) \frac{1}{\left(1 - \frac{\lambda_1 - \lambda_{-1}}{\bar{\zeta} - \underline{\zeta}} \right)}. \quad (42)
\end{aligned}$$

Similarly, we have

$$\mathbb{P}[\text{Like} \mid \text{Review}, q = 0, Q = 0] = \frac{1}{\bar{\zeta} - \underline{\zeta}} \left(\bar{\zeta} - \lambda_1 - \frac{p}{2} + \frac{1}{2}\bar{\theta} \right) \frac{1}{\left(1 - \frac{\lambda_1 - \lambda_{-1}}{\bar{\zeta} - \underline{\zeta}} \right)}. \quad (43)$$

Comparing Eq. (42) and Eq. (43), we obtain

$$\mathbb{P}[\text{Like} \mid \text{Review}, q = 1, Q = 1] > \mathbb{P}[\text{Like} \mid \text{Review}, q = 0, Q = 0].$$

Therefore, the conditions of both Corollary 1 and Corollary 2 hold which concludes the proof.

Proof of Theorem 5

Full history: We show the proof for $Q^* = 0$ as the proof of $Q^* = 1$ is similar. Suppose the subset $T_i \subseteq T$ is partitioned into T_{i_l} and T_{i_h} . Therefore, with a more refined rating system the speed of learning changes by

$$\begin{aligned} & \mathbb{P}[h \in T_{i_l} | q = 0, Q = 0] \log \left(\frac{\mathbb{P}[h \in T_{i_l} | q = 0, Q = 0]}{\mathbb{P}[h \in T_{i_l} | q = 0, Q = 1]} \right) \\ & + \mathbb{P}[h \in T_{i_h} | q = 0, Q = 0] \log \left(\frac{\mathbb{P}[h \in T_{i_h} | q = 0, Q = 0]}{\mathbb{P}[h \in T_{i_h} | q = 0, Q = 1]} \right) \\ & - \mathbb{P}[h \in T_i | q = 0, Q = 0] \log \left(\frac{\mathbb{P}[h \in T_i | q = 0, Q = 0]}{\mathbb{P}[h \in T_i | q = 0, Q = 1]} \right). \end{aligned} \quad (44)$$

Since $\mathbb{P}[h \in T_i | q = 0, Q] = \mathbb{P}[h \in T_{i_l} | q = 0, Q] + \mathbb{P}[h \in T_{i_h} | q = 0, Q]$, we can use the following inequality

$$x \log \frac{x}{x'} + y \log \frac{y}{y'} \geq (x + y) \log \frac{x + y}{x' + y'}, \quad \forall x, x', y, y' \in (0, 1) \quad (45)$$

to conclude that Eq. (44) is non-negative. Finally, note that Eq. (45) holds because of Jensen's inequality. This completes the proof.

Summary statistics: The proof follows from Theorem 4 and a similar argument to full history.

Proof of Theorem 6

We break the proof into two parts. In the first part we show that learning occurs and in the second part we find the speed of learning.

Part 1 (Learning): The proof is similar to that of Theorem 3. We break the proof into three steps. First, we show a lemma to pin down the evolution of beliefs and probability of the summary statistic given $Q = 0$ and $Q = 1$. Second, we show that the probabilities are separated. Third, we show the exponential tail bound for the beliefs which in turn shows almost sure convergence.

Step 1: Letting

$$p_Q(i, \tau) \triangleq \mathbb{P}[S_\tau = i | Q], \quad Q \in \{0, 1\},$$

the following lemma establishes a recursive relation which pinpoints the values of $p_1(i, \tau)$ and $p_0(i, \tau)$.

Lemma 2. *The sequences $\{p_1(i, \tau)\}$ and $\{p_0(i, \tau)\}$ satisfy*

$$p_Q(i, \tau) = \sum_{j=1}^K p_Q(i - j, \tau - 1) G_Q(j, q(i - j, \tau - 1)), \quad (46)$$

where $q(i, \tau) = \frac{p_1(i, \tau)}{p_1(i, \tau) + p_0(i, \tau)}$.

