Reviews or Ratings: Quantifying Information Loss from Coarsening*

Matthew J. H. Murphy†

This version: Apr 9, 2025

This paper studies optimal design of a review system that elicits information from reviewers who willingly submit reviews. I show how a platform seeking to learn an unknown state should trade off between the informativeness and frequency of submission of reviews. Finer reviews provide more information per review, but their complexity leads reviewers to submit them less frequently. Signals capture the realized utility of past consumers, which depends on the state of the world and an idiosyncratic individual component. I study how taste heterogeneity for the product impacts the performance of these systems. When consumers are more homogeneous, simple reviews perform worse. Moreover, when homogeneity increases, the optimal binary review is asymmetric. Regardless of the degree of homogeneity, when the optimal threshold is used a binary review is preferred to a full signal if it is at least 3.25 times as frequent.

*I am grateful to Aislinn Bohren, Kevin He, and the rest of the Penn Theory group for extremely helpful feedback.

†University of Pennsylvania. Email: mjhm@sas.upenn.edu

1 Introduction

Transmitting information is costly. Yet, in many instances economic agents rely on other agents freely providing their information about some underlying uncertainty. In many settings, it is deemed infeasible or inappropriate to compensate individuals for sharing their information. In some cases this is due to concerns that compensation may bias the information that agents provide (e.g., Amazon prohibiting vendors from paying consumers for reviews). In others, compensation is simply not a viable option (e.g., a school eliciting students' course reviews). In these settings, the collector of information is reliant on the goodwill of agents providing their own information. Due to this tension—a principal wants information, but lacks a way to directly incentivize agents to provide it—in many settings platforms attempt to reduce barriers to information sharing as much as possible.

One natural way to reduce frictions in information-sharing is to ask for less information. Intuitively, reducing the number of questions that a reviewer must answer increases the likelihood that they will complete the review. This method of reducing frictions is not without a major drawback: the shorter review will elicit and thus provide less information. At the same time, if reviewers are more willing to provide coarser information, then this trade-off is not trivial.

This paper's main contribution is explicitly modelling and quantifying this trade-off. I show when a platform prefers to elicit coarsened information, and what coarsened information is optimal. I propose a model where eliciting more information from reviewers makes them less likely to submit their signal. A principal (platform), aiming collect information about an unknown state, must trade off between more frequent and better information. I quantify this trade-off when the number of potential reviewers is large. I show that the principal's preference over different review systems is independent of its decision problem, but is heavily dependent on the distribution of reviewers' signals. The former implies that the platform can apply the same methods across products and settings without having to closely monitor other aspects of the environment. This is especially relevant for large organizations, where the designers of review systems are not necessarily those using the information. The latter means that the there is a strong incentive for the platform to understand how the distribution of signals affects the trade-off between frequency and quality. The model succinctly provides conditions under which coarsened reviews lead to minimal loss of information, and then uses these conditions to justify why certain types of review systems are prevalent. In particular, my results predict the optimality of positively-skewed review systems and the prevalence of binary ratings systems.

In the model, a platform elicits information from a pool of agents (reviewers) who each have a signal that is informative of the underlying state of the world. These reviewers are interpreted as those who have previous experience about the state, and their signal represents their realized utility from this experience. The platform chooses a mapping from signals into reviews. Reviewers observe the mapping, and each chooses whether to submit their signal or not. This choice depends on a private willingness to review and the number of possible reviews they can leave: reviewers are less likely to leave a submission when they must decide between many possible reviews.

Motivated by online reviews, I focus on the case when there is a large number of reviewers. As the number of reviewers grows, the state will eventually be known with near certainty as long as a nonzero proportion leave reviews. What matters, then, is the *rate* at which the state is learned by the platform. This rate depends on both the average informativeness of a single review and the probability that a reviewer decides to submit one. Since this rate is independent of the platform's decision problem, this setting admits a decision problem-free ranking of review systems.

Technically, I leverage large deviations techniques in order to characterize this rate, for each review system. These techniques allow me to characterize the trade-off between frequency and quality of a review. The developed ranking is separable across this trade-off, enabling independent analysis. Since the rate of review can easily be computed empirically, I focus on characterizing the informativeness of a single review for different review systems. A key characteristic of online review platforms is that a single review is almost uninformative: individual taste differences mean that, on average, any one reviewer's experience is a very noisy signal about the underlying quality of the product. In this "scarce information" context, I derive a straightforward quantification for how much information is lost in terms of how much more frequently the coarse review to be submitted relative to a full-signal review for the platform to prefer the coarse review.

In practice, many discrete rating systems degenerate to a review system resembling a binary system. It has been well-documented (see Hu, Zhang, and Pavlou (2009)) that many 5-star review systems degenerate to a bimodal distribution. Even in this simple setting, the platform must determine how much information is lost when a simple positive/negative review is used as well as determine what is the optimal definition of "negative". I apply the results on quantifying information loss to answer these questions. Specifically, I study how the answers are influenced by the dispersion of taste shocks in the reviewers' signals. Here, larger dispersion reflects more heterogeneous experiences within the population for the product.

A key insight is that a more homogeneous population of reviewers implies that even when information is scarce, some reviewers have very informative signals. When coarse reviews are used, these very informative signals are mixed with a much larger mass of uninformative signals, drowning out their informational content. On the other hand, when reviewers are very heterogeneous, the variance of signal informativeness is small so coarsening does not mix as different information. As a result, more information is lost due to coarsening when the population in homogeneous.

Another key insight of the model is that the optimal threshold for a negative review is extremely important for minimizing information-loss. When reviewers are sufficiently heterogeneous, the optimal binary review system is simple: the platform optimally asks reviewers to symmetrically distinguish between a positive and a negative experience. However, this symmetric review is *not* optimal when reviewers are homogeneous: as agents become increasingly homogeneous, the optimal binary review system becomes increasingly asymmetric. If the platform were to use a symmetric system for homogeneous populations, the amount of information that it would lose would be unbounded. However, the optimal threshold bounds information loss. For any degree of homogeneity, if a reviewer is 3.25 times as likely to submit a binary review as their full signal, the optimal binary system outperforms the full system.

These results suggest that even extremely coarse reviews can be constructed such that they retain a relatively large amount of information. The framework also helps explain why the ratings systems that are observed in practice are implemented, even when they seem counter-intuitive. The stark asymmetry in review systems used by platforms may indeed be optimal, despite the appearance of sub-optimality. For instance, reviewers of many platforms have converged to skewing their review to be very positive (e.g., many users default to 5-stars on Uber and Airbnb), unless their experience was very negative. I show that if reviewers exhibit less negative than positive dispersion (i.e., they agree more on what constitutes a negative experience than a positive one), it is always optimal for the platform to isolate large negative experiences, while grouping all other experiences together. This results in a much larger number of positive reviews than negative reviews, even when the state is unfavourable.

The outline of the paper is as follows. Section 2 presents and discusses the model. Section 3 formalizes the trade-off between review informativeness and reviewers' frequency of review and determines optimal review systems when information is scarce. Section 4 applies these results to study the impact of heterogeneity on review systems. Section 5 discusses several extensions and Section 6 concludes. All proofs are included in the Appendix.

1.1 Related Literature

This paper is closely related to the literature studying learning dynamics in recommendation and review systems.

A large strand of this literature focuses on social learning dynamics. Ifrach et al. (2019) studies a standard social learning setting with binary reviews. Besbes and Scarsini (2018) looks at a similar setting, and focuses on when boundedly-rational agents use of summary statistics of reviews will lead to eventual learning of the state. Most similar to my work is Acemoglu et al. (2022), which studies learning rates for different review systems in the context of social learning. Although both their study and this work are interested in learning rates, there are several major differences. In particular, instead of focusing on the effect of social learning, I focus on quantifying the trade-off between different review systems. Additionally, I focus on a setting where a key difference in different review systems is the rate at which reviews are left. In the setting of Acemoglu et al. (2022), where reviews are automatically submitted, a finer review system is always preferred to a coarser system; in my setting this is not the case.

There is also a related literature that takes a different approach to studying recommendation and review systems. For instance, Che and Hörner (2018) studies a platform which aims to understand the quality of a product by "pushing" it to consumers. The signal structure in that model is perfect good news, but the platform cannot prove the good news to consumers. Another, very different approach is taken in Garg and Johari (2019), which uses large deviations techniques to optimally differentiate many underlying states by controlling the rate at which positive reviews are given for each quality level. A major difference between this work and Garg and Johari (2019) is that in that model, the designer has full control over the information system, whereas in my case the designer is constrained by the signal structure of the reviewers.

My work is also informed by the empirical literature on ratings. Looking at the common five-star system, it has been well documented that, across platforms, the distribution of submitted ratings is positively-skewed and bimodal (see Chen, Yoon, and Wu (2004) and Hu, Zhang, and Pavlou (2009)). Hu, Pavlou, and Zhang (2017) studies how this distribution is due to a self-selection effect in the population. Fradkin, Grewal, and Holtz (2021) and Fradkin and Holtz (2023) study possible interventions to improve the rate and quality of review submission. These works find that while these interventions (including incentivizing reviews) can increase the rate at which reviews are submitted, but do not tend to improve the overall informativeness of the review system. For a survey of empirical work aiming to understand review systems, see Magnani (2020). Recent empirical work has also provided alternative reasons to use coarsened information. For instance, Botelho et al. (2025) recently showed that moving from a five-star system to a binary rating decreased racial discrimination in an online platform that matches workers with customers. My work compliments this literature by providing an information-theoretic argument for using simple review systems.

My paper contributes also to the literature studying and applying rates of learning to understand economic contexts. This literature is broad, as learning rates have use in many settings. For instance, Rosenberg and Vieille (2019) and Hann-Caruthers, Martynov, and Tamuz (2018) study how quickly actions converge in a social learning setting, and Harel et al. (2021) and Dasaratha and He (2024) study the impact of networks on social learning.

Technically, this paper applies large deviations techniques to study rates of learning. There is a developed statistical literature that uses similar techniques to compare statistical experiments. Chernoff (1952) provides foundational results in the context of hypothesis testing. Torgersen (1981), which develops a generic ordering over different statistical experiments for finite decision problems when the number of signals is large, extends the ordering of Blackwell (1951). Large deviations techniques have been applied in several cases in the economics literature. Moscarini and Smith (2002) studies more sensitively the rate of learning, in order to determine characteristics of the demand for information. Frick, Iijima, and Ishii (2024b) ranks different misspecifications by the rate at which they slow learning.[1] Mu et al. (2021) develops a generalization of the ordering of Blackwell (1951) in the context of many signals. Finally, Fedorov, Mannino, and Zhang (2009) looks at similar trade-offs to those that I study in the context of hypothesis testing.

2 Model

There are two states of the world, $\Theta = \{L, H\}$. The state θ is drawn at the start and does not change. There is commonly held prior belief q on the state being $\theta = L$.

A principal, which I refer to as the platform, seeks to take an action depending on the state. It has finite action set A and state-dependent utility function $u : A \times \Theta \to \mathbb{R}$.

Reviewers have information about the state. There are N reviewers, indexed by $i \in \{1, ..., N\}$. Each reviewer has a signal $S_i \in \mathbb{R}$, drawn from measure space $(\Omega, \mathcal{F}, P)$. Conditional on θ the S_i are independent, with densities given by f_θ. Signals are informative, so that $f_L \neq f_H$. In addition to their signal, reviewers each have a willingness to review $W_i \in \mathbb{R}$. The W_i are independent from each other, and from the S_i.

The principal chooses a review system, which is a mapping from signals to reviews, $r : \mathbb{R} \to \mathbb{R}$ (with r measurable). If the platform chooses r, and a reviewer with signal s_i submits her signal, the platform observes $r(s_i)$. Let $R_r = r(\mathbb{R})$ denote the set of reviews possible under r.

To submit a review, a reviewer bears a cost, that can be interpreted as either a cognitive or temporal cost. This is given as $c(r) = C(|R_r|)$ for some increasing function $C : \mathbb{N} \cup \{\infty\} \to \mathbb{R}$. This captures the intuition that it is "easier" to submit

[1] Frick, Iijima, and Ishii (2023) and Frick, Iijima, and Ishii (2024a) apply similar techniques to different settings.

reviews when there are fewer choices.[2]

After the platform makes its choice of review system, reviewers' private signals and willingness to review are drawn, and reviewers then decide whether to submit their review. Reviewer i receives intrinsic utility $w_i - c(r)$ from reviewing, and 0 utility from not. It follows that reviewers will use a simple threshold rule: if $w_i \geq c(r)$ they will review, and if $w_i < c(r)$ they will not. Let $p_r := P(W_i \geq c(r))$ be the ex ante probability that a reviewer submits a review r.

After the N reviewers make their review decisions, the platform observes the submitted reviews, update its beliefs, and then takes an action to maximize its expected utility. Explicitly, let $I \subseteq \{1, ..., N\}$ be the set of reviewers who choose to submit a review. The platform observes $\mathcal{S}(I) = (\tilde{r}(s_1), ..., \tilde{r}(s_N))$, which is the vector of *submitted* reviews:

$$\tilde{r}(s_i) = \begin{cases} r(s_i) & \text{if} \quad i \in, \\ \emptyset & \text{if} \quad i \notin I. \end{cases}$$

After observing $\mathcal{S}(I)$, the platform updates its belief that the state is L. Let $\pi(\mathcal{S}) = \pi(L|\mathcal{S}(I))$ denote the posterior belief on the state being L after observing reviews $\mathcal{S}$. Here, and below, I drop the reliance on I for notational clarity.

After updating beliefs, the platform then chooses an action to maximize its expected utility given beliefs $\pi(\mathcal{S})$:

$$u^*(\mathcal{S}) := \max_{a \in A} \{\pi(\mathcal{S})u(a, L) + (1 - \pi(\mathcal{S}))u(a, H).\}$$

Since $\mathcal{S}$ is a random vector, $u^*(\mathcal{S})$ is stochastic. Define $u^*(r, N) := \mathbb{E}[u^*(\mathcal{S})]$, where this expectation is taken over the S_i and W_i. The objective of the platform is to maximize $u^*(r, N)$:

$$\max_r u^*(r, N).$$

[2] Much recent work has shown that individuals have preferences against complexity. See, for instance, Oprea (2020). In some cases, complexity has been conceptualized in a way dependent on the number of objects under consideration (e.g., Puri (2022)). Recently, Wang and Li (2025) studied the impact of increased cognitive load on user engagement in such systems, and finds that decreases in complexity lead to more user engagement.

2.1 Discussion of Model

This model flexibly captures situations where an agent must use information elicited from others in order to make informed decisions but the agent cannot force or otherwise incentivize the group with information to share their signals.

Importantly, I do not take a stance on the goal of the platform. In some settings, the platform's goal is to share the information that it collects with other agents (e.g., consumers of a product that is of unknown quality). The model naturally nests this setting, but does not assume that the platform's goals are aligned with future consumers. In this way, the model sheds insight on a variety of settings, from online retailers to general information dissemination platforms.

This highlights a major distinction between the platform in my model and previous models of review systems. For instance, the platform's goal in designing a rating/recommendation in Acemoglu et al. (2022) and Che and Hörner (2018) is to maximize the expected utility of consumers. In my work, the platform's incentives may not be aligned directly with consumers. This difference is due to the primary focus in those works being how platforms can incentivize early consumers to explore a product of unknown quality, whereas I abstract from the problem of experimentation. For this reason, I view the present work as complimentary with the aforementioned studies. Previous works provide guidance on how a platform should design the information that it provides to early consumers to incentivize experimentation, while my study suggests how the platform should collect information in order to further its own goals. In practice, it can simultaneously maximize the information that it collects for itself and then consider what information is optimal to show to consumers; the two objectives are not incompatible.

One of the main restrictions in the set-up above is the assumption that, given a review system r, the reporting rate p_r is uniform across the population of reviewers. In practice, this is a restrictive assumption: it is natural that reviewers with more extreme signals are more likely to submit reviews.[3] To account for this, I extend the model to allow for non-uniform reporting rates in Section 5.1. The main insights of the baseline model extend to that setting, and so I begin with a discussion of the uniform reporting rates for clarity of exposition.

[3]See, for instance Lafky (2014).

As a final note of discussion on the model, I assume throughout that there are no processing costs for the platform to interpret the submitted reviews. One of the practical benefits for using simple rating schemes is that they are easier to digest and summarize than full-length reviews. My model abstracts from this, and assumes that there is no cost to interpreting the reviews. Closely tied to this is the possibility of misinterpretation or miscommunication that are inherent to complicated reviews (e.g., free-text). I assume that when a reviewer submits a review, it is reported perfectly. Abstracting from these forces suggests that my model overstates the amount of information loss that stems from using coarsened signals, since removing these forces makes full-signal reports more accurate.

3 Quantifying Information Loss

In this section, I quantify the trade-off between informational content and frequency of report for different review systems by applying large-deviations techniques. I first show how a platform should trade-off between quality and frequency in review systems (with an appropriate definition of quality). I then characterize how much information a review systems loses relative to the full review system when information is scarce.

3.1 Trading-off between Quality and Frequency in Reviews

Fix a review system r. Denote the measure of reviews under r conditional on state L as γ_L^r. That is, for any measurable $B \subseteq R_r$ (where measurability is inherited from the Borel σ-algebra),

$$\gamma_L^r(B) := \mathbb{P}_L(r^{-1}(B)) = \int_{\{s:r(s)\in B\}} f_L(s)ds.$$

Similarly, let γ_H^r denote the measure corresponding to the distribution of reviews in state H. These measures reflect the distribution of reviews r *conditional* on reviews being submitted.

