Clicks or Comments? The quality-quantity trade-off of review systems*

Matthew J. H. Murphy†

This version: May 27, 2025

I study the optimal design of review systems. A platform, seeking to learn an unknown state, faces a trade-off between the informativeness and frequency of user-generated reviews. Detailed reviews provide more information per review, but reviewers submit them less frequently than simple reviews. I characterize the informational content of review systems, which depends on the precision of each review and the rate at which they are submitted by reviewers. I use this characterization to derive the platform's optimal review system when reviewer's signals are imprecise. I apply these results to study how reviewers' taste heterogeneity affects the optimal binary review system. Increased taste heterogeneity affects reviewers' information by increasing the dispersion of their signals. When reviewers' tastes are sufficiently heterogeneous, the optimal binary review is symmetric. When reviewers' preferences are sufficiently homogeneous, however, the optimal review is asymmetric. Regardless of the level of heterogeneity, the optimal binary review is preferred to a fully detailed review system if it is submitted at least 3.25 times as frequently.

*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, economic agents rely on others' information. In many settings, it is infeasible to compensate agents 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 not a viable option (e.g., a school eliciting students' course reviews). In these settings, agents freely choose whether to provide the principal with their information. Due to this tension—the principal wants information, but lacks a way to directly incentivize agents to provide it—the principal should reduce the cost of sharing information. One natural way to reduce costs is to ask for less information. Reducing the number or complexity of questions that a reviewer must answer increases the likelihood that they will complete the review (Bean and Roszkowski 1995). However, this method of reducing costs is not without a significant drawback: the simpler review elicits, and thus provides, less information.

This paper models and quantifies this trade-off. I show when a platform prefers to elicit more frequent coarse information rather than scarcer detailed information. Moreover, I characterize the optimal coarse review structure when signals are imprecise. In my model, a platform aims to learn an payoff-relevant unknown state. Reviewers each have a signal about the state. The platform chooses a "review system", which is a mapping from signals into reviews. Reviewers observe the mapping and choose whether or not to submit their signal. This choice depends on a private benefit from reviewing and the number of possible reviews: reviewers are less likely to leave a review when they must decide between many possible options. Since eliciting more detailed information makes reviewers less likely to submit their signal, the platform trades off between more frequent and more detailed information.

Motivated by online reviews, I focus on the case when the number of reviewers is large. I leverage large deviations techniques in order to characterize the rate at which the platform learns the state, for each review system. This rate determines the platform's preference over review systems and depends on both the average precision of a single review and the probability that a reviewer leaves a review. I quantify the precision of each review system relative to the full review. This "relative information" measures how much less frequently the full review must be submitted for the platform to prefer the simpler review system.

A key characteristic of online reviews is that a single review is almost uninformative: any one reviewer's experience is a very noisy signal about the underlying quality of the product. In this imprecise information context, I explicitly characterize the relative information of finite review systems. This allows me to tractably compute the optimal review system.

An important implication of my model is that the platform's preference over review systems is independent of its decision problem. The platform's utility is determined by how frequently it takes a sub-optimal action. When the number of reviewers is large, this probability decreases at the same rate regardless of the platform's decision problem. Independence across decision problems 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 information produced by review systems is used for many different problems.

While the optimal review system is independent of the platform's decision problem, it critically depends on reviewers' information. In order to study how differences in reviewers' information affect review systems, I interpret reviewers' signals as their realized utility from experiencing a product of unknown quality. I apply my results to demonstrate how the heterogeneity of reviewers' idiosyncratic preferences impacts the optimal binary review system. When reviewers' preferences are heterogeneous, their signals are more dispersed. For some products (e.g., movies), taste heterogeneity is large, while for others (e.g., toasters), reviewers' tastes are homogeneous.

Simple reviews lose information because they combine signals that induce different beliefs. They lose more information when the signals they combine induce very different beliefs. When reviewers are heterogeneous, no reviewer is representative. This means that signal informativeness varies little. Hence, even binary review systems perform well. However, when reviewers are homogeneous, some reviewers have very informative signals (e.g., their toaster broke). Simple reviews mix these signals with a large mass of uninformative signals, drowning out the informative signals. As a result, when reviewers are homogeneous, binary reviews perform worse.