Proof. The proof is similar to that of Lemma 1. ■

Step 2: Given the separation assumption, we can find numbers $G_1(j)$ and $G_0(j)$ for $j \in [K]$ such that

$$\sum_{j=i}^K G_1(j) = \min_q \sum_{j=i}^K G_1(j, q) > \max_q \sum_{j=i}^K G_0(j, q) = \sum_{j=i}^K G_0(j), \quad 2 \leq i \leq K.$$

We show that the probability of average summary statistics with $Q = 1$ majorizes a multinomial random variable with alphabet $1, \dots, K$, where the probability of letter $i \in [K]$ is $G_1(i)$. Similarly the probability of summary statistics i with $Q = 0$ is majorized by a multinomial random variable. Formally, we have

$$\begin{aligned} S_\tau &\succeq \text{Multinomial}(1, \dots, K; G_1(1), \dots, G_1(K); \tau), & Q = 1, \\ S_\tau &\preceq \text{Multinomial}(1, \dots, K; G_0(1), \dots, G_0(K); \tau), & Q = 0. \end{aligned}$$

Equivalently, letting

$$S_1(l, \tau) \triangleq \mathbb{P}[S_\tau \leq l \mid Q = 1] = \sum_{j=1}^l p_1(j, \tau), \quad S_0(l, \tau) \triangleq \mathbb{P}[S_\tau \leq l \mid Q = 0] = \sum_{j=1}^l p_0(j, \tau),$$

we have

$$\begin{aligned} S_1(l, \tau) &\leq \tilde{S}_1(l, \tau) \triangleq \mathbb{P}[\text{Multinomial}(1, \dots, K; G_1(1), \dots, G_1(K); \tau) \leq l] \\ S_0(l, \tau) &\geq \tilde{S}_0(l, \tau) \triangleq \mathbb{P}[\text{Multinomial}(1, \dots, K; G_0(1), \dots, G_0(K); \tau) \leq l] \end{aligned}$$

Proof of Step 2: We prove the claim for $Q = 1$ as the proof for $Q = 0$ is identical. The proof follows by induction on τ . If it holds for $\tau - 1$, then using Lemma 2 we can write

$$\begin{aligned} S_1(l, \tau) &= \sum_{i=1}^l \sum_{j=1}^K p_Q(i-j, \tau-1) G_1(j, q(i-j, \tau-1)) \\ &\leq S_1(l-1, \tau-1) - \sum_{j=1}^{K-1} p_1(l-j, \tau-1) \sum_{r=j+1}^K G_1(r), \\ &= \sum_{j=1}^K S_1(l-j, \tau-1) G_1(j) \leq \sum_{j=1}^K \tilde{S}_1(l-j, \tau-1) G_1(j) = \tilde{S}_1(l, \tau). \end{aligned} \tag{47}$$

The proof of Step 1 finishes by noting that for $\tau = 1$ we have

$$\begin{aligned} S_1(1, 1) &= G_1\left(1, \frac{1}{2}\right) \leq G_1(1) = \tilde{S}_1(1, 1), \quad S_1(K, 1) = 1 = \tilde{S}_1(K, 1), \\ S_1(i, 1) &= \sum_{j=1}^i G_1\left(j, \frac{1}{2}\right) \leq \sum_{j=1}^i G_1(j) = \tilde{S}_1(i, 1), \quad 2 \leq i \leq K-1. \end{aligned}$$

Step 3: Letting $\Delta = \left(\sum_{j=1}^K jG_1(j)\right) - \left(\sum_{j=1}^K jG_0(j)\right)$, for any $\epsilon > 0$, we have

$$\mathbb{P}[q_\tau > \epsilon \mid Q = 0] \leq \frac{2}{\epsilon} e^{-\tau \frac{\Delta^2}{2K^2}}, \quad \forall \tau \geq 1, \text{ and } \mathbb{P}[1 - q_\tau > \epsilon \mid Q = 1] \leq \frac{2}{\epsilon} e^{-\tau \frac{\Delta^2}{2K^2}}, \quad \forall \tau \geq 1. \quad (48)$$

Moreover, we have $q_\tau \rightarrow Q^*$ almost surely.