The performance of a review system depends on the rate at which the state is revealed under the review system. Intuitively, the more different are γ_H^r and γ_L^r, the faster will be learning. The measure of difference in this context is the Hellinger transform between the two distributions. Consider, for a fixed review $r_i \in R_r$, the log-likelihood ratio of observing r_i in states L and H:

$$\log\left(\frac{d\gamma_L^r(r_i)}{d\gamma_H^r(r_i)}\right). \tag{1}$$

In state H, learning will be good if (1) small, on average. The Hellinger transform is moment-generating function of (1) in state H. The smaller the value of the Hellinger transform, the more likely the platform will think that the state is H when the true state is H after observing reviews.

Definition 1. The *learning efficiency* of review system r is defined to be 1 minus the minimum of the Hellinger transform between γ_L^r and γ_H^r:

$$\nu(r) := 1 - \min_{\lambda\in[0,1]} \int_{R_r} \left(\frac{d\gamma_L^r}{d\gamma_H^r}\right)^{\lambda} d\gamma_H^r \in [0,1].$$

The learning efficiency is a simple transformation of the minimal value of the Hellinger transform. The less overlap between the two distributions, the larger the learning efficiency becomes. At one extreme, when $d\gamma_L^r > 0 \iff d\gamma_H^r = 0$, then $\nu(r) = 1$, so that learning occurs after one review is received. At the other extreme, when signals are entirely uninformative, $\gamma_L^r = \gamma_H^r$, $\nu(r) = 0$, and learning never occurs.

To see why the learning efficiencies $\nu(r)$ are important, consider first the case that all reviews are submitted ($p_r = 1$). Then, beliefs after observing N reviews will be determined by the behaviour of the sum of the log-likelihood ratios. When N is large, the probability that the platform takes an action that is not optimal given the state is determined by the tail behaviour of these log-likelihood ratios. This behaviour is captured by $\nu(r)$, so that the learning efficiency captures the rate at which the platform learns the state. As a result, *regardless* of the platform's utility function u, $\nu(r)$ determines how quickly its expected utility converges its the full-information utility. Hence, if $\nu(r) > \nu(r')$, for large N the platform will have higher expected utility under r than under r'.[4]

When reviews are not always submitted, the rate at which the platform learns

[4]This discussion summarizes a known result in the literature (see, for instance, Chernoff (1952) and Torgersen (1981)).

the state is not $\nu(r)$. Intuitively, if reviews are submitted half of the time, the rate of learning should be halved. The main result of this section is that this intuition is correct: regardless of the decision problem, for large N, $p_r\nu(r)$ determines the utility of the platform. This suggests that while the platform must trade-off between the rate at which reviewers are submitted and the quality of the reviews, the trade-off is easily separable.

Theorem 1. *Fix two review systems r, r' with corresponding rates of review $p_r, p_{r'}$ such that $p_r\nu(r) > p_{r'}\nu(r')$. For any action set A and utility function u, there exists an $\overline{N}$ such that for all $N \geq \overline{N}$, such that $u^*(r, N) > u^*(r', N)$.*

Theorem 1 provides a straightforward way for a platform to trade-off between frequency of reviews and their quality. Since $p_r\nu(r)$ is independent of the decision problem of the platform, the platform need not tailor its review design on the specifics of each problem. This is a particularly important consideration for large organizations where designers of the review system may be different from those using the information.

The proof of Theorem 1 utilizes existing results in the large deviations literature which formalize the intuition above for the comparison when $p_r = 1$. In order to leverage those results, I construct a statistical experiment with full reporting whose learning efficiency is $p_r\nu(r)$ and show that this experiment is equivalent to the review system with stochastic reporting.

In general, the p_r are easy to empirically compute for platforms: a large platform can easily implement an A/B test to determine the relative reporting rates for two different review systems. This suggests that it is beneficial to focus directly on the computation and comparison of the $\nu(r)$. An equivalent condition to that of Theorem 1 is, if

$$\frac{\nu(r')}{\nu(r)} < \frac{p_r}{p_{r'}}, \tag{2}$$

then there exists an $\overline{N}$ such that for all $N \geq \overline{N}$, such that $u^*(r, N) > u^*(r', N)$. The value $\nu(r')/\nu(r)$ determines the benchmark (in terms of relative frequency of reporting) that the platform should use in choosing between two review systems. Consider the full review $r_f(s) = s$. By setting $r' = r_f$, the value (2) becomes a measure of how much information is lost in review r relative to the full review r_f.

Definition 2. The *learning loss* of a review function r is $\kappa(r) := \frac{\nu(r_f)}{\nu(r)} \in [1, \infty)$.

The learning loss $\kappa(r)$ of a review system r is how much more frequently a review of type r needs to be submitted relative to a full review for the platform to prefer using reviews of type r to full reviews. For instance, if $\kappa(r) = 2$, then r needs to be submitted twice as frequently as a full review for the platform to prefer r. To compare two different review systems r and r', the platform simply uses $\kappa(r)/\kappa(r')$. The main work of the following section is to derive properties of the learning loss in different settings and for different review systems.

Remark 1. It is important for the separability of $\nu(r)$ and p_r that the reporting rate p_r is uniform across the population of reviewers. If this is not the case (i.e., individuals with more extreme signals are more likely to leave a review), then the two are no longer completely separable. However, insights from the general model still apply in that setting. See Section 5.1 for an extension that allows an individual reviewer's reporting rate to depend on their realized signal.

Remark 2. That the p_r represent the stochastic rate at which signals are reported is important. An alternative to this approach is to ask how many signals n_r from r "equals" a signal $n_{r'}$ from r'. I discuss this alternative and its relationship to my specification in Section 5.2.

3.2 Finite Reviews and Scarce Information

When individual signals are not very informative, learning from any review system will be slow. In this setting, optimizing the rate of learning will be even more important. This section studies the case of scarce information, and classifies $\kappa(r)$ in this setting.

In order to parameterize informativeness, I now assume that a reviewer's signal is given by $S_i = u_\theta + \epsilon_i$, where

$$u_\theta = \begin{cases} -\mu & \text{if} \quad \theta = L \\ \mu & \text{if} \quad \theta = H \end{cases}$$

The parameter μ determines the separation between the two distributions: the larger the value of μ the easier it is to distinguish the two distributions. For large μ the two distributions overlap little different, and for small μ the two are almost indistinguishable. The ϵ_i reflects reviewer taste idiosyncrasies: different reviewers will experience the same product differently.

Assumption 1. The idiosyncratic taste noise ϵ_i are i.i.d., with continuous, full-support, log-concave density given by f.

In this setting, the densities of signals in state L and H are given by

$$f_L(s) = f(s + \mu) \quad \text{and} \quad f_H(s) = f(s - \mu).$$

The two primitives are f and μ. Given a review system r, write the learning efficiency of r when the separation is μ as $\nu_f(\mu; r)$. Similarly, write the learning loss of r as $\kappa_f(\mu; r)$. When it is clear from the context, I drop the dependence on f.

Threshold Rules: In general, the platform's choice of a finite review with $|R_r| = k$, is decide on a partition of $\mathbb{R}$ into k sections. However, review systems that are used in practice are simple partitions: a more negative experience leads to a more negative rating. For arbitrary f, these will not in general be optimal. The assumption of log-concavity guarantees the optimality of these "threshold" review systems.

Explicitly, a review system r is called a k-threshold rule if there exist $\tau_0 \leq \tau_1 \leq ... \leq \tau_k$ with $\tau_0 = -\infty$ and $\tau_k = \infty$ such that

$$r(s) = \sum_{i=1}^{k} r_i \mathbb{1}\left\{s \in (\tau_{i-1}, \tau_i]\right\}$$

In this case, I write $\tau^k = (\tau_1, ..., \tau_{k-1})$, and write r_{τ^k} for the review system defined by thresholds τ^k. Optimality of threshold rules follows from (i) the log-concavity of f guaranteeing that the more negative the signal, the larger the belief that state is $\theta = L$, and (ii) it is always being in the platform's interest to group signals that are more informative of one state than another together.

Lemma 1. *Suppose that f is log-concave. Then for any μ, consider a review system r with $|R_r| = k$. There exists a k-threshold rule r_{τ^k} such that $\nu(\mu; r_{\tau^k}) \geq \nu(\mu; r)$.*

If a platform wishes to implement an optimal review system with k reviews, it optimizes over the $k - 1$ thresholds τ_i.

Scarce Information: Reviews are almost uninformative exactly when $\mu \approx 0$. For any review system (including r_f), $\lim_{\mu\to 0} \nu(\mu; r) = 0$. However, learning loss remains well-defined in the limit, and so for the remainder of the section I study the properties of

$$\kappa(0; r_{\tau^k}) := \lim_{\mu\to 0} \kappa(\mu; r_{\tau^k}) = \lim_{\mu\to 0} \frac{\nu(\mu; r_f)}{\nu(\mu; r_{\tau^k})}. \tag{3}$$

Although $\kappa(0; r_{\tau^k})$ is not well-defined, I abuse notation and write it to represent the limit in (3).[5] This limit represents a threshold in the relative rates of reporting that make the platform prefer eliciting reviews r_{τ^k} to eliciting full reviews. If the limit (3) is large, then when information is scarce, a binary review system with thresholds τ^k performs poorly relative to full reviews. On the other hand, if (3) is small, then even when reviewers are slightly more likely to leave a binary review, the platform would prefer asking for coarse information.

The main result of this section is the explicit characterization of (3). Before I introduce the result, some intuition is useful. When $\mu = 0$, $\nu(\mu; r_f) = 0$, since there is no difference in the distribution of signals in the two states. When the separation μ increases from 0, on regions where f is nearly constant, signals will continue to be uninformative. However, on regions where f is very curved, even a small shift in μ causes signals in that region to become very informative, because the posteriors that these signals induce will shift greatly. A larger curvature of f is then associated with a larger value of (3). The measure of curvature that captures this effect is the *Fisher Information* of f (with respect to μ):

$$I_f := \mathbb{E}\left[\left(\frac{f'(s)}{f(s)}\right)^2\right].$$

A similar logic determines the rate at which $\nu(\mu; r_{\tau^k})$ increases as the separation μ increases from zero. Since the review is discrete, the effect of a change in μ will only impact the reviews at the thresholds $\tau_1, ..., \tau_{k-1}$. Consider that for a single review r_i, a marginal increase in the separation will cause the probability that r_i is reported in state L to increase by $f(\tau_i) - f(\tau_{i-1})$. The probability that r_i is reported in state H

[5] The optimal $(\tau^k)^*$ will depend on μ. However, it is sufficient to set $(\tau^k)^* = \lim_{\mu\to 0}(\tau^k)^*(\mu)$ as $\lim_{\mu\to 0} \frac{\nu(\mu;(\tau^k)^*(\mu))}{\nu(\mu;(\tau^k)^*)} = 1$.

increases by exactly the opposite: $f(\tau_{i-1}) - f(\tau_i)$. The larger the absolute value of this difference, the more informative review r_i will be. This effect is magnified if r_i is uncommon: it is easier then to attribute the review r_i to this change.

The following assumption is a regularity assumption on f, guaranteeing that the Fisher information is well-defined.[6]

Assumption 2. The density of idiosyncratic taste shocks f is absolutely continuous almost everywhere, as is its derivative.

These effects come together in the following major result of this section:[7]

Theorem 2. *Suppose that f that satisfies Assumptions 1 and 2. Let F denote the distribution function of f. Then, learning loss with scarce information is given by*

$$\kappa(0; r_{\tau^k}) = I_f \Bigg/ \left(\sum_{i=1}^{k} \frac{\left(f(\tau_i) - f(\tau_{i-1})\right)^2}{F(\tau_i) - F(\tau_{i-1})} \right) \in [1, \infty).$$

Theorem 2 gives a straightforward way to determine the degree of information loss for different review functions. It is proved by showing independently the rate at which $\nu(\mu; r_f)$ and $\nu(\mu; r_{\tau^k})$ go to zero as a function of the separation of the distributions.

An important corollary of Theorem 2 is an explicit characterization of the optimal finite review system with $|R_r| = k$.

Corollary 1. *The maximizer of*

$$\sum_{i=1}^{k} \frac{\left(f(\tau_i) - f(\tau_{i-1})\right)^2}{F(\tau_i) - F(\tau_{i-1})}.$$

is the optimal k-threshold rule r_{τ^k} under scarce information.

Corollary 1 shows that the distribution of taste shocks in the population has important implications for the design of review systems. The next section applies these results to analyze the impact of one key characteristic of taste shocks: heterogeneity.

[6]This condition can be weakened. For instance, a slightly weaker assumption is that f is twice continuously differentiable almost everywhere, with the first and second derivative of $f(x+\mu)^\lambda f(x-\mu)^{1-\lambda}$ with respect to μ integrable for μ in a neighbourhood of 0.

[7]With the notational convention that $f(\infty) = f(-\infty) = 0$, and $F(\infty) = 1, F(-\infty) = 0$.

Remark 3. An alternative to parameterizing the distribution of signals by the separation μ would be to vary the variance:

$$f_L(s;\sigma) \propto f\left(\frac{s+\mu}{\sigma}\right) \quad \text{and} \quad f_H(s;\sigma) \propto f\left(\frac{s-\mu}{\sigma}\right).$$

Scarce information in this setting requires taking $\sigma \to \infty$. The choice between the separation μ and variance σ parameterization is stylistic: the two yield the same results. A parameterization of (μ, σ) yields the same information as $(\frac{\mu}{\sigma}, 1)$. Hence, there is a direct analog of Theorem 2 for the variance parameterization.

To justify the choice of one over the other, a drawback of varying σ is that the notation regarding an optimal threshold becomes more cumbersome. Instead of having a well-defined limit τ, instead the convergence is in terms of $\frac{\tau(\sigma)}{\sigma}$. I emphasize again that the two are equivalent.

4 Application: Binary Reviews and Heterogeneity

In this section, I show how the results from the previous sections can be used to show how optimal review systems vary with important economic characteristics. In particular, I analyze the impact of taste heterogeneity. I focus primarily on optimal binary reviews for clarity of exposition.

An important economic characteristic when eliciting information is how heterogeneous or homogeneous is the population of individuals whose information is being elicited. In the context of product reviews, different products are likely to induce a different degree of homogeneity in the population. Preferences over products like movies (or other forms of art) are likely to be quite dispersed because individual taste matters to a large degree. However, for simple products like toasters, preferences are likely to be much less dispersed, since experiences are much more uniform.[8] Since homogeneity of experience differs so greatly across products, it is important to understand how it affects review systems.

[8] One possible way to distinguish between goods with low and high taste homogeneity would be the classification of "search" and "experience" goods used in the empirical literature. First introduced by Nelson (1970), search goods are those goods whose attributes are well-defined and easily found, whereas experience goods are those which must be experienced before an opinion can be formed (see also Magnani (2020) for a discussion). Search goods, because of their well-defined attributes,are likely to be more homogeneous (does the good perform as it should), while experience goods are likely to be more heterogeneous (since individual experience matters more).

In order to understand how taste heterogeneity affects review systems, I restrict attention to a one-parameter family of distributions where the parameter can be interpreted as the degree of homogeneity within the population of reviewers:

$$f_\alpha(s) \propto \exp(-|s|^\alpha). \tag{4}$$

As discussed in Remark 3, because I focus on the case of scarce information, the variance of signals does not impact learning loss. Larger tails correspond to greater dispersion, and hence greater heterogeneity in idiosyncratic tastes. The parameter α in the (4) captures this tail behaviour. The larger the value of α, the smaller are tails, and so, the smaller the dispersion of taste shocks. Henceforth, I refer to α as homogeneity.

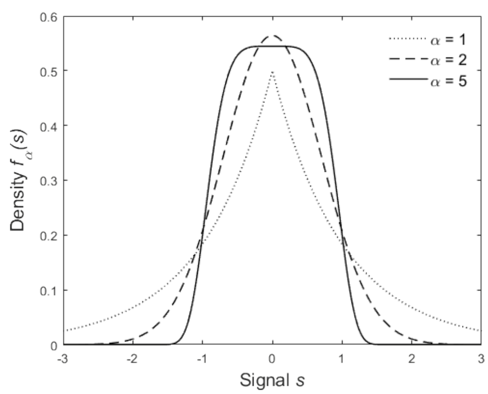


Figure 1: f_α for $\alpha = 1, 2, 5$.

Remark 4. For $\alpha \geq 1$, $f_\alpha(s)$ is log-concave, and so from Lemma 1 a threshold rule is optimal. This family is sometimes referred to as the Generalized Normal Distributions and incorporates three well-known distributions as special cases: f_1 is the Laplace distribution; f_2 is the Normal distribution; and as $\alpha \to \infty$, f_α converges to the uniform distribution on $(-1, 1)$. The uniform distribution reflects the largest degree of homogeneity possible. The normal distribution reflects a small level of homogeneity, and the Laplace distribution reflects the largest possible degree of heterogeneity. Since variance does not affect learning loss, the scale/variance parameter is normalized to 1. Figure 1 exhibits $f_\alpha(s)$ for different values of α.

Studying this class of distributions allows me to show how heterogeneity impacts information loss and the optimal choice of threshold for binary review systems. To motivate why properties of the optimal threshold are important, I first show what happens when the platform uses a naive threshold: asking reviewers if their experience was positive or negative. This behaviour is captured in r_0, the review system with the symmetric threshold of 0.

Perhaps surprisingly, when the population is very heterogeneous, this threshold leads to small learning loss. However, as homogeneity grows, this does not continue: the naive threshold leads to unbounded loss in information.