Moreover, the design of the optimal binary review varies depending on the level of reviewers' homogeneity. When reviewers are sufficiently heterogeneous, the optimal binary review system is simple: the platform asks reviewers if their experience was positive ("good") or negative ("bad"). When reviewers are sufficiently homogeneous, the platform uses an asymmetric review: it isolates either informative "horrible" or "amazing" experiences (i.e., it directly asks if their toaster broke). For very homogeneous populations, the naive "good" or "bad" binary review performs arbitrarily poorly relative to the full review. However, the optimal review bounds performance. 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 review.

My framework also explains why counter-intuitive review systems persist. For instance, platforms' reviews are positively skewing reviews (e.g., many users default to 5-stars on Uber and Airbnb, see Hu, Zhang, and Pavlou (2009)). I show that if reviewers' preferences exhibit less negative than positive heterogeneity (i.e., reviewers agree more on what constitutes a negative experience than a positive one), it is optimal for the platform to isolate "horrible" experiences, while grouping all other signals. This optimal review system results in a much larger number of positive reviews than negative reviews, even when the product is of low quality.

At a technical level, my paper contributes to the literature on comparing statistical experiments. My characterization of the trade-off between the frequency and quality in review systems extends existing results on how quickly different experiments lead to learning the state. In particular, I characterize this rate when signals are imprecise.

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 explicitly characterizes optimal review systems when information is imprecise. Section 4 applies these results to study the impact of taste 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 leads to eventual learning. 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 that 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 across review systems is the rate at which reviews are left. In the setting of Acemoglu et al. (2022) reviews are automatically submitted and as a result detailed review systems are always preferred to simpler systems; in my setting this is not the case.

A wider literature aims to study review systems outside of the context of standard social learning settings. For instance, Che and Hörner (2018) studies a platform that aims to determine 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. Che, Kim, and Zhong (2024) studies how statistical discrimination can arise in markets where consumers use ratings schemes to choose between products. A different approach is taken in Garg and Johari (2019), which uses large deviations techniques to optimally differentiate many underlying states by controlling the proportion of positive binary reviews for each quality level. A major difference between this work and Garg and Johari (2019) is that there, the designer has full control over the information system, whereas in my case the designer is constrained by the information of reviewers.

My work is also informed by the empirical literature on online reviews. For the common five-star system, it has been well documented that, across platforms, the distribution of submitted ratings is positively-skewed and bimodal (Chen, Yoon, and Wu 2004; Hu, Zhang, and Pavlou 2009). Hu, Pavlou, and Zhang (2017) studies how this distribution is due to a self-selection effect in the population of reviewers. 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 directly incentivizing reviews) can increase the rate at which reviews are submitted, they do not tend to improve the overall informativeness of the review system. This highlights the need to carefully design review systems, even when incentivizes are possible. For a survey of empirical work studying review systems, see Magnani (2020). Recent empirical work has also provided alternative reasons to use coarse review systems. For instance, Botelho et al. (2025) shows 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, and suggesting that skewed reviews may be optimal.

My paper contributes also to the literature studying and applying rates of learning to understand economic settings. 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 these 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 I study in the context of hypothesis testing.

2 Model

A state $\theta \in \Theta := \{L, H\}$ is drawn according to common prior $q \in (0, 1)$ that the state is L. A platform chooses an action from finite action set $\mathcal{A}$. The platform's payoff depends on its action and the state, $u : \mathcal{A} \times \Theta \to \mathbb{R}$.

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

There are N reviewers, indexed by $i \in \{1, ..., N\}$. Each reviewer observes a signal $S_i \in \mathbb{R}$, drawn from state-contingent density f_θ. Signals are independent conditional on the state. Let s_i denote a generic realization of the random variable S_i.

The platform chooses a review system, which is a measurable function mapping signals to reviews, $r : \mathbb{R} \to \mathbb{R}$. For instance, the platform can ask reviewers if their signal is above or below some threshold τ: $r(s_i) = \mathbb{1}\{s_i > \tau\}$. If the platform chooses review system r, and a reviewer with signal s_i submits her signal, the platform observes $r(s_i)$. Let $\mathcal{R}_r = \text{supp}(r(S_i))$ denote the set of reviews possible under r.