Proof of Step 3: First note that $\Delta > 0$, as we have

$$\sum_{j=1}^K jG_1(j) = \sum_{j=1}^K \sum_{i=j}^K G_1(i) > \sum_{j=1}^K \sum_{i=j}^K G_0(i) = \sum_{j=1}^K jG_0(j).$$

Letting $\gamma = \frac{1}{2} \left(\sum_{j=1}^K jG_1(j) + \sum_{j=1}^K jG_0(j)\right)$ and using Step 2 along with Chernoff-Hoeffding's bound, we have

$$S_1(\gamma\tau, \tau) \leq \tilde{S}_1(\gamma\tau, \tau) = \mathbb{P} \left[\text{Multinomial}(1, \dots, K; G_1(1), \dots, G_1(K); \tau) \leq \tau \frac{\Delta}{2} \right] \leq 2e^{-\tau \frac{\Delta^2}{2K^2}}.$$

We also have

$$S_0(\gamma\tau, \tau) \geq 1 - 2e^{-\tau \frac{\Delta^2}{2K^2}}. \quad (49)$$

We show the claim for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. We have

$$\begin{aligned} \mathbb{P}[q_\tau > \epsilon \mid Q = 0] &= \sum_{i=0}^{K\tau} p_0(i, \tau) \mathbf{1}\{q(i, \tau) > \epsilon\} = \sum_{i=0}^{K\tau} p_0(i, \tau) \mathbf{1}\left\{\frac{q(i, \tau)}{1 - q(i, \tau)} > \frac{\epsilon}{1 - \epsilon}\right\} \\ &= \sum_{i=0}^{\tau\gamma} p_0(i, \tau) \mathbf{1}\left\{\frac{q(i, \tau)}{1 - q(i, \tau)} > \frac{\epsilon}{1 - \epsilon}\right\} + \sum_{i=\tau\gamma}^{K\tau} p_0(i, \tau) \mathbf{1}\left\{\frac{q(i, \tau)}{1 - q(i, \tau)} > \frac{\epsilon}{1 - \epsilon}\right\} \\ &\leq \sum_{i=0}^{\tau\gamma} p_0(i, \tau) \mathbf{1}\left\{\frac{q(i, \tau)}{1 - q(i, \tau)} > \frac{\epsilon}{1 - \epsilon}\right\} + (1 - S_0(\gamma\tau, \tau)) && \text{Using } \mathbf{1}\{\cdot\} \leq 1 \\ &\leq \sum_{i=0}^{\tau\gamma} p_0(i, \tau) \mathbf{1}\left\{\frac{q(i, \tau)}{1 - q(i, \tau)} > \frac{\epsilon}{1 - \epsilon}\right\} + 2e^{-\tau \frac{\Delta^2}{2K^2}} && \text{Using Eq. (48)} \\ &\leq \sum_{i=1}^{\tau\gamma} \frac{1 - \epsilon}{\epsilon} p_1(i, \tau) + 2e^{-\tau \frac{\Delta^2}{2K^2}} \leq \frac{1 - \epsilon}{\epsilon} 2e^{-\tau \frac{\Delta^2}{2K^2}} + 2e^{-\tau \frac{\Delta^2}{2K^2}} = \frac{2}{\epsilon} e^{-\tau \frac{\Delta^2}{2K^2}}. && \text{Using Eq. (48)} \end{aligned}$$

This completes the proof of the first part. The second part follows from using Borell-Cantelli Lemma together with the exponential tail bound, similar to that of Theorem 3.