Lemma 2. *Consider learning loss under the naive threshold 0, $\kappa_{f_\alpha}(0; r_0)$. Then,*

(i) $\kappa_{f_1}(0; r_0) = 1$*; and*

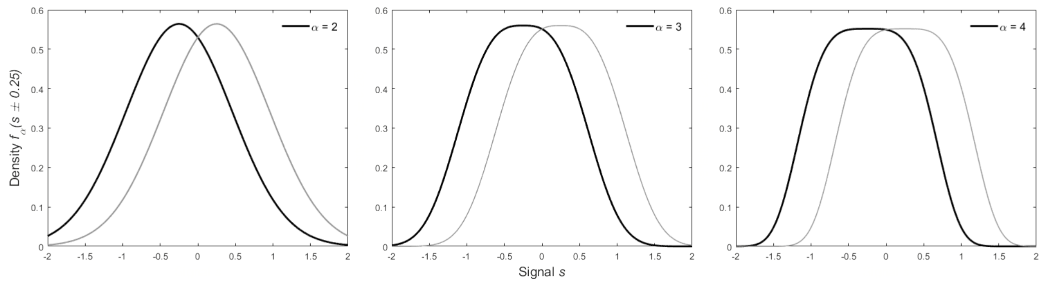
(ii) $\kappa_{f_\alpha}(0; r_0)$ *is increasing in* $\alpha \geq 1$ *with* $\lim_{\alpha \to \infty} \kappa_{f_\alpha}(0; r_0) = \infty$*.*

Lemma 2 shows that under the naive binary review, information loss is increasing in the degree of homogeneity of the population. The intuition is as follows. When idiosyncratic tastes are less dispersed, while many signals are very uninformative, there are also more signals that are very informative (i.e., some reviewers are very representative). See Figure 2, which shows this visually. As homogeneity α increases, simultaneously more signals are very uninformative and more signals are very informative.[9] In posterior space, this is reflected in a larger spread of the induced posteriors. Using the symmetric review r_0 mixes very informative signals with very uninformative signals, limiting the information contained in either review.

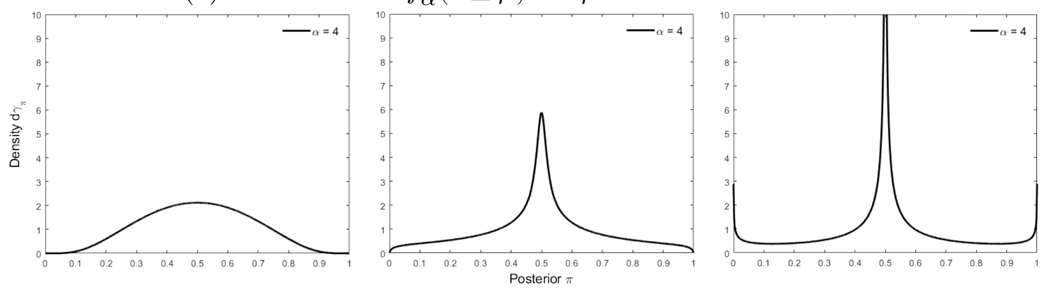
This highlights the central tension in choosing an optimal threshold: extracting more information from some signals while extracting less from others. For values of α that are close to 1, $\kappa_\alpha(0)$ is close to the lower bound of 1, so it is unlikely that a different threshold could improve upon this. However, as discussed above, when α is large, a small separation μ induces a small mass of extremely informative signals. This suggests that the incentive to choose an asymmetric threshold increases as α grows.

As shown below, this intuition is correct. As homogeneity α increases, the asymmetry of the optimal threshold increases. Importantly, the optimal threshold limits learning loss under the binary review. While learning loss is still increasing in the degree of homogeneity of the population, this optimal learning loss information loss

[9] Appendix B explores this relationship in more detail.



(a) Distributions $f_\alpha(s \pm \mu)$ for $\mu = 0.25$ and different α.



(b) Distributions $f_\alpha(s \pm \mu)$ for $\mu = 0.25$ and different α.

Figure 2: Different distributions of $f_\alpha(s \pm \mu)$ and corresponding ex-ante distributions of posteriors (for uniform prior) for $\mu = 0.25$ and different values of α.

is bounded above. If the population is 3.25 times as likely to submit a binary review than their full signal, the optimal binary system should be preferred, regardless of the degree of homogeneity. This highlights the importance of the analysis: the choice of the threshold has extremely important implications for the performance of binary reviews.

In what follows, let $\tau^*(\alpha)$ be the nonnegative value that maximizes[10]

$$\frac{f_\alpha^2(\tau)}{F_\alpha(\tau)} + \frac{f_\alpha^2(\tau)}{(1 - F_\alpha(\tau))}. \tag{5}$$

Proposition 1. *For $\alpha \geq 1$, there is a unique maximizer $\tau^*(\alpha)$ to (5) on $[0, \infty)$. It has the following properties:*

(i) If $\alpha \leq 2$, $\tau^*(\alpha) = 0$, while if $\alpha > 2$, $\tau^*(\alpha) \neq 0$;

(ii) $\lim_{\alpha \to \infty} \tau^*(\alpha) = 1$; and

[10]Since f_α is symmetric, there will be also a nonpositive maximizer $-\tau^*(\alpha)$. In the statement of Proposition 1, everything is written in terms of the nonnegative optimal τ^* for clarity of exposition. Comparable statements for the nonpositive optimal threshold hold. The next section shows that introducing (arbitrarily small) asymmetry in f_α breaks this multiplicity.

(iii) $\kappa_{f_\alpha}(0; r_{\tau^*(\alpha)}) < 2e^2 \int_1^\infty \frac{e^{-s}}{s} ds$ *for all* α*, with equality in the limit. This value is less than* 3.25*.*

The normal distribution represents the degree of homogeneity at which the optimal threshold stops being symmetric. It is exactly those populations that are more homogeneous than the normal distribution that have a nonzero optimal threshold: When the degree of homogeneity is large the platform optimally asks individuals either (i) whether they had a very good experience or an experience that was not exceptional (positive τ^*), or (ii) whether they had a very bad experience or an experience that was not horrendous (negative τ^*).

Figure 3 plots comparative statics for $\tau^*(\alpha)$: both $\tau^*(\alpha)$ and $F_\alpha(\tau^*(\alpha))$ are increasing in α. The interpretation of the first is straightforward: the optimal threshold is increasing. The interpretation of the second is slightly more subtle: the probability of the more common review increases in α, *regardless of the state*. This means that (using the positive optimal threshold), for large α, even when the state is high, the majority of reviews will be negative. This fraction increases to 1 in the limit, showing that for large degrees of homogeneity, the optimal review system appears to degenerate. This suggests that even review systems where the majority of reviewers leave the same review may not be sub-optimal. These results are summarized in Proposition 2. Figure 3 also shows that, despite the improvements coming from the optimal threshold, learning loss is still increasing in homogeneity.

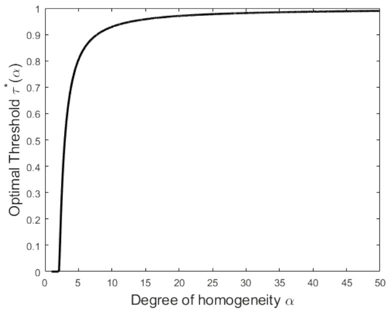
Proposition 2. *Consider the non-negative optimal threshold* $\tau^*(\alpha)$*. The following are true:*

(i) The optimal threshold is increasing in the homogeneity of the population: $\tau^*(\alpha)$ *is strictly increasing in* α *for* $\alpha \geq 2$*; and*

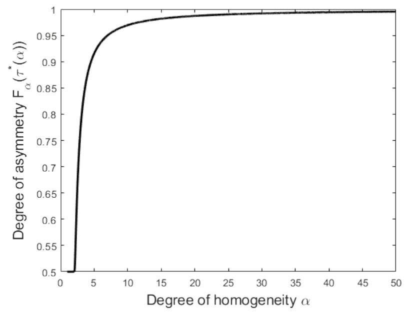
(ii) Optimal asymmetry is increasing in the homogeneity of the population: $F_\alpha(\tau^*(\alpha))$ *is strictly increasing in* α *for* $\alpha \geq 2$*.*

4.1 Asymmetric Heterogeneity

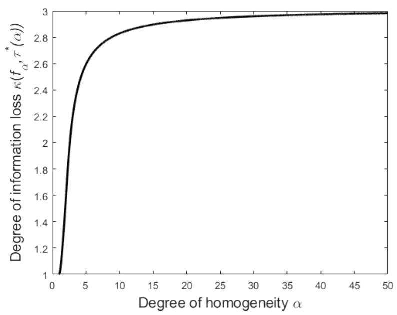
The multiplicity of optimal thresholds in the above results is concerning because the resulting reviews result in very different behaviour. If the platform were unable to consistently coordinate reviewers, then the review system could easily break-down.



(a) Optimal thresholds as a function of homogeneity.



(b) Optimal degree of asymmetry, as a function of α.



(c) Minimal information loss, as a function of homogeneity.

Figure 3: Plots showing the comparative statics of Proposition 2.

This section shows that this multiplicity disappears when any asymmetry in the distribution of taste shocks is introduced. The asymmetry that I introduce is asymmetry in the degree of homogeneity for positive and negative idiosyncratic tastes. For any nonzero difference in homogeneity on the two sides of the distribution, the platform optimally chooses a threshold on the side of the distribution that is more homogeneous.

Apart from breaking multiplicity, this is an important setting to study because asymmetric homogeneity is a relevant characteristic in many cases. In hiring a taxi or ride-share service, the qualities that lead to a very bad experience are likely to not differ greatly across individuals: no customer wants to be late or get into an accident. On the other hand, the qualities that lead to a very good experience may differ across individuals: some (not all) people like having a long discussion with their driver, some (not all) like music to be played.[11] In these contexts, where negative experiences are

[11] As another example, consider again search and experience goods. For most search goods, in-

more homogeneous, the platform optimally uses negative reviews only for very bad experiences and uses positive reviews for any experiences that are at least mediocre. This suggests a robust skewing towards positive reviews. This is a characteristic that is observed in many review systems: most reviews on Airbnb are 5-stars, as are those on Uber (see for instance Hu, Zhang, and Pavlou (2009)). The below result suggests that this skewing is consistent with the optimal review.

Let α^* be the value for which $f_\alpha(0)$ is maximized.[12]

Proposition 3. *Let*

$$f_{\alpha_L,\alpha_H}(s) \propto \mathbb{1}\{s \geq 0\}e^{-|s|^{\alpha_H}} + \mathbb{1}\{s < 0\}e^{-|s|^{\alpha_L}}$$

with $\alpha^* \leq \alpha_H, \alpha_L$. If $\alpha_L > \alpha_H$, then, the optimal threshold $\tau^*(\alpha_L, \alpha_H)$ is negative. If $\alpha_H > \alpha_L$, then the optimal threshold is positive.

Figure 4 highlights the intuition behind Proposition 3. Consider using a positive threshold (dotted line). At the negative threshold that yields the same value of the density (dashed line), it is easily seen that there is less mass to the left of that threshold than there is to the right of the positive candidate. This logic applies to any such threshold, so that the optimal threshold must be negative. A similar logic applies if, instead of increasing the homogeneity on one side of the distribution, the variance of signals on one side of the distribution were increased. In that case, to, the optimal threshold would necessarily be negative.

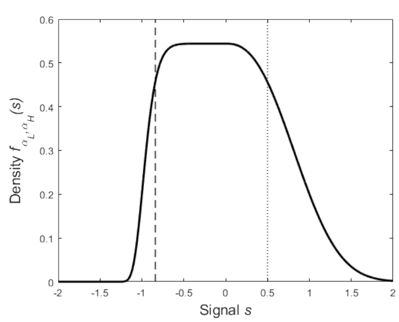


Figure 4: f_{α_L,α_H}, for $\alpha_L = 10, \alpha_H = 2.5$.

creased negative homogeneity is expected: most reviewers will be in agreement about whether a good doesn't meet expectations, but there may be disagreement about how it over-performs.

[12]It is easy to verify that $\alpha^* \approx 2.166$.

As mentioned, Proposition 3 helps provide justification for why in practice platforms like ride-sharing apps exhibit many more positive reviews than negative reviews. If people agree on negative experiences more than they do positive experiences, for optimal learning, the platform *should* to isolate extremely negative experiences and thus choose a system that frequently degenerates on a large portion of the population leaving the same review, regardless of the underlying state.

4.2 Heterogeneity with Multiple Bins

Similar qualitative results apply to the finite review systems with k bins, where $k > 2$. Figure 5 exhibits the optimal three-bin system for different degrees of homogeneity α. In particular, learning loss is increasing in the degree of homogeneity, and the optimal asymmetry of the system is also increasing in the degree of homogeneity. An important difference here over the binary system is that *both* sides of the distribution can be isolated. What is important is that as homogeneity grows, it is increasingly important to "throw out" large portions of the population. The third bin does not provide any information to the platform (since it is equally likely in each state). Its presence is important, however, because it allows the platform to effectively skim uninformative reviews in order to isolate strong signals on both sides of the distribution.

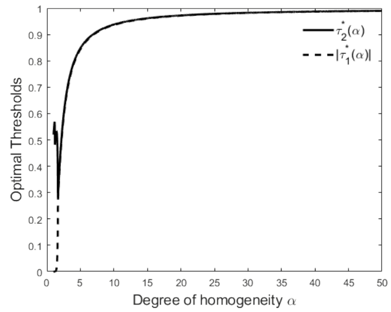
5 Extensions

5.1 Non-Uniform Reporting Rates

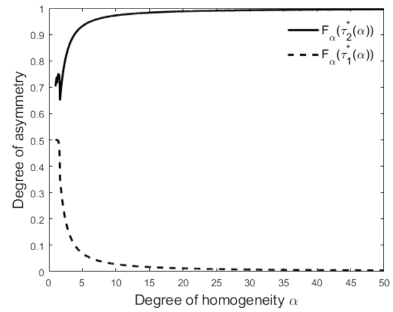
I assume above that given a review system r, the rate of reporting p_r is uniform across the population of reviewers. This is a restrictive assumption; in practice a reviewer's experience is likely to affect the probability that they submit a review. It is reasonable that a reviewer with an extreme review is more likely to leave a review, and in some contexts reviews may differentially report positive or negative experiences.[13] In this section I show how to incorporate non-uniform reporting rates into the framework studied above. While the qualitative insights from previous sections remain, several new forces emerge.

Suppose that a reviewer with signal s facing review r has a probability of leaving a review given by $p(r, s)$, dependent on both the review system *and* the reviewer's

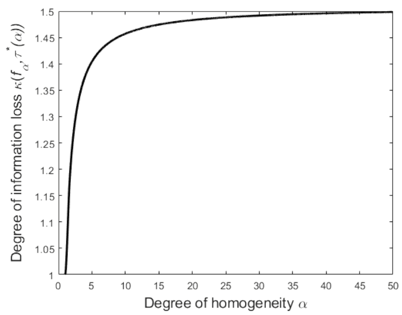
[13] For instance, Lafky (2014) suggests that individuals are more likely to report extreme experiences than moderate experiences because they derive a benefit from informing others, and extreme experiences are more informative.



(a) Optimal thresholds as a function of homogeneity.



(b) Optimal degree of asymmetry, as a function of α.



(c) Minimal information loss, as a function of homogeneity.

Figure 5: Plots for optimally chosen three-bin review system.

signal. As before, it is the difference in review measures across the two states that determines the rate of learning of the platform. In this setting, however, it is not possible to separate the propensity of reviewing from the underlying signal distribution. Explicitly, define the measure $\gamma_L^{(p,r)}$ of reviews in state L (implicitly dependent on μ) as

$$\gamma_L^{(p,r)}(B) := \int_{\{s:r(s)\in B\}} p(r,s)f(s+\mu)ds.$$

Similarly define the measure $\gamma_H^{(p,r)}$ of reviews in state H. As before, these measures are used to determine the rate at which the platform learns the state (the analogue of $p_r\nu(r)$) in this setting:

$$\rho(\mu; r, p) := 1 - \min_{\lambda\in[0,1]} \left[\int_{R_r} \left(\frac{d\gamma_L^{(p,r)}}{d\gamma_H^{(p,r)}} \right)^{\lambda} d\gamma_H^{(p,r)} + \left(\gamma_L^{(1-p,r)}(\mathbb{R})\right)^{\lambda} \left(\gamma_H^{(1-p,r)}(\mathbb{R})\right)^{1-\lambda} \right].$$

There are two significant differences between ρ and $p_r\nu(r)$. First, there is a mechanical change: since $p(r,\cdot)$ varies with the signal realization, different signals receive different weight. Second, there is a conceptual change: "null reviews" (reviews that are not submitted) can now be informative of the state. Null reviews are an additional signal: if reviewers are more likely to be submitted when the state is L, a lack of reviews is an indication that the state is H.

Theorem 1 generalizes to this setting, where the ρ take the role of $p_r\nu(r)$.

Lemma 3. *Fix two review systems r, r' such that $\rho(\mu; r, p) > \rho(\mu; r', p)$. For any action set A and a utility function u, there exists an $\overline{N}$ such that for all $N \geq \overline{N}$, such that $u^*(r, N) > u^*(r', N)$.*

Lemma 3 shows that even with full flexibility in reporting rates a generic ordering over review systems exists. Without additional structure on the propensities to review p, little more can be said about the comparison. When arbitrary functions p are allowed, then the underlying distribution of signals that the review systems draw from differ, and anything can happen. In order to gain structure, I restrict attention to functions p that are separable between the impact of the review function and the signal realization.