Reviewers bear a cost to submit a review, which can be interpreted as either cognitive or temporal. The cost is $c(r) = C(|\mathcal{R}_r|)$ for some increasing function $C : \mathbb{N} \cup \{\infty\} \to \mathbb{R}_+$: it is less costly to submit reviews with fewer choices.[2] Reviewer i also receives a payoff $w_i \geq 0$ from reviewing. This benefit from reviewing reflects the expressive or altruistic desire of a reviewer to share her information and is distributed according to random variable W_i. The W_i are independent across reviewers and are independent from signals. She receives utility $w_i - c(r)$ from reviewing, and 0 utility from not. It follows that reviewers use a simple threshold rule: if $w_i \geq c(r)$ reviewer i will review, and if $w_i < c(r)$ she will not. Let $p_r := \mathbb{P}(W_i \geq c(r))$ be the ex-ante probability that a reviewer submits a review r.

The timing is as follows. First, the platform chooses a review system. Then, reviewers' private signals and benefit from reviewing are drawn, and reviewers decide whether to submit their reviews. After this, the platform observes the submitted reviews, updates its beliefs about the state, and 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:

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

After observing $\mathcal{S}(I)$, the platform updates its belief. Let $\pi(r, \mathcal{S}) = \pi(L|r, \mathcal{S}(I))$

[2] Much recent work has shown that individuals have preferences against complexity (e.g., Oprea (2020)). The number of objects under consideration is a driver of complexity (Puri 2022). Recently, Wang and Li (2025) studied the impact of increased cognitive load on user engagement on online platforms. That work finds that decreases in complexity lead to more user engagement.

denote the posterior belief on the state being L after observing reviews $\mathcal{S}$, where going forward I drop the reliance on I for clarity. After updating its beliefs, the platform chooses an action to maximize its expected utility given beliefs $\pi(\mathcal{S})$:

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

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

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

2.1 Discussion of Model

In many contexts principals must use others' information to make informed decisions. My model provides a framework where asking simpler questions makes agents more likely to agree to provide their information. Importantly, I do not take a stance on the goal of the platform. In some settings, the platform's goal is to simply share the information that it collects with other agents. In other settings, the platform may want to skew the information that it provides to future agents in order to encourage different behaviour. Since the model does not assume that the platform's goals are aligned with future agents, it 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 system in Acemoglu et al. (2022) and Che and Hörner (2018) is to maximize the expected utility of consumers. The primary focus of these papers is how platforms can incentivize consumers to explore a product of unknown quality. In my work, the platform's incentives may not be aligned directly with consumers. I abstract from the problem of experimentation in order to directly compare review systems. In practice, a platform can optimally collect information for itself while considering what is optimal to show consumers: the two objectives are compatible.

A key assumption is that the probability a reviewer submits a review is independent of a reviewer's signal realization. This significantly simplifies the exposition of my results, but in practice reviewers with more extreme signals are more likely to submit reviews (Lafky 2014; Hu, Pavlou, and Zhang 2017). I extend the model to allow for this dependence in Section 5.1. The main insights of the baseline model extend to that setting.

One practical benefit to using simple review systems is that they are easier to digest and process than full-length reviews. I abstract from processing costs for the platform to interpret submitted reviews. I also assume that there is no possibility of miscommunication: when a reviewer submits a review, it is reported perfectly. Abstracting from these forces allows me to focus on the key quantity-quality trade-off between simple and detailed reviews. These forces provide an additional reason for the platform to use simple reviews.

3 The Informational Content of Review Systems

In this section, I quantify the informational content of review systems when there are many reviewers. I first show how a platform trades off between quality (appropriately defined) and frequency, then characterize the optimal review system when the reviewers' signals are imprecise.

3.1 Trading-off between Quality and Frequency in Review Systems

Given a review system r, let γ_θ^r denote the distribution of reviews conditional on state θ. That is, for any measurable subset of possible reviews $B \subseteq \mathcal{R}_r$ (where measurability is inherited from the Borel σ-algebra),

$$\gamma_\theta^r(B) := \mathbb{P}_\theta(r^{-1}(B)) = \int_{\{s:r(s)\in B\}} f_\theta(s)ds.$$

The measures γ_L^r and γ_H^r reflect the distribution of reviews *conditional* on reviews being submitted.