Part 2 (Speed of Learning): We show this part for $Q^* = 0$ as the proof for $Q^* = 1$ is identical. We will use Sanov's Theorem (Cover and Thomas [2012, Chapter 11]) stated below:

Let Q be a probability distribution on finite set $\mathcal{X} = \{1, \dots, K\}$. Also, let E be a (closed) set of

probability distribution on \mathcal{X} , then we have

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \left(\sum_{\substack{n_1, \dots, n_K \\ \sum_{i=1}^K n_i = \tau \\ (\frac{n_1}{\tau}, \dots, \frac{n_K}{\tau}) \in E}} \binom{\tau}{n_1, \dots, n_K} \left(\prod_{i=1}^K Q^{(i)n_i} \right) \right) = - \min_{P \in E} D(P || Q),$$

Similar to the proof of Theorem 4, letting \mathcal{H}_τ indicate the set of all histories for which the average of ratings is $E_0 \triangleq \sum_{j=1}^K jG_0(j, 0)$, we obtain

$$\begin{aligned} & \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \left(\frac{q_\tau}{1 - q_\tau} \right) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \left(\frac{\sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_\tau | Q = 1]}{\sum_{H_\tau \in \mathcal{H}_\tau} \mathbb{P}[H_\tau | Q = 0]} \right) \\ & = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \left(\frac{\sum_{\substack{n_1, \dots, n_K \\ \sum_{j=1}^K n_j = \tau}} \binom{\tau}{n_1, \dots, n_K} \left(\prod_{j=1}^K G_1(j, 1)^{n_j} \right)}{\sum_{\substack{n_1, \dots, n_K \\ \sum_{j=1}^K n_j = \tau \\ \sum_{j=1}^K j n_j = \tau E_0}} \binom{\tau}{n_1, \dots, n_K} \left(\prod_{j=1}^K G_0(j, 0)^{n_j} \right)} \right) \\ & = - \min_{\substack{a_1, \dots, a_K \\ \sum_{j=1}^K a_j = 1 \\ \sum_{j=1}^K j a_j = E_0}} D([a_1, \dots, a_K] || [G_1(1, 1), \dots, G_1(K, 1)]), \end{aligned}$$

where we used Sanov's Theorem with $E = \{(a_1, \dots, a_K) : \sum_{j=1}^K a_j = 1, \sum_{j=1}^K j a_j = E_0\}$ to obtain the limit. Finally, note that this speed of learning is smaller (not larger) than

$$D([G_0(1, 0), \dots, G_0(K, 0)] || [G_1(1, 1), \dots, G_1(K, 1)]),$$

which is the speed of learning for vector summary statistics.

Proof of Theorem 7

To study the effects of targeted rating systems we consider m types of customers where the valuations of customers of type $i \in [m]$ are drawn iid from distributions $f_\theta^{(i)}$ and $f_\zeta^{(i)}$ for θ and ζ , respectively. We also let p_i denote the probability that a new customer belongs to group i for $1 \leq i \leq m$. Similar to Eq. (4), for any $i \in [m]$, we let $\mathbb{P}^{(i)}[\cdot | q_t, Q]$ denote the probability of different actions when the distribution of θ and ζ are $f_\theta^{(i)}$ and $f_\zeta^{(i)}$. Using these notations, we next show the proof of this Theorem for $Q^* = 0$ (the proof for $Q^* = 1$ follows from a similar argument).

Full history: Learning the true quality follows from Theorem 1. We next characterize the speed of learning. We have $m \times (K_2 + K_1 + 1)$ total possible actions, i.e., $\{1, \dots, m\} \times (\{-K_1, \dots, K_2\} \cup \{N\})$.