Definition 3. The propensity to review function $p(r,s)$ is said to be *separable* if $p(r,s) = p_r q(s)$ for some functions $p : \mathbb{N} \cup \{\infty\} \to [0,1]$ and $q : \mathbb{R} \to [0,1]$.

This specification represents a two-stage procedure. The first stage reflects individuals who are willing to ever submit a review of type r (a fraction p_r of the population). In the second stage, these individuals make their decision based off of their signal realization: among these, only fraction $q(s)$ if those with signal s will decide to leave the review.[14] This reflects that even reviewers who sometimes leave reviews will not always leave them, and this choice may be influence by their experience.[15]

To understand the new forces that enter when reporting rates are not uniform, I now develop an analogue of Theorem 2 in the case that propensities to review

[14] A micro-foundation for this two-stage procedure is given in Appendix **??**.

[15] Importantly, q depends on the signal and *not* the taste shock of the individual. This latter case would simply reduce to the setting of the previous sections.

are separable. Before I present the result, additional notation is needed. Fix the distribution of taste heterogeneity f and the signal reporting rate q. Define the functional E as the integral of a function g with respect to the measure whose density is given by $q \cdot f$:

$$E(g) := \int_{-\infty}^{\infty} g(s)q(s)f(s)ds.$$

Note that E is not an expectation because in general $q \cdot f$ is not probability distribution (unless $q \equiv 1$ almost everywhere). Similarly, given a review function r, and a review $\tilde{r} \in R_r$, denote by

$$E(g;\tilde{r}) := \int_{-\infty}^{\infty} g(s)\mathbb{1}\{r(s) \in \tilde{r}\}q(s)f(s)ds$$

the integral with respect to the restricted measure. With this notation, Theorem 2 generalizes to the following result.

Theorem 3. *Suppose that f satisfies Assumptions 1 and 2 and $p(r,s) = p_r q(s)$ where q is continuous. Let $\ell_f := \frac{f'}{f}$ denote the linear score of f. Then learning-loss for a finite review system r is given by*

$$\frac{p_r}{p_f} \cdot \lim_{\mu \to 0} \frac{\rho(\mu; p_f \cdot q)}{\rho(\mu; r, p_r \cdot q)} = \frac{E\left(\ell_f^2\right) + \frac{p_f E(\ell_f)^2}{1 - p_f E(1)}}{\sum_{\tilde{r} \in R_r} \frac{E(\ell_f;\tilde{r})^2}{E(1;\tilde{r})} + \frac{p_r E(\ell_f)^2}{1 - p_r E(1)}}. \tag{6}$$

The left-hand side of (6) is the generalization of κ from Section 4. In the case that $q \equiv 1$ and r is a threshold rule, Theorem 3 is a generalization of Theorem 2 for non-threshold finite reviews. The first term of the numerator and the sum in the denominator are the direct generalizations of the numerator and denominator from Theorem 2.

The impact of null reviews is reflected in the second term of the numerator and denominator. Importantly, the information contained in null reviews is related to the rate at which a review is submitted. In Figure 6, the top two panels show the distribution of submitted reviews. As p_r increases, these distributions are scaled. However, this is not the case for null reviews, which are shown in the bottom two panels. As the fewer reviewers report ($p_r \to 0$), the mass of null reviews becomes more similar in the two states, since $1 - q(s)p_r$ approaches 1 uniformly (drowning out any asymmetry due to q). Hence, null reviews become less informative as p_r decreases. In particular, in cases when few people submit reviews (i.e., when p_r is small), the intuition from uniform reporting rates applies in the same way to the setting with non-uniform reporting rates since null reviews are not informative.

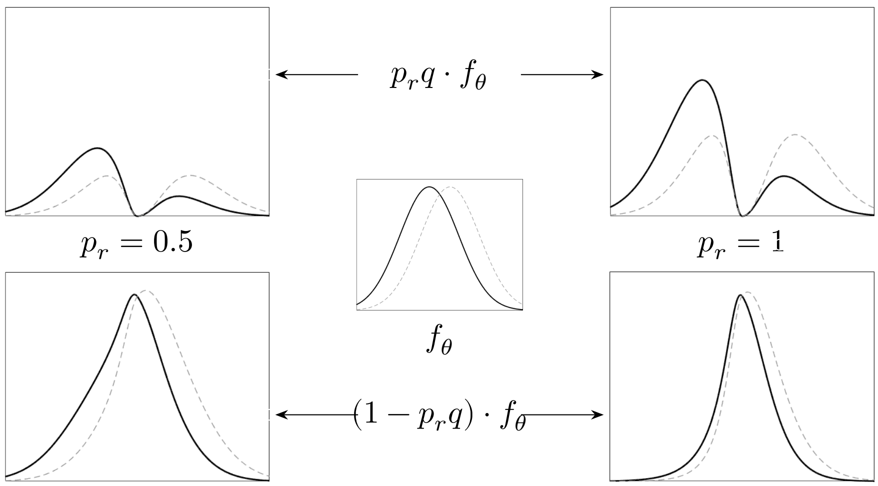


Figure 6: Distributions of reviews for $f_\theta(s)$ normal, with $\mu = 0.25$, and $q(s) = 1 - \frac{1}{1+2(x^2+4\mathbb{1}\{x\leq 0\}x^2)}$ and for $p_r = 0.5$ (left), 1 (right). The top panels show the distribution of submitted reviews, and the bottom panels show the distribution of null reviews.

5.2 Relationship between Production and Consumption of Information

As discussed above, the rates of reporting p (whether uniform or not) are taken from an ex-ante, and not an ex-post, perspective. An ex-post perspective would be to directly compare n_r reviews of type r and $n_{r'}$ reviews of type r'.[16] In practice, this latter decision falls to consumers who must decide between multiple types of reviews on a website. While the ex-ante and ex-post preferences over sources of information are related, they are not identical.

Consider two different review systems, r and r', with $\nu(r) \gg \nu(r')$. If $p_r \ll p_{r'}$, then the platform may be worried about receiving any reviews of type r, and so might prefer r'. However, on average, there will still be many reviews of type r

[16] This is the perspective that has traditionally been taken in the literature. For instance, see Chernoff (1952), Moscarini and Smith (2002), and Mu et al. (2021).

submitted, and so to a consumer deciding between Np_r reviews from system r and $Np_{r'}$ reviews from system r', the system with "better" reviews are preferred. This intuition is correct: the ex-post perspective leads to a stronger preference for finer reviews. However, when information is scarce, this difference disappears and the two orders agree. Appendix C formalizes the model and the differences, and shows these relationships.

6 Conclusion

In this paper I quantify, in a rate of learning sense, how much information is lost when signals are coarsened. I show how a platform should trade-off between the quality of a review and how frequently it is reported. I also show how to choose optimal review systems when information is scarce. I apply the framework to the context of online reviews and find that the degree of taste homogeneity that the product being reviewed induces in the population has important implications for how much information is lost when signals are coarsened and for the structure of optimal binary systems. As the population becomes more homogeneous, more information is lost. A symmetric binary system can lose an arbitrary amount of information relative to the full signal, but this loss can be greatly mitigated by optimally using an asymmetric system. The same qualitative results extend to other review systems with finite bins. These results explain why certain review systems are used in practice—in particular explaining the prevalence of positively-skewed systems and binary schemes—and highlight what forces platforms should consider when designing review systems.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022): "Learning from reviews: The selection effect and the speed of learning". *Econometrica* 90(6), pp. 2857–2899.

Besbes, O. and M. Scarsini (2018): "On information distortions in online ratings". *Operations Research* 66(3), pp. 597–610.

Blackwell, D. (1951): "Comparison of Experiments". *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 93–102.

Botelho, T. L., S. Jun, D. Humes, and K. A. DeCelles (2025): "Scale dichotomization reduces customer racial discrimination and income inequality". *Nature*, pp. 1–9.

Che, Y.-K. and J. Hörner (2018): "Recommender systems as mechanisms for social learning". *The Quarterly Journal of Economics* 133(2), pp. 871–925.

Chen, P. Y., J. Yoon, and S.-y. Wu (2004): "The impact of online recommendations and consumer feedback on sales". *International Conference on Information Systems, ICIS 2004*.

Chernoff, H. (1952): "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations". *The Annals of Mathematical Statistics*, pp. 493–507.

Dasaratha, K. and K. He (2024): *Aggregative Efficiency of Bayesian Learning in Networks*. arXiv: 1911.10116 [econ.TH]. URL: https://arxiv.org/abs/1911.10116.

Fedorov, V., F. Mannino, and R. Zhang (2009): "Consequences of dichotomization". *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry* 8(1), pp. 50–61.

Fradkin, A., E. Grewal, and D. Holtz (2021): "Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on Airbnb". *Marketing Science* 40(6), pp. 1013–1029.

Fradkin, A. and D. Holtz (2023): "Do incentives to review help the market? Evidence from a field experiment on Airbnb". *Marketing Science* 42(5), pp. 853–865.

Frick, M., R. Iijima, and Y. Ishii (2023): "Learning Efficiency of Multiagent Information Structures". *Journal of Political Economy* 131(12), pp. 3377–3414.

— (2024a): "Monitoring with rich data". *arXiv preprint arXiv:2312.16789*.

— (2024b): "Welfare comparisons for biased learning". *American Economic Review* 114(6), pp. 1612–1649.

Garg, N. and R. Johari (2019): "Designing optimal binary rating systems". *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1930–1939.

Hann-Caruthers, W., V. V. Martynov, and O. Tamuz (2018): "The speed of sequential asymptotic learning". *Journal of Economic Theory* 173, pp. 383–409.

Harel, M., E. Mossel, P. Strack, and O. Tamuz (2021): "Rational groupthink". *The Quarterly Journal of Economics* 136(1), pp. 621–668.

Hu, N., P. A. Pavlou, and J. Zhang (2017): "On self-selection biases in online product reviews". *MIS quarterly* 41(2), pp. 449–475.

Hu, N., J. Zhang, and P. A. Pavlou (2009): "Overcoming the J-shaped distribution of product reviews". *Communications of the ACM* 52(10), pp. 144–147.

Ifrach, B., C. Maglaras, M. Scarsini, and A. Zseleva (2019): "Bayesian social learning from consumer reviews". *Operations Research* 67(5), pp. 1209–1221.

Lafky, J. (2014): "Why do people rate? Theory and evidence on online ratings". *Games and Economic Behavior* 87, pp. 554–570.

Magnani, M. (2020): "The economic and behavioral consequences of online user reviews". *Journal of Economic Surveys* 34(2), pp. 263–292.

Moscarini, G. and L. Smith (2002): "The law of large demand for information". *Econometrica* 70(6), pp. 2351–2366.

Mu, X., L. Pomatto, P. Strack, and O. Tamuz (2021): "From Blackwell dominance in large samples to Rényi divergences and back again". *Econometrica* 89(1), pp. 475–506.

Nelson, P. (1970): "Information and consumer behavior". *Journal of Political Economy* 78(2), pp. 311–329.

Oprea, R. (2020): "What makes a rule complex?" *American economic review* 110(12), pp. 3913–3951.

Puri, I. (2022): "Simplicity and risk". *The Journal of Finance*.

Rosenberg, D. and N. Vieille (2019): "On the efficiency of social learning". *Econometrica* 87(6), pp. 2141–2168.

Torgersen, E. N. (1981): "Measures of information based on comparison with total information and with total ignorance". *The Annals of Statistics* 9(3), pp. 638–657.

Wang, Y. and K. Li (2025): "Not easy to "like": How does cognitive load influence user engagement in online reviews?" *Decision Support Systems*, p. 114436. ISSN: 0167-9236. DOI: https://doi.org/10.1016/j.dss.2025.114436.

A Proofs

All proofs are contained in this appendix, broken out by section.

A.1 Proofs for Section 3

In this appendix I provide the proof of the major result from Section 3. Before I prove the main result, I first state the following result from Torgersen (1981):

Theorem (Torgersen 1981). *Suppose that reviews are always submitted (i.e., $p_r = 1$ for all review systems r). Fix two review systems r, r' with $\nu(r) > \nu(r')$ and a decision problem for the platform. There exists an $\overline{N}$ such that for all $N \geq \overline{N}$, such that $u^*(r,N) > u^*(r',N)$.*

Proof of Theorem 1. For ease of notation, write $\nu_r := \nu(r)$ to denote the learning efficiency of a review r. The pair (p_r, r) can be thought of as defining a statistical experiment over the two states of the world.

In order to apply the above theorem to this setting, we must translate (p_r, r) into a statistical experiment without stochastic reporting, and show that its learning efficiency is given by $p_r \nu_r$, where ν_r is the learning efficiency of the statistical experiment $(1, r)$.

Let $\gamma_L^{(p_r,r)}$ be the measure defined over $R_r \cup \{\emptyset\}$ representing the distribution of reviews in state L. That is,

$$\begin{aligned}\gamma_L^{(p_r,r)}(B) :=& p_r \cdot \int_{\{s:r(s)\in B\}} f_L(s)ds + (1-p_r)\cdot \mathbb{1}\{\emptyset \in B\}\\ =& p_r \cdot \gamma_L^r(B\cap R_r) + (1-p_r)\cdot \mathbb{1}\{\emptyset\in B\}.\end{aligned}$$

The only difference between this measure and the measure γ_L^r is the weight $1-p_r$ placed on $\{\emptyset\}$, which represents a null review (a reviewer choosing to not submit a review). Similarly,

$$\gamma_H^{(p_r,r)}(B) := p_r \cdot \gamma_H^r(B\cap R_r) + (1-p_r)\cdot \mathbb{1}\{\emptyset\in B\}.$$

Notice then, for a given value of λ, it is the case that

$$\begin{aligned}\int_{R_r\cup\{\emptyset\}} \left(d\gamma_L^{(p_r,r)}\right)^{\lambda}\left(d\gamma_H^{(p_r,r)}\right)^{1-\lambda} =& \int_{R_r} \left(p_r\cdot d\gamma_L^r\right)^{\lambda}\left(p_r \cdot d\gamma_H^r\right)^{1-\lambda} + (1-p_r)\\ =& p_r\int_{R_r}\left(d\gamma_L^r\right)^{\lambda}\left(d\gamma_H^r\right)^{1-\lambda} + 1 - p_r.\end{aligned}$$

Hence,

$$\min_{\lambda\in[0,1]}\int_{R_r\cup\{\emptyset\}} \left(d\gamma_L^{(p_r,r)}\right)^{\lambda}\left(d\gamma_H^{(p_r,r)}\right)^{1-\lambda} = 1 - p_r\left(1 - \min_{\lambda\in[0,1]}\int_{R_r}\left(d\gamma_L^r\right)^{\lambda}\left(d\gamma_H^r\right)^{1-\lambda}\right).$$

Concluding,

$$\nu_{(p_r,r)} = 1 - \min_{\lambda\in[0,1]} \int_{R_r\cup\{\emptyset\}} \left(d\gamma_L^{(p_r,r)}\right)^{\lambda} \left(d\gamma_H^{(p_r,r)}\right)^{1-\lambda} = p_r \cdot \nu_r,$$

which is what we set out to show. □

Proof of Lemma 1. This follows from the Monotone Likelihood Ratio Property of log-concave densities and Proposition 5, which is stated and proved in Appendix **??**. □

Proof of Theorem 2. The proof follows from an application to Theorem 3, in the case that $q \equiv 1$ and the review system r is a threshold system. To see this, observe that (with the notation of the proof of Theorem 3), if $\tilde{r}$ consists of those signals that fall between τ_{i-1} and τ_i, it is the case that

$$H_+'(0,\tilde{r})^2 = \left(\int_{\tau_{i-1}}^{\tau_i} p_r f'(s)ds\right)^2 = p_r^2 \left(f(\tau_i) - f(\tau_{i-1})\right)^2 \quad \text{and}$$
$$H_+(0,\tilde{r}) = \left(\int_{\tau_{i-1}}^{\tau_i} p_r f(s)ds\right) = p_r \left(F(\tau_i) - F(\tau_{i-1})\right).$$

The result follows from simply taking the ratio. □

A.2 Proofs for Section 4

In this Appendix I provide the proofs of the results in Section 4. To begin the analysis, some preliminary calculations are needed. Closely related to our family interest is the Gamma function.