Review system r is informative if γ_L^r and γ_H^r are different. As the difference increases, the review systems aggregates information more quickly. The distribution of log-likelihood ratios $\log\left(\frac{d\gamma_L^r}{d\gamma_H^r}\right)$ encodes this difference. Conditional on $\theta = H$, a review is informative of the wrong state if $\log\left(\frac{d\gamma_L^r}{d\gamma_H^r}\right) \geq 0$. For any λ, from Markov's inequality,

$$\mathbb{P}_H\left(\log\left(\frac{d\gamma_L^r}{d\gamma_H^r}\right) \geq 0\right) \leq \mathbb{E}_H\left[e^{\lambda \log\left(\frac{d\gamma_L^r}{d\gamma_H^r}\right)}\right] = \mathbb{E}_H\left[\left(\frac{d\gamma_L^r}{d\gamma_H^r}\right)^{\lambda}\right].$$

Hence, the moment generating function of log-likelihood ratios provides a bound on the probability of reviews that are indicative of the wrong state. Since this relationship holds for any $\lambda \in \mathbb{R}$, the minimum of the moment generating function provides the tightest bound. When the number of reviews is large, this bound is tight.[3] Hence, the minimum of this moment generating function determines the rate at which the platform ceases to make mistakes about the state. If this value is small, it is more likely that the platform is correct in its belief after observing reviews.

Definition 1. The *learning efficiency* of review system r is the following measure of distance between γ_H^r and γ_L^r:

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

The learning efficiency is a simple transformation of the minimum of the moment-generating function of log-likelihoods. It encodes how quickly the platform learns the state. The learning efficiency of a review system r is larger if r separates the states more effectively. If $d\gamma_L^r > 0 \iff d\gamma_H^r = 0$, learning occurs immediately and $\nu(r) = 1$. If reviews are uninformative ($\gamma_L^r = \gamma_H^r$), $\nu(r) = 0$ and learning never occurs.

The speed at which the platform learns the state controls the platform's expected utility when the number of reviewers is large. Consider first the case that all reviews are submitted ($p_r = 1$). The platform's utility is determined by the probability that it takes a sub-optimal action. Since sub-optimal actions are caused by incorrect beliefs, $\nu(r)$ determines this probability. 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's expected

[3]For any informative review system, the expected value of the log-likelihood ratio is negative. When the number of reviews is large, the average belief drifts towards the true state, and so it is the probability of unlikely beliefs that binds. This is why the Kullback-Leibler divergence is not the correct notion of difference in this case.

utility is higher under r than under r'.[4]

Decreasing the probability that reviews are submitted p_r slows the learning of the platform. Intuitively, if reviews are submitted half of the time, the platform learns half as quickly. Theorem 1 formalizes this intuition: regardless of the decision problem, $p_r\nu(r)$ determines the utility of the platform for large N. In particular, the platform separably trades off between the quality of reviews and the rate at which they are submitted.

Theorem 1. *Fix two review systems r, r' such that $p_r\nu(r) > p_{r'}\nu(r')$. For any finite action set $\mathcal{A}$ and utility function u such that the platform's decision problem is not trivial ($\operatorname{argmax}_{a\in\mathcal{A}} u(a, L) \cap \operatorname{argmax}_{a\in\mathcal{A}} u(a, H) = \emptyset$), there exists an $\overline{N}$ such that for all $N \geq \overline{N}$, such that $u^*(r, N) > u^*(r', N)$.*

The proof of Theorem 1 leverages existing results in the large deviations literature for the case that $p_r = 1$. To utilize 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. The subtlety of the proof is in the definition of $\nu(r)$. This measure admits the separability between reporting rates and learning efficiency that is not present in previous applications of large deviations techniques.

Theorem 1 implies that the platform need not tailor its review design to the specifics of each problem because $p_r\nu(r)$ is independent of the decision problem of the platform. This is particularly relevant when the platform is designing a review system that is to be used across a number of decision problems, as in the case of large organizations.