Therefore, Theorem 2 shows that learning speed is

$$\begin{aligned} & \sum_{i=1}^m \sum_{h \in \{-K_1, \dots, K_2\} \cup \{N\}} p_i \mathbb{P}^{(i)} [h \mid q = 0, Q = 0] \log \left(\frac{p_i \mathbb{P}^{(i)} [h \mid q = 0, Q = 0]}{p_i P^{(i)} (h \mid q = 0, Q = 1)} \right) \\ &= \sum_{i=1}^m p_i D \left(\mathbb{P}^{(i)} [\cdot \mid q = 0, Q = 0] \parallel \mathbb{P}^{(i)} [\cdot \mid q = 0, Q = 1] \right). \end{aligned}$$

We next show that learning speed is higher than learning speed without types. Using convexity of KL-divergence and Jensen's inequality, we obtain

$$\begin{aligned} D(\mathbb{P}[\cdot \mid q = 0, Q = 0] \parallel \mathbb{P}[\cdot \mid q = 0, Q = 1]) &= D \left(\sum_{i=1}^m p_i \mathbb{P}^{(i)} [\cdot \mid q = 0, Q = 0] \parallel \sum_{i=1}^m p_i \mathbb{P}^{(i)} [\cdot \mid q = 0, Q = 1] \right) \\ &\leq \sum_{i=1}^m p_i D \left(\mathbb{P}^{(i)} [\cdot \mid q = 0, Q = 0] \parallel \mathbb{P}^{(i)} [\cdot \mid q = 0, Q = 1] \right), \end{aligned}$$

which completes the proof of this part.

Summary statistics: Learning the true quality follows from Theorem 3. Using Theorem 4, for each type $i \in [m]$ the speed of learning is given by

$$\lim_{t_i \rightarrow \infty} \frac{1}{t_i} \log l_{t_i} = -D(\mathbb{P}[\cdot \mid h \in T, q = 0, Q = 0] \parallel \mathbb{P}[\cdot \mid h \in T, q = 1, Q = 1]), \quad (50)$$

where t_i denotes the sequence of customers that belong to type i . Each customer that joins the platforms observes the summary statistics of all of m rating systems and forms a belief, which using Bayes' rule can be written as $l_t = \prod_{i=1}^m l_{t_i}$. Using Eq. (50) along with $\frac{t_i}{t} \rightarrow p_i$ almost surely (using SLLN), we obtain

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log l_t = - \sum_{i=1}^m p_i D(\mathbb{P}[\cdot \mid h \in T, q = 0, Q = 0] \parallel \mathbb{P}[\cdot \mid h \in T, q = 1, Q = 1]).$$

Again, using Jensen's inequality completes the proof.

Proof of Theorem 8

We break the proof into two steps. In the first step we reformulate the platform's revenue and highlight that the speed of learning is a key component of platform's revenue. In the second step we show that increasing speed of learning, increases platform's revenue.

Step 1: Maximizing the expected revenue of platform is equivalent to maximizing

$$\begin{aligned}
& \mathbb{E}_{\theta, \zeta} \left[\left(\frac{1}{2} - p + \theta + \zeta \right) \mathbf{1} \{ \theta + \zeta - p \geq 0 \} + \frac{1}{2} (1 - p + \theta + \zeta) \mathbf{1} \{ \theta + \zeta - p \in (-1, 0) \} \right] \\
& + \frac{1}{2} \mathbb{E}_{\theta, \zeta} [(- (p - \theta - \zeta)) \mathbb{P}_{t, \Omega_t} [q_t \geq p - \theta - \zeta \mid Q = 0] \mathbf{1} \{ \theta + \zeta - p \in (-1, 0) \}] \\
& + \frac{1}{2} \mathbb{E}_{\theta, \zeta} [(- (1 - (p - \theta - \zeta))) \mathbb{P}_{t, \Omega_t} [q_t \leq p - \theta - \zeta \mid Q = 1] \mathbf{1} \{ \theta + \zeta - p \in (-1, 0) \}]. \quad (51)
\end{aligned}$$