Definition 4. Denote by $\Gamma(\alpha)$ the standard Gamma function, given by

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1}e^{-x}dx.$$

Let $\Gamma(\alpha,t)$ denote the upper incomplete gamma function:

$$\Gamma(\alpha,t) := \int_t^\infty x^{\alpha-1}e^{-x}dx.$$

Finally, let $\gamma(\alpha, t)$ denote the lower incomplete gamma function

$$\gamma(\alpha, t) := \int_0^t x^{\alpha-1} e^{-x} dx = \Gamma(\alpha) - \Gamma(\alpha, t).$$

Lemma 4. *It is the case that*

$$f_\alpha(x) = \frac{\alpha}{2\Gamma(1/\alpha)} e^{-|x|^\alpha}.$$

Moreover, the Fisher information of f_α is given by

$$I_{f_\alpha}(0) = \frac{\alpha^2}{\Gamma(1/\alpha)} \Gamma\left(2 - \frac{1}{\alpha}\right).$$

Proof of Lemma 4. The first part of the statement is just ensuring that f_α is indeed a probability distribution.

$$\int_{-\infty}^{\infty} e^{-|x|^\alpha} dx = 2 \int_0^{\infty} e^{-x^\alpha} dx.$$

Make now the substitution $u = x^\alpha$. Then, $\frac{1}{\alpha} u^{\frac{1}{\alpha}-1} du = dx$ so that

$$2 \int_0^{\infty} e^{-x^\alpha} dx = 2 \int_0^{\infty} \frac{1}{\alpha} u^{\frac{1}{\alpha}-1} e^{-u} du = \frac{2}{\alpha} \Gamma\left(\frac{1}{\alpha}\right).$$

This proves the first claim. To show the second claim, recall that the Fisher Information is also equal to

$$I_f = \mathbb{E}\left[\left(\frac{f'(s)}{f(s)}\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial s^2} \log(f(s))\right].$$

This means that it is the case that

$$\begin{aligned} I_{f_\alpha} =& \frac{\alpha}{2\Gamma(1/\alpha)} \int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial x^2} |x|^\alpha\right) e^{-|x|^\alpha} dx = \frac{\alpha}{\Gamma(1/\alpha)} \int_0^{\infty} \left(\frac{\partial^2}{\partial x^2} x^\alpha\right) e^{-x^\alpha} dx \\ =& \frac{\alpha}{\Gamma(1/\alpha)} \int_0^{\infty} \alpha(\alpha-1) x^{\alpha-2} e^{-x^\alpha} dx. \end{aligned}$$

Now, perform again the substitution of $u = x^\alpha$.

$$
\begin{aligned}
=&\frac{\alpha}{\Gamma(1/\alpha)}\int_0^\infty \alpha(\alpha-1)\frac{1}{\alpha}u\cdot u^{-\frac{2}{\alpha}}\cdot u^{\frac{1}{\alpha}-1}e^{-u}dx\\
=&\frac{\alpha}{\Gamma(1/\alpha)}\int_0^\infty (\alpha-1)u^{-\frac{1}{\alpha}}e^{-u}dx = \frac{\alpha(\alpha-1)}{\Gamma(1/\alpha)}\Gamma\left(1-\frac{1}{\alpha}\right).
\end{aligned}
$$

Since it is the case that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$,

$$
\Gamma\left(1-\frac{1}{\alpha}\right) = \frac{\alpha}{\alpha-1}\Gamma\left(2-\frac{1}{\alpha}\right),
$$

Concluding,

$$
I_{f_\alpha}(0) = \frac{\alpha^2}{\Gamma(1/\alpha)}\Gamma\left(2-\frac{1}{\alpha}\right).
$$

□

In all proofs, denote $\kappa_\alpha(\tau)$ by $\kappa(\alpha;\tau)$. This is just done for clarity: κ is a function of α.

Proof of Lemma 2. First, a simple computation shows that $\kappa(\alpha;0) = \Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(2-\frac{1}{\alpha}\right)$. This shows that $\kappa(1;0) = 1$, as $\Gamma(1) = 1$.

I now proceed to the other claims. Let $\psi(\alpha)$ denote the digamma function

$$
\psi(\alpha) := \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}.
$$

It is well-known that ψ is strictly increasing on $(0,\infty)$.

Now, let $\kappa'(\alpha;0)$ denote $\frac{\partial}{\partial\alpha}\kappa(\alpha;0)$. A straightforward computation yields that

$$
\kappa'(\alpha;0) = \frac{\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(2-\frac{1}{\alpha}\right)}{\alpha^2}\left(\psi\left(2-\frac{1}{\alpha}\right)-\psi\left(\frac{1}{\alpha}\right)\right).
$$

Since ψ is strictly increasing, it is immediately seen that $\kappa'(\alpha;0) > 0$.

It is left to show that $\lim_{\alpha\to\infty}\kappa(\alpha;0) = \infty$. First, it can be easily verified that $\psi(2) < 1$ and $\lim_{\alpha\to\infty}\frac{1}{\alpha}\psi\left(\frac{1}{\alpha}\right) = -1$. These two facts together show that κ' can be seen as

$$\frac{\kappa'(\alpha;0)}{\kappa(\alpha;0)} \approx \frac{1}{\alpha}.$$

That is, $\lim_{\alpha\to\infty} \alpha\frac{\kappa'(\alpha;0)}{\kappa(\alpha;0)} = 1$. This means that $\frac{\kappa(\alpha;0)}{\alpha} \to 1$, and hence $\kappa'(\alpha;0) \to 1$. This shows that κ diverges, completing the proof. □

Proof of Proposition 1. As noted, I restrict attention to $\tau \in [0,\infty)$ for the proof. The exact analogues apply for $\tau \leq 0$.

For ease of reference I restate that the objective that the platform aims to maximize is given by.

$$g_\alpha(\tau) := \frac{f_\alpha(\tau)^2}{F_\alpha(\tau)(1-F_\alpha(\tau))} \tag{7}$$

I proceed in several steps. I first show that $\tau^*(\alpha) \neq 0$ for $\alpha > 2$, and that it 0 is a candidate for $\tau^*(\alpha)$ for $\alpha \leq 2$.

Lemma 5. *It is the case that* 0 *is a local maximum of* $g_\alpha(\tau)$ *for* $\alpha \leq 2$*, and a local minimum of* $g_\alpha(\tau)$ *for* $\alpha > 2$*.*

Proof of Lemma 5. The derivative of (7) with respect to τ is given by (dropping the subscript α)

$$\frac{f(\tau)}{F(\tau)(1-F(\tau))}\left[2f'(\tau) - f(\tau)^2\left(\frac{1}{F(\tau)} - \frac{1}{1-F(\tau)}\right)\right].$$

As f is of full support, $\frac{f(\tau)}{F(\tau)(1-F(\tau))} > 0$, and so can be ignored when searching for a critical value, and hence a maximum. The first order condition to be a local maximum then becomes

$$2f'(\tau) = f(\tau)^2\left(\frac{1}{F(\tau)} - \frac{1}{1-F(\tau)}\right). \tag{8}$$

For any $\alpha > 1$, 0 will be a solution to (8), and hence a candidate for a maximizer. This is as $f'_\alpha(0) = 0$ and $F_\alpha(0) = \frac{1}{2}$.

This means that to verify the claim the second order condition is needed. At 0, it can be verified that the second-order condition for a maximizer (since $F_\alpha(0) = \frac{1}{2}$) is

$$\frac{1}{2} f_\alpha(0)^4 + \frac{1}{8} f'_\alpha(0)^2 + \frac{1}{8} f_\alpha(0) f''_\alpha(0) \leq 0.$$

Notice, for all $\alpha > 2$, we have that $f'_\alpha(0) = 0$ and $f''_\alpha(0) = 0$. This means that, as $f_\alpha(0) > 0$, for all $\alpha > 2$, $\tau = 0$ is a local minimum, and hence is not optimal. Similarly, for $\alpha < 2$, it is the case that $\lim_{\tau \to 0} f''_\alpha(\tau) = -\infty$, while $f'_\alpha(0), f_\alpha(0) > 0$ This means that 0 is indeed a local maximum for $\alpha < 2$. The case $\alpha = 2$ can be computed numerically, and it can be seen that this is indeed a local maximum in this case. □

I now show that there is at most one possible candidate for the maximizer of $g_\alpha(\tau)$ on $\tau \in [0, \infty)$. This will show uniqueness of $\tau^*(\alpha)$.

Lemma 6. *The following are true:*

(i) For $\alpha > 2$, there is exactly one value of $\tau \in (0, \infty)$ that satisfies (8) and it is a local maximum of $g_\alpha(\tau)$; and

(ii) For $\alpha \leq 2$, no value of $\tau \in (0, \infty)$ satisfies (8).

Proof of Lemma 6. Throughout I drop the subscript of α for notational clarity.

Case 1: $\alpha > 2$.
Straightforward algebra allows (8) to be rewritten as

$$\frac{2f'(\tau)}{1 - 2F(\tau)} = \frac{f^2(\tau)}{F(\tau)(1 - F(\tau))}. \tag{9}$$

Notice that the right-hand side of (9) is exactly $g(\tau)$. This means that the value of g can also be used to determine the sign of its derivative. Specifically, the critical points of the objective are exactly those that intersect the curve given by

$$w(\tau) := \frac{2f'(\tau)}{1 - 2F(\tau)}.$$

It is exactly when $g(\tau) = w(\tau)$ that $g'(\tau) = 0$. Moreover,

$$g'(\tau) > 0 \iff g(\tau) > w(\tau).$$

Due to this relationship, the characteristics of $w(\tau)$ imply certain characteristics of $g(\tau)$. In particular, I now show that $w(\tau)$ has one unique maximum on $(0, \infty)$. To do this, I iterate on the process above. The sign of $w'(\tau) = 0$ is determined by the sign of

$$2f''(\tau)(1 - 2F(\tau)) + 4f'(\tau)f(\tau) \tag{10}$$

in particular, $w'(\tau) = 0$ if and only if

$$\frac{2f'(\tau)}{1 - 2F(\tau)} = -\frac{f''(\tau)}{f(\tau)} =: h(\tau).$$

Moreover, $w'(\tau) > 0$ if and only if $h(\tau) > w(\tau)$ (notice that this is the reverse relationship of w and g).

While w and g are difficult to manage because the presence of F in their definition (and there is no closed-form expression for the incomplete Gamma functions), $h(\tau) = (\alpha^2 - \alpha)x^{\alpha-2} - \alpha^2 x^{2\alpha-2}$. This (relative to g and w) is an extremely simple function. It can be seen that for all $\alpha > 2$, h has the following properties: (i) $h(0) = 0$, and $h(\tau) > 0$ for τ in a neighbourhood of 0; (ii) h is single-peaked; (iii) $h(1) < 0$; and (iv) $h(\tau) > w(\tau)$ in a neighbourhood of 0. Properties (i)-(iii) are easily to verify, and property (iv) follows from $w'(\tau) > 0$ in a neighbourhood of 0 (which directly implies that $w(\tau) < h(\tau)$).

Claim 1a: $w(\tau)$ has at one critical value on $(0, \infty)$. It is a local maximum, and this critical value is on $(0, 1)$.

Recall that if $w(\tau) < h(\tau)$, then $w'(\tau) > 0$, and if $w(\tau) > h(\tau)$, then $w'(\tau) < 0$. First, observe that all critical values of w must be on $(0, 1)$ because h is negative of $[1, \infty)$ (so $w(\tau) \neq h(\tau)$ for any $\tau \geq 1$).

Second, let τ_h be the value of τ that maximizes h. Notice that h is increasing on $(0, \tau_h)$ and decreasing on (τ_h, ∞). Suppose now that $w(\tau') = h(\tau')$ for some $\tau' \in (0, \tau_h)$. There must be smallest value of τ' for which this holds, as $w(\tau)$ is strictly increasing in a neighbourhood of 0. Then for this τ', it is the case that $w(\tau') = 0$ and $h(\tau') > 0$. This means that for some $\epsilon > 0$ it is the case that for all $\tau \in (\tau' - \epsilon, \tau')$ it is the case that $w(\tau) > h(\tau)$. But τ' is the smallest positive value where the two cross, which is a contradiction. This means that it cannot be the case that $w'(\tau) = 0$ when $h'(\tau) > 0$. This is as, for w to cross h when h is increasing it must be the case that w is larger than h before the crossing.

Since eventually $w(\tau) > h(\tau)$, it must be the case then that either $w = h$ at the peak of h, or on the region where h is decreasing. In either case it must be the case that w goes from increasing to decreasing, and can not cross h again. This proves the claim.

Claim 1b: $g(\tau)$ has a most one critical value on $(0, \infty)$. It is a local maximum, and this critical value is on $(0, 1)$.

Recall that, if $g(\tau) > w(\tau)$, then $g'(\tau) > 0$, and if $g(\tau) < w(\tau)$, then $g'(\tau) < 0$. Notice that because zero is a local minimum of $g(\tau)$ (by Lemma 5), $g(\tau)$ is increasing in a neighbourhood of zero (since we are concerned only with $\tau \geq 0$. Moreover, it is the case that $g(0) > 0 = w(0)$. This second equality holds because $f''(0) = 0$ for $\alpha > 2$. Denote by τ_w the value of τ that maximizes w on $(0, \infty)$.

Now, suppose that $g(\tau') = w(\tau')$ for some $\tau' \in (\tau_w, \infty)$. This is because it implies that $g(\tau) > w(\tau)$ for $\tau = (\tau', \tau' + \epsilon)$ for some $\epsilon > 0$. However, this means that $g'(\tau) > 0$ on this region. Because w is decreasing on this region, it implies that $g'(\tau) > 0$ on (τ', ∞), contradicting that $\lim_{\tau \to \infty} g(\tau) = 0$. This means that the two cannot intersect when $w'(\tau) < 0$.

By a similar logic, it must be the case that $g(\tau) = w(\tau)$ for some $\tau \in (0, \tau_w]$. Call the first point where they intersect τ^* (well-defined because they are not equal at 0). Notice that for all $\tau \in (\tau^*, \tau^* + \epsilon)$, it must be the case that g is decreasing.[17] But, g must continue to decrease until they intersect again. So, they cannot intersect until w decreases. We have, however, ruled out intersection in that case, so it must be that τ^* is unique, proving the claim and finishing Case 1.

Case 2: $\alpha < 2$.

In the case of $\alpha < 2$ showing that there is no critical value on $(0, \infty)$ is simpler and

[17] This is obvious in the case that $\tau^* < \tau_w$. In the case that $\tau^* = \tau_w$, it follows from a similar argument to showing that there is no intersection when w is decreasing.

so I omit the details.

In this case, (i) $\lim_{\tau \to 0} w(\tau) = \infty$ and (ii) $h(\tau)$ is monotonically decreasing. These jointly imply that $w(\tau)$ is monotonically decreasing to 0. This in turn implies that $g(\tau)$ is monotonically decreasing to zero (since they cannot intersect when w decreases in a similar argument to above). This shows that in the case that $\alpha < 2$ it must be that $\tau = 0$ is the only possible critical value.

Case 3: $\alpha = 2$.
This is dealt with in a similar manner to Case 2. The major difference is that $w(\tau)$ does not diverge at 0. Instead, it is easy to show that $w(0) = h(0)$ and w is decreasing in a region of 0, so that it must be monotonically decreasing. This case then reduces to Case 2. □

Now that I have showed that τ^* exists and is unique, it is sufficient to look at the solution to the first order condition (8) on $(0, \infty)$. The last things to check are that $\tau^*(\alpha) \to 1$, and the bounding value of κ.

Claim: $\lim_{\alpha \to \infty} \tau^*(\alpha) = 1$.

First, notice that $\tau^*(\alpha) < 1$ for all $\alpha \geq 1$, by Lemma 6.

So, I simply show that for all $\tau < 1$, there exists an α_τ such that: $\alpha \geq \alpha_\tau$, $g'_\alpha(\tau) > 0$. To do this, explicitly write the inner term

$$
\begin{aligned}
2f'(\tau) - f(\tau)^2 \left(\frac{1}{F(\tau)} - \frac{1}{1 - F(\tau)} \right) \\
&= -\alpha\tau^{\alpha-1} \cdot \frac{\alpha}{\Gamma(1/\alpha)} e^{-\tau^\alpha} \\
&\quad + \left(\frac{\alpha}{2\Gamma(1/\alpha)} e^{-\tau^\alpha} \right)^2 \left[\frac{\Gamma(1/\alpha)}{\frac{1}{2}\Gamma(1/\alpha, \tau^\alpha)} - \frac{\Gamma(1/\alpha)}{\frac{1}{2}\Gamma(1/\alpha) + \frac{1}{2}\gamma(1/\alpha, \tau^\alpha)} \right],
\end{aligned}
$$

where here $\Gamma(1/\alpha, \cdot)$ and $\gamma(1/\alpha, \cdot)$ are the upper- and lower-incomplete gamma functions, respectively. Because we are only interested in the sign of this expression, it is possible to simplify. It is equivalent to look at the sign of

$$
-2\tau^{\alpha-1} + e^{-\tau^\alpha} \left[\frac{1}{\Gamma(1/\alpha, \tau^\alpha)} - \frac{1}{\Gamma(1/\alpha) + \gamma(1/\alpha, \tau^\alpha)} \right]. \tag{11}
$$

That is, $g'_\alpha(\tau) > 0$ exactly when (11) is positive. When $\tau < 1$, the first term converges to 0, so we need only ensure that the second term stays bounded from 0. First, for fixed $\tau < 1$, $e^{-\tau^\alpha} \to 1$. Second, as $\alpha \to \infty$, it is easy to verify that $f_\alpha(\tau) \to \frac{1}{2}\mathbb{1}\{\tau \in [-1,1]\}$ (except for the measure zero set of $\{-1,1\}$). This means that

$$\lim_{\alpha\to\infty}\left(\frac{1}{F(\tau)} - \frac{1}{1-F(\tau)}\right) = \left(\frac{2}{\tau+1} - \frac{2}{1-\tau}\right) > 0$$

for all fixed $\tau \in (0,1)$. This means that neither portions of the second term go to zero, so that for any τ in this interval, $g'_\alpha(\tau)$ is eventually positive.

This means that $\tau^*(\alpha)$ must eventually grow to 1. It is not hard to use the same arguments to show that $\tau^*(\alpha)^\alpha$ converges to the solution of $\Gamma(0,x) - \frac{1}{2x}e^{-x} = 0$.

Claim: $\kappa(\alpha;\tau^*(\alpha)) < 2e^2\int_1^\infty \frac{e^{-x}}{x}dx$.

For all $\alpha \geq 4$, it can be seen that $\kappa(\alpha,1)$ is less than this limit. For $\alpha < 4$, using the naive threshold of 0 will also return a value that is less than this limit. This shows the final claim, and completes the proof of Proposition 1. □

Proof of Proposition 2. In order to prove Proposition 2(i), I prove the following, stronger, lemma.