In practice, review frequencies p_r are simple for platforms to empirically compute through A/B testing. This suggests a focus on the comparison of learning efficiencies $\nu(r)$ for different review systems r. Theorem 1 is equivalent to: if

$$\frac{\nu(r)}{\nu(r')} > \frac{p_{r'}}{p_r}, \tag{1}$$

then there exists an $\overline{N}$ such that for all $N \geq \overline{N}$, such that $u^*(r, N) > u^*(r', N)$. Heuristically, if r' is submitted much less frequently than r ($p_{r'} \ll p_r$), then the

[4] This discussion summarizes a known result in the literature (Chernoff 1952; Torgersen 1981).

platform prefers r. The ratio $\nu(r)/\nu(r')$ formalizes how small $\frac{p_{r'}}{p_r}$ must be for r to be favoured. Consider in particular the full review $r_f(s) = s$. By setting $r' = r_f$, the value (1) becomes a measure of how much information review r contains relative to the full review r_f. I next define a measure of relative information in order to compare review systems to the full review system.

Definition 2. The *relative information* of a review function r is $\kappa(r) := \frac{\nu(r)}{\nu(r_f)} \in [0, 1]$.

No review system can contain more information per review than the full review system. The relative information $\kappa(r)$ of a review system r is a measure of how much information is lost when coarsening from the full review to r. It characterizes how much less frequently the full review r_f must be submitted than r for the platform to prefer eliciting reviews of type r to full reviews. For instance, if $\kappa(r) = 0.5$, then r must be submitted twice as often as the full review for the platform to prefer r. The relative information compares the informational content across review systems in terms of how frequently they are reported. To directly compare two different review systems r and r', the platform uses $\kappa(r)/\kappa(r')$.

Remark 1. It is important for the separability of $\nu(r)$ and p_r that the reporting rate p_r is independent of a reviewer's realized signal. As mentioned previously, I extend the model to allow an individual reviewer's reporting rate to depend on her realized signal in Section 5.1.

Remark 2. The p_r represent the stochastic rate at which signals are reported. An alternative is to ask how many signals n_r from r "equals" $n_{r'}$ signals from r'. I discuss this alternative in Section 5.2.

3.2 Threshold Systems and Imprecise Reviews

An important class of review systems is the class of *threshold systems*. In practice, review systems are monotone: a more negative experience leads to a more negative review. Threshold systems capture this behaviour.

Definition 3. A review system r is a k-threshold review system if there exist $\tau_0 = -\infty < \tau_1 < ... < \infty = \tau_k$ such that

$$r(s) = \sum_{i=1}^{k} r^i \mathbb{1}\left\{s \in (\tau_{i-1}, \tau_i]\right\}, \text{ for distinct } r^i \in \mathbb{R}.$$

I write $r_{\boldsymbol{\tau}}$ for the review system defined by thresholds $\boldsymbol{\tau} = (\tau_1, ..., \tau_{k-1})$.

Threshold systems are induced by natural partitions of $\mathbb{R}$: all signals contained in the interval $(\tau_{i-1}, \tau_i]$ are mapped to the same review r^i. In this section I show that threshold systems are optimal for a wide class of signal structures and characterize the optimal k-threshold system when the noise in reviewers' signals is large.

In order to study threshold review systems, I parameterize the informativeness of an individual review by assuming that a reviewer's signal is given by $S_i = \mu_\theta + \sigma\epsilon_i$, where

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

A reviewer's signal can be interpreted as her realized utility from a product whose quality is dependent on the state. Under this parameterization the μ_θ reflect the average utility from the good. The ϵ_i reflect the idiosyncratic taste component in the reviewers' experience of the good: different reviewers experience the same product differently. The parameter σ controls the noise level of the taste shocks. Let f be the density of ϵ_i. In this setting, f_L and f_H are given by

$$f_L(s) = \frac{1}{\sigma} f\left(\frac{s+\mu}{\sigma}\right) \quad \text{and} \quad f_H(s) = \frac{1}{\sigma} f\left(\frac{s-\mu}{\sigma}\right).$$

The following is a regularity assumption for f to guarantee that the platform's problem is well-behaved in this setting.

Assumption 1. The density of idiosyncratic taste shocks f has full support and is continuous. Moreover, f absolutely continuous almost everywhere, as is its derivative.