The platform's expected revenue from customer t is proportional to $\mathbb{P}[j_t = 1]$ (recall that platform gains a fixed revenue from each customer that joins the platform). Therefore, the platform's problem is to maximize $\mathbb{P}[j_t = 1]$. Using Eq. (2), this probability can be written as

$$\begin{aligned}
\mathbb{P}_c [j_t = 1] &= \mathbb{P} [\mathbb{E}_{Q, \theta_t, \zeta_t, \Omega_t, t} [\mathbf{1} \{ b_t = 1 \} (\theta_t + \zeta_t + Q - p)] - c_t \geq 0] \\
&= F_c (\mathbb{E}_{Q, \theta_t, \zeta_t, \Omega_t, t} [(\theta_t + \zeta_t + Q - p) \mathbf{1} \{ b_t = 1 \}]).
\end{aligned}$$

Since $F_c(\cdot)$ is non-decreasing, the platform's problem is to maximize

$$\mathbb{E}_{Q, \theta_t, \zeta_t, \Omega_t, t} [(\theta_t + \zeta_t + Q - p) \mathbf{1} \{ b_t = 1 \}].$$

Using Eq. (3), we have

$$\begin{aligned}
& \mathbb{E}_{Q, \theta_t, \zeta_t, \Omega_t, t} [(\theta_t + \zeta_t + Q - p) \mathbf{1} \{ b_t = 1 \}] = \mathbb{E}_{Q, \theta_t, \zeta_t, \Omega_t, t} [(\theta_t + \zeta_t + Q - p) \mathbf{1} \{ q_t > p - \theta_t - \mathbb{E}[\zeta] \}] \\
& = \frac{1}{2} \mathbb{E}_{\theta, \zeta} [- (p - \theta - \zeta) \mathbb{P}_{t, \Omega_t} [q_t > p - \mathbb{E}[\zeta] - \theta \mid Q = 0]] \\
& + \frac{1}{2} \mathbb{E}_{\theta, \zeta} [(1 - (p - \theta - \zeta)) \mathbb{P}_{t, \Omega_t} [q_t > p - \zeta - \mathbb{E}[\theta] \mid Q = 1]] \\
& = \frac{1}{2} (1 - p + \mathbb{E}_{\theta, \zeta} [\theta + \zeta]) + \frac{1}{2} \mathbb{E}_{\theta, \zeta} [- (p - \theta - \zeta) \mathbb{P}_{t, \Omega_t} [q_t \geq p - \theta - \zeta \mid Q = 0]] \\
& + \frac{1}{2} \mathbb{E}_{\theta, \zeta} [- (1 - (p - \theta - \zeta)) \mathbb{P}_{t, \Omega_t} [q_t \leq p - \theta - \zeta \mid Q = 1]]. \quad (52)
\end{aligned}$$

We can further write

$$\begin{aligned}
& \mathbb{E}_{Q, \theta_t, \zeta_t, \Omega_t, t} [(\theta_t + \zeta_t + Q - p) \mathbf{1} \{ b_t = 1 \}] \\
& = \frac{1}{2} (1 - p + \mathbb{E}_{\theta, \zeta} [\theta + \zeta]) + \frac{1}{2} \mathbb{P}_{\theta, \zeta} [p - \theta - \zeta \leq 0] \mathbb{E}_{\theta, \zeta} [- (p - \theta - \zeta) \mid p - \theta - \zeta \leq 0] \\
& + \frac{1}{2} \mathbb{P}_{\theta, \zeta} [p - \theta - \zeta \geq 1] \mathbb{E}_{\theta, \zeta} [- (1 - (p - \theta - \zeta)) \mid p - \theta - \zeta \geq 1] \\
& + \frac{1}{2} \mathbb{P}_{\theta, \zeta} [p - \theta - \zeta \in [0, 1]] \mathbb{E}_{\theta, \zeta} [- (p - \theta - \zeta) \mathbb{P}_{t, \Omega_t} [q_t \geq p - \theta - \zeta \mid Q = 0] \mid p - \theta - \zeta \in [0, 1]] \\
& + \frac{1}{2} \mathbb{P}_{\theta, \zeta} [p - \theta - \zeta \in [0, 1]] \mathbb{E}_{\theta, \zeta} [- (1 - (p - \theta - \zeta)) \mathbb{P}_{t, \Omega_t} [q_t \leq p - \theta - \zeta \mid Q = 1] \mid p - \theta - \zeta \in [0, 1]] \\
& = \mathbb{E}_{\theta, \zeta} \left[\left(\frac{1}{2} - p + \theta + \zeta \right) \mathbf{1} \{ \theta + \zeta - p \geq 0 \} + \frac{1}{2} (1 - p + \theta + \zeta) \mathbf{1} \{ \theta + \zeta - p \in (-1, 0) \} \right] \\
& + \frac{1}{2} \mathbb{E}_{\theta, \zeta} [(- (p - \theta - \zeta)) \mathbb{P}_{t, \Omega_t} [q_t \geq p - \theta - \zeta \mid Q = 0] \mathbf{1} \{ \theta + \zeta - p \in (-1, 0) \}] \\
& + \frac{1}{2} \mathbb{E}_{\theta, \zeta} [(- (1 - (p - \theta - \zeta))) \mathbb{P}_{t, \Omega_t} [q_t \leq p - \theta - \zeta \mid Q = 1] \mathbf{1} \{ \theta + \zeta - p \in (-1, 0) \}].
\end{aligned}$$