Lemma 7. *$\tau^*(\alpha)^\alpha$ is strictly increasing in α for $\alpha \geq 2$.*

Proof of Lemma 7. First, notice that

$$\frac{\partial \tau^*(\alpha)^\alpha}{\partial\alpha} > 0 \iff \left(\frac{\partial^2}{\partial\tau^\alpha\partial\alpha}\frac{f_\alpha^2(\tau)}{F_\alpha(\tau)(1-F_\alpha(\tau))}\right)\Big|_{\tau=\tau^*(\alpha)} > 0.$$

That is, letting $x = \tau^\alpha$, the sign of

$$\frac{\partial}{\partial\alpha}\left[-2\frac{x}{x^{1/\alpha}}e^x + \left[\frac{1}{\Gamma(1/\alpha,x)} - \frac{1}{\Gamma(1/\alpha)+\gamma(1/\alpha,x)}\right]\right] \tag{12}$$

must be determined at the optimal value of $x^*(\alpha) = \tau^*(\alpha)^\alpha$. Notice that (12) is not exactly equal to the derivative, because there is an additional positive term that multiplies this. However, because what is in the brackets here is zero at the optimal threshold, this term can be ignored (because it will not influence sign of this object). This just follows from the envelope theorem. So, we must show that (12) is positive at $x^*(\alpha)$.

Another simplification is to replace α with $1/\alpha$. That is, let $\beta := 1/\alpha$. A straightforward computation that yields that the sign of (12) is the opposite of the sign of

$$\begin{aligned} b_\beta(x) :=& 2\log(x)\frac{x}{x^\beta}e^x \\ &- \frac{\int_x^\infty y^{\beta-1}\log(y)e^{-y}dy}{\Gamma(\beta,x)^2} + \frac{\int_0^\infty y^{\beta-1}\log(y)e^{-y}dy + \int_0^x y^{\beta-1}\log(y)e^{-y}dy}{(\Gamma(\beta)+\gamma(\beta,x))^2}. \end{aligned}$$

This is because $\frac{\partial}{\partial\alpha}\beta < 0$. So, I show that $b_\beta(x^*(\beta)) < 0$ (abusing notation here writing $x^*(\beta))$ for $\beta < \frac{1}{2}$. As a final choice to make notation easier, let λ_β denote the measure that is induced by $y^{\beta-1}e^{-y}$ (that is, $d\lambda_\beta(y) = y^{\beta-1}e^{-y}dy$). Then, $b_\beta(x)$ can be written as

$$2\log(x)\frac{x}{x^\beta}e^x - \frac{\int_x^\infty \log(y)d\lambda_\beta(y)}{\Gamma(\beta,x)^2} + \frac{\int_0^\infty \log(y)d\lambda_\beta(y) + \int_0^x \log(y)d\lambda_\beta(y)}{(\Gamma(\beta)+\gamma(\beta,x))^2}. \tag{13}$$

Note that in the notation above, we have that

$$\Gamma(\beta) = \int_0^\infty d\lambda_\beta(y), \quad \Gamma(\beta,x) = \int_x^\infty d\lambda_\beta(y), \quad \text{and} \quad \gamma(\beta) = \int_0^x d\lambda_\beta(y).$$

Now, at $x^*(\beta)$, from (12), it is the case that

$$2x^{1-\beta}e^x = \frac{1}{\Gamma(\beta,x)} - \frac{1}{\Gamma(\beta)+\gamma(\beta,x)}.$$

This means that it is possible to write (13) at $x^*(\beta)$ (I write x instead of $x^*(\beta)$ for conciseness) as

$$\begin{aligned} &\log(x)\left(\frac{1}{\Gamma(\beta,x)} - \frac{1}{\Gamma(\beta)+\gamma(\beta,x)}\right) \\ &\quad - \frac{\int_x^\infty \log(y)d\lambda_\beta(y)}{\Gamma(\beta,x)^2} + \frac{\int_0^\infty \log(y)d\lambda_\beta(y) + \int_0^x \log(y)d\lambda_\beta(y)}{(\Gamma(\beta)+\gamma(\beta,x))^2}. \\ =&\frac{\log(x)\int_x^\infty d\lambda_\beta(y)}{\Gamma(\beta,x)^2} - \frac{\log(x)\left(\int_0^\infty d\lambda_\beta(y) + \int_0^x d\lambda_\beta(y)\right)}{\Gamma(\beta)+\gamma(\beta,x)} \\ &\quad - \frac{\int_x^\infty \log(y)d\lambda_\beta(y)}{\Gamma(\beta,x)^2} + \frac{\int_0^\infty \log(y)d\lambda_\beta(y) + \int_0^x \log(y)d\lambda_\beta(y)}{(\Gamma(\beta)+\gamma(\beta,x))^2} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\int_x^\infty (\log(x) - \log(y))\, d\lambda_\beta(y)}{\Gamma(\beta, x)^2} \\
&\quad + \frac{\int_0^\infty (\log(y) - \log(x))\, d\lambda_\beta(y) + \int_0^x (\log(y) - \log(x))\, d\lambda_\beta(y)}{(\Gamma(\beta) + \gamma(\beta, x))^2}.
\end{aligned}
$$

Since log is an increasing function, it must be that $\int_x^\infty (\log(x) - \log(y))\, d\lambda_\beta(y) < 0$ whenever $\beta < \frac{1}{2}$, so that $x^*(\beta) > 0$. Moreover, it must always be the case that $\Gamma(\beta, x) < \Gamma(\beta) + \gamma(\beta, x)$ whenever $x > 0$. This means that at $x^*(\beta)$ it is the case that

$$
\frac{\int_x^\infty (\log(x) - \log(y))\, d\lambda_\beta(y)}{\Gamma(\beta, x)^2} < \frac{\int_x^\infty (\log(x) - \log(y))\, d\lambda_\beta(y)}{\left(\Gamma(\beta) + \gamma(\beta, x)\right)^2}.
$$

This means that (13) evaluated at $x^*(\beta)$ is less than

$$
\begin{aligned}
&\frac{\int_x^\infty (\log(x) - \log(y))\, d\lambda_\beta(y)}{(\Gamma(\beta) + \gamma(\beta, x))^2} \\
&\quad + \frac{\int_0^\infty (\log(y) - \log(x))\, d\lambda_\beta(y) + \int_0^x (\log(y) - \log(x))\, d\lambda_\beta(y)}{(\Gamma(\beta) + \gamma(\beta, x))^2} \\
&= \frac{2\int_0^x (\log(y) - \log(x))\, d\lambda_\beta(y)}{(\Gamma(\beta) + \gamma(\beta, x))^2} < 0.
\end{aligned}
$$

This proves that $x^*(\beta)$ is decreasing in β, so that $\tau^*(\alpha)^\alpha$ is increasing in α for all $\beta < \frac{1}{2}$. This in turn shows that $\tau^*(\alpha)^\alpha$ is increasing in α for all $\alpha \geq 2$. □

As noted, Proposition 2(i) follows as an immediate corollary of this lemma, since $\tau^*(\alpha) < 1$ for all α.

Proof of Proposition 2(ii):

In what follows, let $P(\alpha, t) := \gamma(\alpha, t)/\Gamma(\alpha)$ denote the *normalized lower incomplete gamma function.*

First, recall that $\tau^*(\alpha)^\alpha$ strictly increasing in α by Lemma 7.

I show the claim separately on two intervals, based off of the behaviour of $f_\alpha(0)$ as a function of α. In what follows, let α^* be defined as the unique value for which $\psi(1/\alpha) + \alpha = 0$. For $\alpha < \alpha^*$, it is the case that $\Gamma(1/\alpha)/\alpha$ is decreasing in α (so that $f_\alpha(0)$ is increasing), and for $\alpha > \alpha^*$ it is the case the $\Gamma(1/\alpha)/\alpha$ is increasing in α (so that $f_\alpha(0)$ is decreasing).

Case 1: $\alpha < \alpha^*$.
First, on this region, for a fixed τ with $0 < \tau < 1$, it must be the case that $f_\alpha(\tau)$ is increasing in α. This is because clearly τ^α is decreasing in α, and this is the region where $\frac{\alpha}{\Gamma(\alpha)}$ is increasing.

Now, $F_\alpha(0) \equiv \frac{1}{2}$, and for all $0 < \tau < \tau'$, it is the case that $f_\alpha(\tau)$ is increasing in α. This means that, for all $\alpha, \alpha' \in (2, \alpha^*)$ with $\alpha < \alpha'$, it is the case that

$$F_\alpha(\tau^*(\alpha)) < F_{\alpha'}(\tau^*(\alpha))$$

From Lemma 7 it is the case that $\tau^*(\alpha') > \tau^*(\alpha)$, so that

$$F_{\alpha'}(\tau^*(\alpha)) < F_{\alpha'}(\tau^*(\alpha'))$$

This shows that $F_\alpha(\tau^*(\alpha))$ is increasing in α on this range, completing the proof in this case.

Case 2: $\alpha > \alpha^*$.
Since we have that $\alpha > \alpha^*$ and we have showed that $\tau^*(\alpha)$ is increasing in α, it must be the case that $f_\alpha(\tau^*(\alpha))$ is decreasing in α for $\alpha > \alpha^*$, as by definition $\Gamma(1/\alpha)/\alpha$ is increasing for all $\alpha > \alpha^*$.

This means that if it is the case that

$$\frac{f_\alpha(\tau^*(\alpha))^2}{F_\alpha(\tau^*(\alpha))(1 - F_\alpha(\tau^*(\alpha)))} \tag{14}$$

is increasing in α, then the claim will hold in this case. This is because, for (14) to be increasing, if the numerator is decreasing then the denominator must be in turn decreasing. This implies that $F_\alpha(\tau^*(\alpha))$ is increasing in α, which is what we set out to show. A stronger condition, which is what we will prove, is that

$$\frac{\left(\frac{\alpha}{\Gamma(1/\alpha)}\right)^2}{(1 + P(1/\alpha, \tau))(1 - P(1/\alpha, \tau))} \tag{15}$$

is increasing in α for fixed τ, where $P(1/\alpha, \tau) := \frac{\Gamma(1/\alpha,\tau)}{\Gamma(1/\alpha)}$ is the normalized incomplete

Gamma function. This is a stronger condition because it is saying is that if we look at values of τ for each α that make $e^{-\tau^\alpha}$ equal, then these values of τ cause this ratio to increase. In particular, consider evaluating (14) at $\tau^*(\alpha)$ for some α. If (15) is increasing in α, then for all $\alpha' > \alpha$ it must be the case that

$$\frac{f_{\alpha'}(\tau^*(\alpha)^{\alpha/\alpha'})^2}{F_{\alpha'}(\tau^*(\alpha)^{\alpha/\alpha'})(1-F_{\alpha'}(\tau^*(\alpha)^{\alpha/\alpha'}))} > \frac{f_\alpha(\tau^*(\alpha))^2}{F_\alpha(\tau^*(\alpha))(1-F_\alpha(\tau^*(\alpha)))}.$$

Since by optimality it is the case that

$$\frac{f_{\alpha'}(\tau^*(\alpha'))^2}{F_{\alpha'}(\tau^*(\alpha'))(1-F_{\alpha'}(\tau^*(\alpha')))} \geq \frac{f_{\alpha'}(\tau^*(\alpha)^{\alpha/\alpha'})^2}{F_{\alpha'}(\tau^*(\alpha)^{\alpha/\alpha'})(1-F_{\alpha'}(\tau^*(\alpha)^{\alpha/\alpha'}))},$$

this will prove the claim.

Now, to show that (15) is indeed increasing in α. Perform the switch to $\beta := 1/\alpha$, as the in proof of Lemma 7. Then, the above can be rewritten as

$$\begin{aligned}&\frac{\left(\frac{1}{\beta\Gamma(\beta)}\right)^2}{(1+P(\beta,x))(1-P(\beta,x))} \\ &= \frac{1}{2\beta^2\Gamma(\beta)}\left(\frac{1}{\Gamma(\beta)+\gamma(\beta,x)}+\frac{1}{\Gamma(\beta)-\gamma(\beta,x)}\right).\end{aligned}$$

Multiplying by 2 and then taking the derivative with respect to β here yields

$$\begin{aligned}&-\frac{2\beta\Gamma(\beta)+\beta^2\int_0^\infty \log(y)d\lambda_\beta(y)}{(\beta^2\Gamma(\beta))^2}\left(\frac{1}{\Gamma(\beta)+\gamma(\beta,x)}+\frac{1}{\Gamma(\beta)-\gamma(\beta,x)}\right) \\ &-\frac{1}{\beta^2\Gamma(\beta)}\left(\frac{\int_0^\infty \log(y)d\lambda_\beta(y)+\int_0^x \log(y)d\lambda_\beta(y)}{(\Gamma(\beta)+\gamma(\beta,x))^2}+\frac{\int_x^\infty \log(y)d\lambda_\beta(y)}{(\Gamma(\beta)-\gamma(\beta,x))^2}\right).\end{aligned}$$

Here I use the definition of $\lambda_\beta(y)$ from the proof of Lemma 7. The goal is to show that this object is negative. Again let $\Gamma'(\beta)/\Gamma(\beta) =: \psi(\beta)$ and observe that for $\alpha > \alpha^*$ by definition it the case that $\frac{1}{\beta}+\psi(\beta) < 0$. Now, straightforward arithmetic shows that the sign of this is equivalent to the sign of

$$-\frac{\int_0^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right)d\lambda_\beta(y)+\int_0^x \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right)d\lambda_\beta(y)}{(\Gamma(\beta)+\gamma(\beta,x))^2}$$

$$-\frac{\int_x^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y)}{(\Gamma(\beta)-\gamma(\beta,x))^2}.$$

I first show that for all $x \in (0,\infty)$ and β, it is the case that.

$$-\int_x^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y) < 0.$$

Now, consider that on this region of β, by definition it is the case that

$$-\left(\frac{2}{\beta}+\psi(\beta)\right) = 2\left(\frac{1}{\beta}+\psi(\beta)\right)+\psi(\beta) < \psi(\beta).$$

This means that

$$-\int_x^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y) < \int_x^\infty \left(-\log(y)+\psi(\beta)\right) d\lambda_\beta(y)$$

Now, the sign of the above is equivalent to the sign of

$$\frac{\int_x^\infty \left(-\log(y)+\psi(\beta)\right) d\lambda_\beta(y)}{\int_x^\infty d\lambda_\beta(y)}.$$

This, however, is clearly decreasing in x, because smaller values of x correspond to the average log value being smaller. Since it is clearly equal to zero at $x=0$, this shows that

$$-\int_x^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y) < 0.$$

This in turn implies that

$$-\frac{\int_x^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y)}{(\Gamma(\beta)-\gamma(\beta,x))^2} < -\frac{\int_x^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y)}{(\Gamma(\beta)+\gamma(\beta,x))^2}.$$

This means that it is sufficient to check the sign of

$$-\int_0^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y) - \int_0^x \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y)$$
$$-\int_x^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y)$$

and ensure that it is negative. But this expression is equal to

$$-2\int_0^\infty \left(\log(y)+\frac{2}{\beta}+\psi(\beta)\right) d\lambda_\beta(y) = -\Gamma'(\beta) - \left(\frac{2}{\beta}+\psi(\beta)\right)\Gamma(\beta) < 0,$$

where this last inequality follows from the observation above:

$$\left(\frac{2}{\beta}+\psi(\beta)\right)\Gamma(\beta) > -\Gamma'(\beta)$$

This completes the proof in this case. □

Proof of Proposition 3. First, notice that

$$f(s) := f_{\alpha_L,\alpha_H}(s) = \frac{1}{\frac{\Gamma(1/\alpha_L)}{\alpha_L}+\frac{\Gamma(1/\alpha_H)}{\alpha_H}}\left(\mathbb{1}\{s\geq 0\}e^{-s^{\alpha_H}} + \mathbb{1}\{s\leq 0\}e^{-(-s)^{\alpha_L}}\right)$$

Recall that, by definition, $\frac{\Gamma(1/\alpha)}{\alpha}$ is increasing in α for all $\alpha \geq \alpha^*$. This means, in particular, that there is a larger mass of positive signals than negative signals in this case.

What must be shown is that

$$\frac{f(\tau)^2}{F(\tau)(1-F(\tau))}$$

is maximized for a negative value of τ. For ease of notation, let $\alpha := \alpha_H$ and $\alpha' := \alpha_L$. Notice that as before, the numerator is strictly increasing in τ for $\tau < 0$, and strictly decreasing in τ for $\tau > 0$. Since $\alpha' > \alpha$, for small values of τ there is larger persistence in the value of f on the negative side of the distribution.[18] Explicitly, given a $\tau > 0$, we will have that $-\tau^{\alpha/\alpha'} < -\tau$ (for $\tau < 1$, which will again be optimal here) will have the same value of the numerator. From here, the claim is two-fold.

Claim 1: The optimal choice of τ for $\tau > 0$ has $\tau \geq \tau^*(\alpha)$.

To see this, consider that the first order condition is given by

$$-2\alpha\tau^{\alpha-1} - e^{-x^\alpha}\left[\frac{1}{\frac{\Gamma(1/\alpha')}{\alpha'}+\int_0^\tau e^{-x^\alpha}dx} - \frac{1}{\int_\tau^\infty e^{-x^\alpha}dx}\right].$$

[18] Of course, for $\tau > 1$ this is reversed, because the tails are smaller on the negative side of the distribution than the positive side of the distribution.