Simple review systems lose information because they combine signals that induce different beliefs. When a single review combines signals that induce very different beliefs, information loss is large. Hence, it is always in the platform's interest to combine signals that induce similar beliefs. Threshold systems naturally combine similar signals. When signals are ordered in terms of informativeness, then threshold systems are optimal. Log-concavity of f ensures that lower signals are more informative of $\theta = L$. That is, log-concavity of f ensures that threshold systems are optimal.

Lemma 1. *Suppose that f is log-concave. For any k and any degree of noise σ, the optimal review system with k reviews is a k-threshold system.*

Given the optimality of threshold systems, I now seek to characterize the optimal threshold review system when the level of noise σ is large. In order to simplify notation, I transform signals: $s \mapsto \frac{s}{\sigma}$. This transformation preserves the information at a fixed σ while normalizing the scale of signals, guaranteeing that review systems converge. In what follows I deal exclusively with normalized signals. The distributions of normalized signals in the two states are given by

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

Given a review system r, write the learning efficiency of r when the level of noise is σ as $\nu_f(\sigma; r)$. Similarly, write the relative information of r as $\kappa_f(\sigma; r)$. When it is clear from the context, I drop the dependence on f. As the noise in individual signals σ increases, the distributions of signals in the two states become similar, limiting the information contained in an individual review. In the limit ($\sigma \to \infty$), individual reviews contain no information ($\lim_{\sigma\to\infty} \nu(\sigma; r) = 0$ for all r). However, the relative information of reviews remains well-defined. In order to characterize the optimal review system when signals are imprecise, I explicitly characterize $\lim_{\sigma\to\infty} \kappa(\sigma; r_{\boldsymbol{\tau}})$ for threshold systems $r_{\boldsymbol{\tau}}$. Slightly abusing notation, I write $\kappa(\infty; r_{\boldsymbol{\tau}})$ for this limit:

$$\kappa(\infty; r_{\boldsymbol{\tau}}) := \lim_{\sigma\to\infty} \kappa(\sigma; r_{\boldsymbol{\tau}}) = \lim_{\sigma\to\infty} \frac{\nu(\sigma; r_{\boldsymbol{\tau}})}{\nu(\sigma; r_f)}.$$

If $\kappa(\infty, r_{\boldsymbol{\tau}})$ is small, then reviews from $r_{\boldsymbol{\tau}}$ must be reported much more frequently than full reviews in order for the platform to prefer $r_{\boldsymbol{\tau}}$ when signals are very noisy. On the other hand, if $\kappa(\infty, r_{\boldsymbol{\tau}})$ is large, then even when reviewers are slightly more likely to leave the simpler review $r_{\boldsymbol{\tau}}$, the platform prefers eliciting coarse information.

Before I introduce the result, some intuition is useful. When noise is large, on regions where f is nearly constant, signals are uninformative. This is because the difference between the two densities f_L and f_H is small. However, on regions where f is very curved, even for very noisy environments, signals are very informative. If f is very curved, the full reviews remains informative even when signals are imprecise. In turn, a large curvature in f depresses $\kappa(\infty, r_{\boldsymbol{\tau}})$. The measure of curvature that captures this effect is the *Fisher Information* of f:[5]

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

A similar logic determines the behaviour of $\nu(\sigma; r_{\boldsymbol{\tau}})$ when σ is large. Since the review system is discrete, the effect of a marginal change in σ only impacts reviews at the thresholds $\tau_1, ..., \tau_{k-1}$. Consider a single review $r^i \in \mathcal{R}_{r_{\boldsymbol{\tau}}}$. A marginal decrease in noise increases the probability that r^i is reported in state L on the order of $(f(\tau_i) - f(\tau_{i-1}))$. Symmetrically, the probability that r^i is reported in state H decreases by the same amount. Review r^i becomes more informative as σ decreases if the magnitude of this difference is large. This effect is magnified if r_i is uncommon because the difference in probabilities in the two states is more easily detected.