This proves the first step.

Step 2: If system I has a higher speed than system II, then the platform's revenue from using system I is higher.

For the second step, note that the terms that depend on the rating system are

$$\begin{aligned} & \frac{1}{2} \mathbb{P}_{\theta, \zeta} [p - \theta - \zeta \in [0, 1]] \mathbb{E}_{\theta, \zeta} [-(p - \theta - \zeta) \mathbb{P}_{t, \Omega_t} [q_t \geq p - \theta - \zeta \mid Q = 0] \mid p - \theta - \zeta \in [0, 1]] \\ & + \frac{1}{2} \mathbb{P}_{\theta, \zeta} [p - \theta - \zeta \in [0, 1]] \mathbb{E}_{\theta, \zeta} [-(1 - (p - \theta - \zeta)) \mathbb{P}_{t, \Omega_t} [q_t \leq p - \theta - \zeta \mid Q = 1] \mid p - \theta - \zeta \in [0, 1]]. \end{aligned}$$

Also, note that if Ω_t^I has a higher learning speed than Ω_t^{II} , then since customers have an improper uniform prior on the number of previous customers (before joining the platform), we obtain

$$\mathbb{P}_{t, \Omega_t^I, \Omega_t^{II}} [q_t^I \leq q_t^{II} \mid Q = 0] = 1, \quad \mathbb{P}_{t, \Omega_t^I, \Omega_t^{II}} [q_t^I \geq q_t^{II} \mid Q = 1] = 1.$$

This leads to

$$\begin{aligned} \mathbb{P}_{t, \Omega_t^I} [q_t^I \geq p - \theta - \zeta \mid Q = 0] & \leq \mathbb{P}_{t, \Omega_t^{II}} [q_t^{II} \geq p - \theta - \zeta \mid Q = 0], \\ \mathbb{P}_{t, \Omega_t^I} [q_t^I \leq p - \theta - \zeta \mid Q = 1] & \leq \mathbb{P}_{t, \Omega_t^{II}} [q_t^{II} \leq p - \theta - \zeta \mid Q = 1]. \end{aligned}$$

Since $p - \theta - \zeta \in [0, 1]$, we have $-(p - \theta - \zeta) \leq 0$ and $-(1 - (p - \theta - \zeta)) \leq 0$, which shows the revenue of the rating system Ω_t^I is higher than the revenue of the rating system with Ω_t^{II} .