As $\frac{\Gamma(1/\alpha')}{\alpha'} > \frac{\Gamma(1/\alpha)}{\alpha}$, the first fraction is smaller, so that the positive component is larger. This means that for the first order condition to be satisfied for $\tau > 0$ we need $\tau > \tau^*(\alpha)$.

Claim 2: For all $\tau > \tau^*(\alpha)$, it is the case that $1 - F(\tau) > F(-\tau^{\alpha/\alpha'})$.

Notice that

$$1 - F(\tau) = \frac{1}{\frac{\Gamma(1/\alpha)}{\alpha} + \frac{\Gamma(1/\alpha')}{\alpha'}} \int_\tau^\infty e^{-x^\alpha} dx = \frac{1}{\frac{\Gamma(1/\alpha)}{\alpha} + \frac{\Gamma(1/\alpha')}{\alpha'}} \cdot \frac{\Gamma\left(\frac{1}{\alpha}, \tau^\alpha\right)}{\alpha}$$

and

$$F(-\tau^{\alpha/\alpha'}) = \frac{1}{\frac{\Gamma(1/\alpha)}{\alpha} + \frac{\Gamma(1/\alpha')}{\alpha'}} \int_{\tau^{\alpha/\alpha'}}^\infty e^{-x^{\alpha'}} dx = \frac{1}{\frac{\Gamma(1/\alpha)}{\alpha} + \frac{\Gamma(1/\alpha')}{\alpha'}} \cdot \frac{\Gamma\left(\frac{1}{\alpha'}, \tau^\alpha\right)}{\alpha'}$$

From here it is sufficient to show that $\frac{\Gamma(1/\alpha, \tau^*(\alpha)^\alpha)}{\alpha} > \frac{\Gamma(1/\alpha', \tau^*(\alpha)^\alpha)}{\alpha'}$ for all $\alpha' > \alpha$. Notice that the value of τ at which this is evaluated is important: it does not hold for all τ (indeed, it clearly will not hold for $\tau = 0$).

Subclaim: $\frac{\Gamma(1/\alpha, \tau^\alpha)}{\alpha} > \frac{\Gamma(1/\alpha' \tau^\alpha)}{\alpha'}$ for $\alpha \geq \alpha^*$, $\tau^\alpha \geq \tau^*(\alpha^*)^{\alpha^*} > 0.001$.

Now, do the standard change of variable for $\beta := 1/\alpha$ and $x := \tau^\alpha$. I show that for $\beta < 1/\alpha^*$ it is the case that

$$\frac{\partial}{\partial \beta} (\Gamma(\beta, x) \cdot \beta) > 0$$

for all $x > 0.001$. The above is equivalent to (after some straightforward algebra in the same vein as above)

$$-\frac{1}{\beta} < \frac{\int_x^\infty \log(y) d\lambda_\beta(y)}{\int_x^\infty d\lambda_\beta(y)} \tag{16}$$

where the $d\lambda_\beta(y)$ notation is the same as in Lemma 7. Notice first that the right-hand side of (16) is (i) increasing in x and (ii) increasing in β. Property (i) holds because is because log is increasing, and so by increasing x the average log value will increase. Property (ii) follows from a similar logic: smaller β put more weight on smaller values of y (and hence smaller values of log).

This means that if (16) holds for $x^* := 0.001$, then the (sub)claim will be proved. From here, several straightforward computations will prove the claim:

$$-2.16 < \frac{\int_{x^*}^{\infty} \log(y) d\lambda_{1/2.7}(y)}{\int_{x^*}^{\infty} d\lambda_{1/2.7}(y)}, \tag{17}$$

dealing with $\beta \in \left(\frac{1}{2.7}, \frac{1}{\alpha^*}\right]$. Then,

$$-2.7 < \frac{\int_{x^*}^{\infty} \log(y) d\lambda_{1/4}(y)}{\int_{x^*}^{\infty} d\lambda_{1/4}(y)}, \tag{18}$$

dealing with then $\beta \in \left(\frac{1}{4}, \frac{1}{\alpha^*}\right]$. Finally, it is the case that

$$-4 < \frac{\int_{x^*}^{\infty} \log(y) d\lambda_0(y)}{\int_{x^*}^{\infty} d\lambda_0(y)}.$$

This means that for all $\beta \in (0, \alpha^*]$ it is the case that (16) holds. This proves Claim 2.

From here, the conclusion is reached by simply putting together Claims 1 and 2: for any possible optimal positive τ, there is a negative threshold that performs strictly better, so the optimal threshold must be negative. □

A.3 Proofs for Section 5.1

Micro-founding Separable Propensities: I suppose now that reviewers are defined by three characteristics. As before, reviewer i receives a conditionally independent signal s_i about the state of the world. Now, however, their willingness to review $w_i = (w_{i,1}, w_{i,2})$ is comprised of two components.

As before, I assume that s_i and w_i are independent, and also that $w_{i,1}$ and $w_{i,2}$ are independent. The reviewer's decision to review now follows a two-stage procedure. First, as before, the reviewer compares $w_{i,1}$ to a *review-dependent* threshold, which captures the heterogeneity in reporting rates across reviews. Let p_r be the probability that a reviewer "passes" this first stage. This proportion p_r can be thought of as the proportion of the population who would ever leave a review of type r: many consumers never leave written reviews, but are open to leaving a five-star review.

Second, the reviewer compares $w_{i,2}$ to a *signal-dependent* threshold. Let $q(s)$

be the probability with which a reviewer with signal s passes this second stage, conditional on passing the first stage, and assume that $q(s)$ is continuous as a function of s. That is, within the population of mass p_r who might leave a review of type r, among those with signal s only proportion $q(s)$ will actually leave the review: not everyone who sometimes writes reviews will always leave a review. Together, these factors combine to mean that the probability that a reviewer with signal s facing review r leaves a review is $p_r q(s)$.

Proof of Theorem 3. I proceed with the computations in many steps. At risk of abusing notation, let

$$f(s, \mu, \lambda) := f(s + \mu)^{\lambda} f(s - \mu)^{1-\lambda}$$

For an abuse of notation, let $p(s) := p(r, s) = p_r q(s)$ when the review function in question is clear. let $\rho(\mu; p, \lambda)$ denote the value of $\rho(\mu; p)$ evaluated at a fixed (but not necessarily the optimal) λ.

It can easily be seen that $\rho(\mu; p), \rho(\mu; r, p) \to 0$. Hence, l'Hopital's rule must be applied. First, we look at $\lim_{\mu \to 0} \frac{\partial}{\partial \mu} \rho(\mu; p)$. The regularity assumptions on f mean that, for a fixed λ, we have that this is given by

$$\begin{aligned}
&-\frac{\partial}{\partial \mu} \rho(\mu; p, \lambda) = \\
&\int \left(\lambda \frac{f'(s + \mu)}{f(s + \mu)} - (1 - \lambda) \frac{f'(s - \mu)}{f(s - \mu)} \right) f(s, \mu, \lambda) p(s) ds \\
&\quad + \lambda \frac{\int ((1 - p(s)) f(s + \mu) ds)^{\lambda - 1}}{\int ((1 - p(s)) f(s - \mu) ds)^{\lambda - 1}} \int ((1 - p(s)) f'(s + \mu) ds) \\
&\quad - (1 - \lambda) \frac{\int ((1 - p(s)) f(s + \mu) ds)^{\lambda}}{\int ((1 - p(s)) f(s - \mu) ds)^{\lambda}} \int ((1 - p(s)) f'(s - \mu) ds) .
\end{aligned}$$

Notice that because of the envelope theorem, we need not consider here the derivative with respect to λ. Consider now the limit as $\mu \to 0$. First, $f(s, \mu, \lambda) \to f(s, \mu, \lambda)$. Hence, the limit is equal to

$$-\lim_{\mu \to 0} \frac{\partial}{\partial \mu} \rho(\mu; p, \lambda) = (2\lambda - 1) \int p(s) \ell_f(s) f(s) ds + (2\lambda - 1) \int (1 - p(s)) f'(s) ds.$$

This is equal to zero because $\ell_f(s)f(s) = f'(s)$, and $\int f'(s)ds = 0$ (because the expectation of the linear score function is zero).

Hence, we must consider the second derivative. Here it can be verified that the derivative with respect to λ will indeed be irrelevant, because this value is zero. This means that we need not consider how λ varies with μ, and again may consider evaluating at λ point-wise.

Because the notation gets so tedious, for readability of the null review terms, let $H_+(\mu, n) := \int(1-p(s))f(s+\mu)ds$ (where n refers to "null"), and similarly $H_-(\mu, n)$. In a similar vein, let $H'_+(\mu, n) := \int(1-p(s))f'(s+\mu)ds$, and similarly $H'_-(\mu, n)$ (and $H''_\pm(\mu, n)$). Computing, we get that

$$
\begin{aligned}
&-\frac{\partial^2}{\partial\mu^2}\rho(\mu; p, \lambda) = \\
&\quad \int \lambda(\lambda-1)\left(\left(\frac{f'(s+\mu)}{f(s+\mu)}\right)^2 + \left(\frac{f'(s+\mu)f'(s-\mu)}{f(s+\mu)f(s-\mu)}\right)\right) f(s,\mu,\lambda)p(s)ds \\
&\quad -\int \lambda(1-\lambda)\left(\left(\frac{f'(s-\mu)}{f(s-\mu)}\right)^2 + \left(\frac{f'(s-\mu)f'(s+\mu)}{f(s-\mu)f(s+\mu)}\right)\right) f(s,\mu,\lambda)p(s)ds \\
&\quad +\int \left(\lambda\frac{f''(s+\mu)}{f(s+\mu)} + (1-\lambda)\frac{f''(s-\mu)}{f(s-\mu)}\right) f(s,\mu,\lambda)p(s)ds \\
&\quad +\lambda(\lambda-1)\frac{H_+(\mu,n)^{\lambda-1}}{H_-(\mu,n)^{\lambda-1}}\left(\frac{H'_+(\mu,n)}{H_+(\mu,n)} + \frac{H'_-(\mu,n)}{H_-(\mu,n)}\right) H'_+(\mu,n) \\
&\quad -(1-\lambda)\lambda\frac{H_+(\mu,n)^{\lambda}}{H_-(\mu,n)^{\lambda}}\left(\frac{H'_+(\mu,n)}{H_+(\mu,n)} + \frac{H'_-(\mu,n)}{H_-(\mu,n)}\right) H'_-(\mu,n) \\
&\quad +\lambda\frac{H_+(\mu,n)^{\lambda-1}}{H_-(\mu,n)^{\lambda-1}} H''_+(\mu,n) + (1-\lambda)\frac{H_+(\mu,n)^{\lambda}}{H_-(\mu,n)^{\lambda}} H''_-(\mu,n).
\end{aligned}
$$

We can then see that the limit is given by

$$
\begin{aligned}
&\lim_{\mu\to 0}\frac{\partial^2}{\partial\mu^2}\rho(\mu; p, \lambda) = \\
&\qquad 4\lambda(1-\lambda)\int\left(\frac{f'(s)}{f(s)}\right)^2 f(s)p(s) + \int f''(s)p(s)ds \\
&\qquad +4\lambda(1-\lambda)\frac{H'_+(0,n)^2}{H_+(0,n)} + H''_+(0,n) \\
&= \quad 4\lambda(1-\lambda)\int\left(\frac{f'(s)}{f(s)}\right)^2 f(s)p(s) + \int p(s)f''(s)ds
\end{aligned}
$$

$$+4\lambda(1-\lambda)\frac{\left(\int(1-p(s))f'(s)ds\right)^2}{\int(1-p(s))f(s)ds}+\int(1-p(s))f''(s)ds.$$

Now, as $\int f''(s)ds = 0$ (because its anti-derivative is f'), $\int f'(s)ds = 0$, and $\int f(s)ds = 1$, this value is equal to the numerator in the statement of the proposition when $\lambda = \frac{1}{2}$. Notice that this value of λ is clearly the maximizer of this object. minimized object is the negative of this, $\lambda = \frac{1}{2}$ will minimize the numerator, so it is the optimal value of λ.

Of course, at this stage it has not been shown that the computations that were computed above correspond to the value of the numerator: the denominator needs to go to 0 at the same rate for this to be satisfied. However, for linearity of the argument it is worthwhile to perform all of the above computations in order. I now turn attention to the denominator. The terms in the denominator are of the following form. From hereon $p(s)$ to refer to $p_r q(s)$. At the end, we shall replace these values with their true values. Define

$$g_{\tilde{r}}(\mu,\lambda) := \left(\int_{\{s:r(s)\in\tilde{r}\}} f(s+\mu)p(s)ds\right)^{\lambda}\left(\int_{\{s:r(s)\in\tilde{r}\}} f(s-\mu)p(s)ds\right)^{1-\lambda}.$$

Similarly, for null reviews we have

$$g_n(\mu,\lambda) := \left(\int f(s+\mu)(1-p(s))ds\right)^{\lambda}\left(\int f(s-\mu)(1-p(s))ds\right)^{1-\lambda}.$$

The value of the denominator for a fixed λ is of the form

$$\rho(\mu;r,p,\lambda) := 1-\sum_{\tilde{r}\in R_r} g_{\tilde{r}}(\mu,\lambda)-g_n(\mu,\lambda).$$

In order to make computations easier to follow, I compute the limits term by term. Notice that the computations for $g_n(\mu,\lambda)$ have already been computed above. For notational clarity, define

$$H_+(\mu,\tilde{r}) = \int_{\{r(s)\in\tilde{r}\}} f(s+\mu)p(s)ds.$$

and similarly all of the related functions $(H_-(\mu,\tilde{r}), H'_\pm(\mu,\tilde{r}), H''_\pm(\mu,\tilde{r}))$. Notice here that the term is $p(s)$ and not $1-p(s)$, because these are submitted reviews. Let

$$\frac{\partial}{\partial \mu} g_{\tilde{r}}(\mu,\lambda) = \lambda \frac{H_+(\mu,\tilde{r})^{\lambda-1}}{H_-(\mu,\tilde{r})^{\lambda-1}} H'_+(\mu,\tilde{r}) - (1-\lambda)\frac{H_+(\mu,\tilde{r})^{\lambda}}{H_-(\mu,\tilde{r})^{\lambda}} H'_-(\mu,\tilde{r}).$$

This means that

$$\lim_{\mu\to 0} \frac{\partial}{\partial \mu} g_{\tilde{r}}(\mu,\lambda) = (2\lambda-1)H'(0,\tilde{r}).$$

Recall that we have already computed the dynamics of $g_n(\mu,\lambda)$ above. This means that, regardless of λ, we have that

$$\begin{aligned}
-\lim_{\mu\to 0} \frac{\partial}{\partial \mu}\rho(\mu;r,p,\lambda) =& (2\lambda-1)\left(\sum_{\tilde{r}} H'_+(0,\tilde{r}) + H'_+(0,n)\right) \\
=& (2\lambda-1)\left(\sum_{\tilde{r}} \int_{\{r(s)\in\tilde{r}\}} f'(s)p(s)ds + \int f'(s)(1-p(s))ds\right) \\
=& (2\lambda-1)\int f'(s)ds = 0.
\end{aligned}$$

The reasons that we need not consider λ are similar to the reasons in the case of the numerator. Because of this, I may focus on point-wise evaluation with respect to λ throughout. In particular, this means that the we must proceed and take another derivative. We have that

$$\begin{aligned}
\frac{\partial^2}{\partial \mu^2} g_{\tilde{r}}(\mu,\lambda) =& \lambda(\lambda-1)\frac{H_+(\mu,\tilde{r})^{\lambda-1}}{H_-(\mu,\tilde{r})^{\lambda-1}}\left(\frac{H'_+(\mu,\tilde{r})}{H_+(\mu,\tilde{r})} + \frac{H'_-(\mu,\tilde{r})}{H_-(\mu,\tilde{r})}\right) H'_+(\mu,\tilde{r}) \\
&-(1-\lambda)\lambda\frac{H_+(\mu,\tilde{r})^{\lambda}}{H_-(\mu,\tilde{r})^{\lambda}}\left(\frac{H'_+(\mu,\tilde{r})}{H_+(\mu,\tilde{r})} + \frac{H'_-(\mu,\tilde{r})}{H_-(\mu,\tilde{r})}\right) H'_-(\mu,\tilde{r}) \\
&+\lambda\frac{H_+(\mu,\tilde{r})^{\lambda-1}}{H_-(\mu,\tilde{r})^{\lambda-1}} H''_+(\mu,\tilde{r}) + (1-\lambda)\frac{H_+(\mu,\tilde{r})^{\lambda}}{H_-(\mu,\tilde{r})^{\lambda}} H''_-(\mu,\tilde{r}).
\end{aligned}$$

Notice that these computations are similar to those used for $g_n(\mu,\lambda)$ in the computation of the numerator's second derivative. Note again that we need not consider the variation in λ here.

This means that

$$\lim_{\mu \to 0} \frac{\partial^2}{\partial \mu^2} g_{\tilde{r}}(\mu, \lambda) = 4\lambda(\lambda - 1)\frac{H'_+(0,\tilde{r})^2}{H_+(0,\tilde{r})} + H''_+(0,\tilde{r}).$$

We are almost done. All that is left is to collect terms and notice again $\int f''(s)ds = 0$:

$$\begin{aligned}
\lim_{\mu \to 0} \frac{\partial^2}{\partial \mu^2} \rho(\mu; r, p, \lambda) =& -\sum_{\tilde{r}} \left(4\lambda(\lambda - 1)\frac{H'_+(0,\tilde{r})^2}{H_+(0,\tilde{r})} + H''_+(0,\tilde{r})\right) \\
&- 4\lambda(\lambda - 1)\frac{H'_+(0,n)^2}{H_+(0,n)} - H''_+(0,n) \\
=& 4\lambda(1-\lambda)\left(\sum_{\tilde{r}} \frac{H'_+(0,\tilde{r})^2}{H_+(0,\tilde{r})} + \frac{H'_+(0,n)^2}{H_+(0,n)}\right).
\end{aligned}$$

Putting this result together with the computations from the numerator above show the claim. □

B Posterior Visualization

In this section I show how the forces that impact the performance of different review systems can be viewed in posterior space. In particular, this visualization provides additional insight to the results of Section 3.2.