Assumption 1 guarantees that the Fisher information is well-defined.[6] These effects come together to characterize relative information when signals are noisy:

Lemma 2. *Assume that f satisfies Assumption 1. Let F denote the cumulative distribution function of f. Then, the relative information of $r_{\boldsymbol{\tau}}$ when information is imprecise is given by*

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

with the convention that $f(\infty) = f(-\infty) = 0$, and $F(\infty) = 1, F(-\infty) = 0$.

A key feature of online reviews is that a single review is very noisy. Learning from any review system is slow when reviewers' signals are imprecise. Optimizing the rate of learning is especially relevant in this setting since small sample effects have a minor impact. Lemma 2 computes the performance of threshold systems when reviewers' signals are imprecise.[7] It is proved by showing independently the

[5]The Fisher information appears in other settings with noisy signals. For instance, in Sadzik and Stacchetti (2015), a larger Fisher information corresponds to easier detection of defections in a principal-agent model with frequent, but noisy, signals.

[6]Absolute continuity can be replaced by assuming 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 being integrable, for μ in a neighbourhood of 0.

[7]Lemma 2 can also be extended to non-threshold finite systems at the cost of notation, see Theorem 3.

rate at which $\nu(\sigma; r_f)$ and $\nu(\sigma; r_{\boldsymbol{\tau}})$ go to zero as a function of the noise in the taste idiosyncrasies. Lemmas 1 and 2 together characterize the optimal review system with k reviews.

Theorem 2. *Assume that f is log-concave and satisfies Assumption 1. The optimal review system with k reviews when noise is large is the k-threshold system $r_{\boldsymbol{\tau}}$ with thresholds that maximize*

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

When noise is large, this system is preferred to full review system r_f if

$$\frac{\sum_{i=1}^{k} \frac{\left(f(\tau_i)-f(\tau_{i-1})\right)^2}{F(\tau_i)-F(\tau_{i-1})}}{I_f} > \frac{p_{r_f}}{p_{r_{\boldsymbol{\tau}}}}.$$

The full review system is preferred to $r_{\boldsymbol{\tau}}$ if the reverse inequality holds.

Theorem 2 provides a straightforward objective for the platform to maximize.[8] Moreover, by characterizing the optimal review system with k reviews, Theorem 2 implicitly characterizes the optimal review system. To determine the optimal review system, the platform first optimizes for a fixed number of reviews and then optimizes over the number of reviews. Importantly, it shows that even when noise is large, the choice of thresholds has important implications for the relative information of review systems. The next section applies these results to analyze the impact of one key characteristic of taste shocks: taste heterogeneity.

4 Application: Taste Heterogeneity and Binary Review Systems

As mentioned, the idiosyncratic shocks ϵ_i are due to the idiosyncratic preferences of reviewers. The distribution of these preference shocks varies across different settings and has important implications for optimal review systems. For some products like movies, preferences are dispersed because individual taste matters to a large degree.

[8] The optimal $\boldsymbol{\tau}^*$ depends on σ. However, it is sufficient to set $\boldsymbol{\tau}^* = \lim_{\sigma\to\infty} \boldsymbol{\tau}^*(\sigma)$ as $\lim_{\sigma\to\infty} \frac{\nu(\sigma; r_{\boldsymbol{\tau}^*(\sigma)})}{\nu(\sigma; r_{\boldsymbol{\tau}^*})} = 1$ under Assumption 1.

For other products like toasters, preferences are much less dispersed.[9] This dispersion, or *heterogeneity*, in reviewers' taste idiosyncrasies has important implications for optimal review systems. In this section I use Theorem 2 to characterize optimal binary review systems as a function of taste heterogeneity.

4.1 Symmetric Heterogeneity

In order to understand how taste heterogeneity affects review systems, I restrict attention to a one-parameter family of distributions whose parameter controls the degree of taste heterogeneity within the population of reviewers:

$$f_\alpha(s) \propto \exp(-|s|^\alpha).$$

When preferences are dispersed, the spread of reviewers' signals is large, regardless of noise. This corresponds to large tails in the distribution of taste shocks. The parameter α controls the tails of the distribution. When α is large, f_α has small tails, so that the population of reviewers is homogeneous. Henceforth, I refer to α as homogeneity.