References

- D. Acemoglu, M. Dahleh, I. Lobel, and A. Ozdaglar (2009), Rate of convergence of learning in social networks. In *Proceedings of the American Control Conference*, pages 2825–2830.
- D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar (2011), Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236.
- M. Amador and P.-O. Weill (2012), Learning from private and public observations of others? actions. *Journal of Economic Theory*, 147(3):910–940.
- A. Banerjee and D. Fudenberg (2004), Word-of-mouth learning. *Games and Economic Behavior*, 46(1):1–22.
- A. V. Banerjee (1992), A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817.
- O. Besbes and M. Scarsini (2015). On information distortions in online ratings. *Columbia Business School Research Paper*.
- S. Bikhchandani, D. Hirshleifer, and I. Welch (1992), A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026.

- T. Blake, C. Nosko, and S. Tadelis (2016), Returns to consumer search: Evidence from ebay. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 531–545.
- Y.-K. Che and J. Horner (2015). Optimal design for social learning.
- A. Y. Chua and S. 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.
- T. M. Cover and J. A. Thomas (2012). *Elements of information theory*. John Wiley & Sons.
- D. Crapis, B. Ifrach, C. Maglaras, and M. Scarsini (2016), Monopoly pricing in the presence of social learning. *Management Science*.
- K. Drakopoulos, A. Ozdaglar, and J. N. Tsitsiklis (2013), On learning with finite memory. *IEEE Transactions on Information Theory*, 59(10):6859–6872.
- R. Durrett (2010), *Probability: theory and examples*. Cambridge university press.
- N. Etemadi (1981), An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 55(1):119–122.
- N. Garg and R. Johari (2017), Pairwise comparisons for online reputation systems.
- W. Hann-Caruthers, V. V. Martynov, and O. Tamuz (2017), The speed of sequential asymptotic learning.
- M. Harel, E. Mossel, P. Strack, and O. Tamuz (2014), When more information reduces the speed of learning. Available at SSRN 2541707.
- J. Horner and N. S. Lambert (2016), Motivational ratings.
- B. Ifrach, C. Maglaras, and M. Scarsini (2014), Bayesian social learning with consumer reviews. *ACM SIGMETRICS Performance Evaluation Review*, 41(4):28–28.
- V. Korchevsky and V. Petrov (2010), On the strong law of large numbers for sequences of dependent random variables. *Vestnik St. Petersburg University: Mathematics*, 43(3):143–147.
- J. Lafky (2014), Why do people rate? theory and evidence on online ratings. *Games and Economic Behavior*, 87:554–570.
- X. Li and L. M. Hitt (2008), Self-selection and information role of online product reviews. *Information Systems Research*, 19(4):456–474.
- I. Lobel and E. Sadler (2015a), Information diffusion in networks through social learning. *Theoretical Economics*, 10(3):807–851.
- I. Lobel and E. Sadler (2015b), Preferences, homophily, and social learning. *Operations Research*, 64(3):564–584.

- D. Mayzlin, Y. Dover, and J. Chevalier (2014), Promotional reviews: An empirical investigation of online review manipulation. *The American Economic Review*, 104(8):2421–2455.
- E. Mossel, A. Sly, and O. Tamuz (2014), Asymptotic learning on bayesian social networks. *Probability Theory and Related Fields*, 158(1-2):127–157.
- E. Mossel, A. Sly, and O. Tamuz (2015), Strategic learning and the topology of social networks. *Econometrica*, 83(5):1755–1794.
- L. Smith and P. Sørensen (2000), Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398.
- A. Talwar, R. Jurca, and B. Faltings (2007), Understanding user behavior in online feedback reporting. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 134–142.
- W. P. Tay, J. N. Tsitsiklis, and M. Z. Win (2008), On the subexponential decay of detection error probabilities in long tandems. *IEEE Transactions on Information Theory*, 54(10):4767–4771.
- S. Vaccari, M. Scarsini, and C. Maglaras (2016), Social learning in a competitive market with consumer reviews.
- X. Vives (1993), How fast do rational agents learn? *The Review of Economic Studies*, 60(2):329–347.
- X. Vives (1995), The speed of information revelation in a financial market mechanism. *Journal of Economic Theory*, 1(67):178–204.
- I. Welch (1992), Sequential sales, learning, and cascades. *The Journal of finance*, 47(2):695–732.