Let $f(s) = \frac{1}{2}(f_L(s) + f_H(s))$ be the ex-ante distribution of signals under a symmetric prior. Moreover, let $\pi(s) := \frac{f_L(s)}{f_L(s)+f_H(s)}$ be the posterior belief placed on state L after observing signal s when there is a symmetric prior.

A straightforward computation shows that the learning efficiency under the full review is an expectation that depends on the posterior and the ex ante distribution:

$$\nu = 1 - \min_{\lambda \in [0,1]} 2 \int \pi(s)^\lambda (1 - \pi(s))^{1-\lambda} f(s) ds.$$

Associated with each signal s is a unique posterior π. Let γ_π be the measure on

$[0,1]$ that reflects the distribution of posteriors that is associated with f.[19] That is,

$$\gamma_\pi(B) := \frac{1}{2}\int_{\{s:\pi(s)\in B\}} f_L(s) + f_H(s)ds.$$

Then, the above can be rewritten as

$$\nu = 1 - \min_{\lambda\in[0,1]} 2\int \pi^\lambda (1-\pi)^{1-\lambda}\gamma_\pi(d\pi). \tag{19}$$

This structure is preserved when review systems are introduced, as long as the review system treats posteriors consistently. That is, as long as it treats two signals that induce the same posterior the same. To this end, call a review system r *consistent* if, for all signals $s, s' \in \mathbb{R}$ such that $\pi(s) = \pi(s')$, then $r(s) = r(s')$.

The structure observed in (19) extends to consistent review systems. The difference is that instead of the function $\pi^\lambda(1-\pi)^{1-\lambda}$ being evaluated at each π, the review system r groups all posteriors that are mapped to the same review.

Proposition 4. *Let r be any consistent review system. Then,*

$$\nu(r) = 1 - \min_{\lambda\in[0,1]} 2\int \left(\mathbb{E}_{\gamma_\pi}\left[\tilde{\pi}|r(\tilde{\pi}) = r(\pi)\right]\right)^\lambda \left(1 - \mathbb{E}_{\gamma_\pi}\left[\tilde{\pi}|r(\tilde{\pi}) = r(\pi)\right]\right)^{1-\lambda}\gamma_\pi(d\pi). \tag{20}$$

Proof. I show this in the case that r is a finite review system, and when γ^r has full support on R_r, since the notation is easier in this case. First, observe that $f_H(s) = f_L(s)\frac{1-\pi(s)}{\pi(s)}$. This means that we can write $\nu(r)$ as

$$\begin{aligned}
1-\nu(r) &= \min_{\lambda\in[0,1]} \sum_{\tilde{r}\in R_r} \mathbb{P}_L(r^{-1}(\tilde{r}))^\lambda \mathbb{P}_H(r^{-1}(\tilde{r}))^{1-\lambda} \\
&= \min_{\lambda\in[0,1]} \sum_{\tilde{r}\in R_r} \left(\int_{\{s\in r^{-1}(\tilde{r})\}} \frac{\pi(s)}{(1-\pi(s))} f_H(s)ds\right)^\lambda \cdot \left(\int_{\{s\in r^{-1}(\tilde{r})\}} f_H(s)ds\right)^{1-\lambda}
\end{aligned}$$

Now, if $\pi(s)$ is injective, it is the case (by straightforward computation) that $\frac{1}{1-\pi}f_H(s) = 2\gamma_\pi(d\pi(s))$ and so also $f_H(s) = 2(1-\pi(s))\gamma_\pi(d\pi(s))$. In general (even

[19] Although γ_π is the measure of posteriors on state L, I use γ_π to indicate that this is the ex-ante measure, and not the measure conditional on state L.

if $\pi(s)$ is not injective), for each λ, then,

$$\begin{aligned}
&\sum_{\tilde{r}\in R_r} \mathbb{P}_L(r^{-1}(\tilde{r}))^{\lambda}\mathbb{P}_H(r^{-1}(\tilde{r}))^{1-\lambda} \\
&\qquad = \sum_{\tilde{r}\in R_r} 2\left(\int_{\tilde{\pi}\in r^{-1}(\tilde{r})} \tilde{\pi}\gamma_\pi(d\tilde{\pi})\right)^{\lambda}\cdot\left(\int_{\tilde{\pi}\in r^{-1}(\tilde{r})} (1-\tilde{\pi})\gamma_\pi(d\tilde{\pi})\right)^{1-\lambda} \\
&\qquad = \sum_{\tilde{r}\in R_r} 2\frac{\left(\int_{\tilde{\pi}\in r^{-1}(\tilde{r})} \tilde{\pi}\gamma_\pi(d\tilde{\pi})\right)^{\lambda}\cdot\left(\int_{\tilde{\pi}\in r^{-1}(\tilde{r})} (1-\tilde{\pi})\gamma_\pi(d\tilde{\pi})\right)^{1-\lambda}}{\int_{\tilde{\pi}\in r^{-1}(\tilde{r})}\gamma_\pi(d\tilde{\pi})}\int_{\tilde{\pi}\in r^{-1}(\tilde{r})}\gamma_\pi(d\tilde{\pi}) \\
&\qquad = \sum_{\tilde{r}\in R_r} 2\left(\mathbb{E}_{\gamma_\pi}\left[\pi|r(\pi)=\tilde{r}\right]\right)^{\lambda}\left(1-\mathbb{E}_{\gamma_\pi}\left[\pi|r(\pi)=\tilde{r}\right]\right)^{1-\lambda}\int_{\tilde{\pi}\in r^{-1}(\tilde{r})}\gamma_\pi(d\tilde{\pi})
\end{aligned}$$

From here, we minimize over λ, and conclude. □

Notice that here the measure with respect to which the integral is taken is simply $\gamma_\pi(d\pi)$, which is *independent* of r. This means that, fixing λ, the loss in information can be thought of as coming exactly from the pre-evaluation of the function $\pi^{\lambda}(1-\pi)^{1-\lambda}$. Since, for any λ, this function is concave, information loss is exactly due to Jensen's inequality. The interpretation is exactly the same as the intuition as before: it is more efficient to separate an informative signal and an uninformative signal than it is to have them grouped into a moderately informative signal. Proposition 4 gives a way to exactly quantify this trade-off.

This analysis can be used to explain why increases in homogeneity increase the asymmetry of the optimal review system. In particular, the larger the spread of the posteriors within a bin, the more information will be lost within that bin, because the larger will be the difference in Jensen's inequality. For small values of μ, different values of α lead to very different distributions of posteriors $d\gamma_\pi$. As observed in Figure 2a, as α increases, simultaneously more signals are very uninformative and more signals are very informative. This is confirmed in Figure 2b, where the distributions of posteriors are plotted for the same values of α. As α increases, the spread in the distribution of posteriors increases, driving the increase in $\kappa_\alpha(0)$. At the same time, the larger mass near certainty (mass near 0 and 1) incentivizes moving the threshold away from symmetry.

B.1 Optimality of Threshold Rules

If r is a consistent review system that has b bins, then it is effectively a partition of the space $[0,1]$. Let $\{B_1, ..., B_b\}$ be the partition of the interval. Then (20) can be rewritten as

$$\nu(r) = 1 - \min_{\lambda \in [0,1]} 2 \sum_{i=1}^{b} \left(\mathbb{E}_{\gamma_\pi} [\pi | \pi \in B_i] \right)^{\lambda} \left(1 - \mathbb{E}_{\gamma_\pi} [\pi | \pi \in B_i] \right)^{1-\lambda} \gamma_\pi (B_i) .$$

With this structure, it is possible to show the intuitive result that optimal coarsening with a fixed number of bins must necessarily be threshold rules in posterior space. First, I explicitly define what it means to be defined by thresholds.

Definition 5. A review system r is called a *b-threshold* review function if $r^{-1}(\tilde{r})$ is convex (except perhaps on a set of measure zero) for each $\tilde{r} \in R_r$.

Optimal review systems are necessarily threshold rules.

Proposition 5. *Let r be an optimal review system with b bins. Then r is b-threshold review function.*

Proof. I show that if r is a finite review with b bins that is not such a threshold rule, then there exists a review with b bins that outperforms it. Moreover, I show that this holds point-wise for each λ, so that certainly it holds over the minimum of λ.

Consider two bins r_1 and r_2 that cannot be separated by any threshold. Let $\pi_i = \mathbb{E}[\pi | r(\pi) = r_i]$. Without loss assume that $\pi_1 \leq \pi_2$.

Consider any point $\overline{\pi}$ that lies between π_1 and π_2 (if they are equal, then $\overline{\pi} = \pi_1 = \pi_2$). By assumption, $\overline{\pi}$ is not a threshold for these two bins. In particular, this means that because $\pi_1 \leq \overline{\pi} \leq \pi_2$, there exists a set Π_1 with positive measure such that for all $\pi \in \Pi_1$, $\pi > \overline{\pi}$ and $r(\pi) = r_1$. That is, consider the following two sets

$$\Pi_1 := \{\pi | r(\pi) = r_1 \text{ and } \pi > \overline{\pi}\} \quad \text{and}$$
$$\Pi_2 := \{\pi | r(\pi) = r_2 \text{ and } \pi < \overline{\pi}\} .$$

Because $\overline{\pi}$ is not a threshold for r_1, r_2, and $\pi_1 \leq \overline{\pi} \leq \pi_2$, these sets both must have positive measure. Suppose that both of their measures are larger than ϵ. Select a subset of each that has mass ϵ: $\Pi_1^\epsilon, \Pi_2^\epsilon$. Let $\pi_i' = \mathbb{E}[\pi | r'(\pi) = r_i]$ denote the new conditional expectations. Because of the construction it must necessarily be the case that $\pi_1 > \pi_1'$ and $\pi_2 < \pi_2'$.

Now, define a new review function r' as follows:

$$r'(\pi) = \begin{cases} r(\pi) & \text{if} \quad \pi \notin \Pi_1^\epsilon, \Pi_2^\epsilon \\ r_1 & \text{if} \quad \pi \in \Pi_2^\epsilon \\ r_2 & \text{if} \quad \pi \in \Pi_1^\epsilon \end{cases}.$$

That is, r' agrees with r, except that it swaps the Π_i^ϵ's. I show that we necessarily have $\nu(r') \geq \nu(r)$. As mentioned, I show that this holds for each λ. To that end, fix a value of $\lambda \in (0,1)$, and let $\varphi(\pi) := \pi^\lambda(1-\pi)^{1-\lambda}$. I now show that the iterated expectation of $\varphi(\pi)$ on the partition induced by r' is lower than on the partition induced by r.

That is, I aim to show that

$$\sum_{i=1}^{b} \varphi\left(\mathbb{E}_{\gamma_\pi}\left[\pi | r(\pi) \in r_i\right]\right) \gamma_\pi(r^{-1}(r_i)) \geq \sum_{i=1}^{b} \varphi\left(\mathbb{E}_{\gamma_\pi}\left[\pi | r'(\pi) \in r_i\right]\right) \gamma_\pi((r')^{-1}(r_i)).$$

Notice now that the two agree except on r_1 and r_2, so that the sum simplifies. Now, consider adding a linear function $h(\pi)$ to $\varphi(\pi)$ such that the unique maximum of $h + \varphi$ is obtained at $\overline{\pi}$. Note that this is possible because φ is strictly concave. In particular, this means that $h+\varphi$ is strictly increasing for $\pi < \overline{\pi}$ and strictly decreasing for $\pi > \overline{\pi}$. Then, because h is linear, it is the case that

$$\begin{aligned} & \sum_{i=1}^{2} \varphi\left(\pi_i\right) \gamma_\pi(r^{-1}(r_i)) \geq \sum_{i=1}^{2} \varphi\left(\pi_i'\right) \gamma_\pi((r')^{-1}(r_i)) \\ \iff \quad & \sum_{i=1}^{2} \left(\varphi\left(\pi_i\right) + h(\pi_i)\right) \gamma_\pi(r^{-1}(r_i)) \geq \sum_{i=1}^{2} \left(\varphi\left(\pi_i'\right) + h(\pi_i')\right) \gamma_\pi((r')^{-1}(r_i)). \end{aligned}$$

As observed, it is the case that $\overline{\pi} \geq \pi_1 > \pi_1'$. This means that

$$\varphi\left(\pi_1\right) + h(\pi_1) > \varphi\left(\pi_1'\right) + h(\pi_1')$$

It is similarly the case that

$$\varphi(\pi_2) + h(\pi_2) > \varphi(\pi_2') + h(\pi_2')$$

Since by construction it is the case that $\gamma_\pi(r^{-1}(r_i)) = \gamma_\pi((r')^{-1}(r_i))$ for $i = 1, 2$, this means that we have achieved a strict improvement. Hence it must be the case that $\nu(r) < \nu(r')$. □

C Relationship between the Platform and Consumers' Preference over Information Structures

In this section, I study the perspective of a consumer choosing which source of information to use when making their decision. In practice, many platforms elicit multiple types of information from reviewers; often one can leave a coarse rating and then supplement the rating with a free-text full review.

In the work above I take the perspective of the platform ex ante choosing what type of review to elicit. In that choice the platform internalizes the randomness in the review decisions of reviewers. If there are multiple types of reviews that have been collected, when it comes time for consumers to choose between them, this randomness has been removed. So, the comparison is not between the rates of acquisition, but directly between the number of reviews of each type of review.

The comparison from the side of consumers is more similar to asking how many coarsened reviews "equals" one review of another type. The distinction between the two is subtle, but as shown below, the removal of risk leads the consumer to require more coarsened reviews. However, as individual reviews become uninformative, this difference disappears and so the metric of learning loss from Theorem 2 applies again.

Lemma 8. *Fix a consumer with a (finite) decision problem, two review systems r, r', and a degree of distinction μ. Then there exists an N large such that for all $n_r, n_{r'} > N$, n_r reviews of type r is preferred to $n_{r'}$ reviews of type r' if*

$$\frac{n_r}{n_{r'}} > \frac{\log(1 - \nu(\mu; r'))}{\log(1 - \nu(\mu; r))}.$$

Proof. Let (n_r, r) denote the statistical experiment that generates n_r conditionally independent signals of drawn according to r. Then, because each of the signals are independent,

$$\begin{aligned}\nu(\mu;(n_r,r)) =& 1-\min_{\lambda\in[0,1]}\int_R \dots \int_R \prod_{i=1}^{n_r} \left(\gamma_L^r(d\tilde{r}_i)\right)^{\lambda} \left(\gamma_H^r(d\tilde{r}_i)\right)^{1-\lambda} \\ =& 1-\min_{\lambda\in[0,1]}\prod_{i=1}^{n_r}\int_R \left(\gamma_L^r(d\tilde{r}_i)\right)^{\lambda} \left(\gamma_H^r(d\tilde{r}_i)\right)^{1-\lambda} \\ =& 1-(1-\nu(\mu;r))^{n_r}.\end{aligned}$$

From here, the result comes from an application of either Torgersen's theorem or Theorem 1. Note that this result follows the same logic as Proposition 3 in Mu et al. (2021). See also Chernoff (1952), which discusses this relationship at some length. □

This condition is similar to, but not exactly, the condition of Theorem 1. As the next Proposition highlights, because the information for consumers is not random, consumers place more weight on "good" information than the platform. However, as information becomes very noisy, this difference shrinks and eventually disappears in the limit. Specifically, the relationship is as follows.

Proposition 6. *Let* $\kappa(\mu;r',r) := \frac{\nu(\mu;r')}{\nu(\mu;r)}$ *and let* $\zeta(\mu;r',r) := \frac{\log(1-\nu(\mu;r'))}{\log(1-\nu(\mu;r))}$. *Then,*

$$\kappa(\mu;r',r) < \zeta(\mu;r',r) \iff \nu(\mu;r') > \nu(\mu;r).$$

However,

$$\lim_{\mu\to 0}\kappa(\mu;r',r) = \lim_{\mu\to 0}\zeta(\mu;r',r).$$

Proof. For all $x, y \in (0,1)$,

$$\frac{\log(1-x)}{\log(1-y)} > \frac{x}{y} \iff x > y.$$

This shows the first claim. For the second, one can apply l'Hopital's rule to the limit of $\mu \to 0$ for

$$\zeta(\mu;r',r) = \frac{\log(1-\nu(\mu;r'))}{\log(1-\nu(\mu;r))}.$$

and find that

$$\lim_{\mu \to 0} \zeta(\mu; r', r) = \lim_{\mu \to 0} \frac{\frac{\partial}{\partial \mu} \nu(\mu; r')}{\frac{\partial}{\partial \mu} \nu(\mu; r)} = \lim_{\mu \to 0} \eta(\mu; r', r).$$

This shows the second claim, completing the proof. □

The platform's choice between ratings systems must necessarily reflect the randomness with which it receives information. This means that, relative to a consumer who does not need to factor-in this randomness, the platform will value more frequent, but uninformative, reviews. This result suggests why in practice platforms collect multiple sources of information: consumers trade-off between sources of information in different ways than do platforms, because of the timing of their choice between sources of information.