Remark 3. For $\alpha \geq 1$, $f_\alpha(s)$ is log-concave, so threshold systems are optimal. This family is referred to as the Generalized Normal Distributions and incorporates three well-known distributions as special cases: f_1 is the Laplace distribution (large heterogeneity); f_2 is the Normal distribution (moderate homogeneity); and as $\alpha \to \infty$, f_α converges almost everywhere to the uniform distribution on $(-1, 1)$ (perfect homogeneity).

I first show how relative information varies with homogeneity when the platform asks reviewers if their signal is positive or negative. When tastes are very heterogeneous, r_0 performs well: relative information is close to 1. However, as homogeneity grows, relative information decreases to zero: information loss is unbounded.

Proposition 1. *Consider relative information under the threshold 0, $\kappa_{f_\alpha}(\infty; r_0)$:*

[9] The classification of "search" and "experience" goods provides a possible way to distinguish between goods with low and high taste dispersion (heterogeneity). First introduced in Nelson (1970), search goods' attributes are well-defined and easily found, whereas experience goods 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 important.

(i) *When reviewers are very heterogeneous, the symmetric binary review contains the same information as the full review:* $\kappa_{f_1}(\infty; r_0) = 1$*; and*

(ii) *As reviewer taste homogeneity increases,* r_0 *performs worse relative to the full review:* $\kappa_{f_\alpha}(\infty; r_0)$ *is decreasing in* $\alpha \geq 1$*. Moreover,* r_0 *performs arbitrarily poorly when reviewers are sufficiently homogeneous:* $\lim_{\alpha\to\infty} \kappa_{f_\alpha}(\infty; r_0) = 0$.

The relative information of the naive binary review r_0 is decreasing in taste homogeneity. That is, as reviewer taste homogeneity increases, reviewers must submit binary reviewers more frequently for r_0 to continue to be preferred to the full review. The intuition for this result is as follows. When reviewers are homogeneous, there is small dispersion in signals. This means that there are both many uninformative signals *and* some very informative signals.[10] The symmetric review r_0 mixes very informative signals with very uninformative signals, limiting the information contained in either review.

This discussion highlights the central tension in choosing an asymmetric binary review: extracting more information from one review while extracting less from the other. If the platform uses review system r_τ for $\tau > 0$ instead of r_0, then (i) the negative review contains a mass of positive signals; but (ii) the positive review contains a smaller mass of uninformative signals. Effect (i) makes the review system less informative, while (ii) improves precision. If the second effect outweighs the first, then it is optimal to use an asymmetric review.

When reviewers' tastes are sufficiently heterogeneous, the cost of decreasing the informativeness of one review is larger than the benefit from making the other review more informative. However, when reviewer homogeneity is large, enough information can be extracted from the informative review to outweigh the cost: the optimal binary review is asymmetric. Importantly, the optimal threshold limits learning loss under the binary review: relative information is bounded below by 1/3.25. That is, if reviewers are 3.25 times as likely to submit a binary review than the fully detailed review, the optimal binary system is preferred, regardless of the degree of homogeneity. This result, together with Proposition 1, highlights the importance of the analysis: the design of the binary review has important implications for its performance.

[10] I.e., the variance of induced posteriors is large.

In what follows, let $\tau^*(\alpha)$ be the non-negative optimal threshold. Since f_α is symmetric, there is also a non-positive maximizer $-\tau^*(\alpha)$. In the statement of Proposition 2, all statements are written in terms of the non-negative optimal threshold for clarity. Comparable statements hold for the non-positive optimal threshold.[11]

Proposition 2. *For $\alpha \geq 1$, there is a unique optimal binary review system with non-negative threshold $\tau^*(\alpha)$. It has the following properties:*

(i) The symmetric binary review is the optimal binary review if and only if reviewers are sufficiently heterogeneous: if $\alpha \leq 2$, $\tau^*(\alpha) = 0$, while if $\alpha > 2$, $\tau^*(\alpha) \neq 0$;

(ii) When reviewers are sufficiently homogeneous, the optimal binary review is very asymmetric: $\lim_{\alpha\to\infty} \tau^*(\alpha) = 1$; and

(iii) The relative information of the optimal binary review is bounded: $\kappa_{f_\alpha}(\infty; r_{\tau^*(\alpha)}) > 1/3.25$ for all α.

The normal distribution corresponds to the degree of taste homogeneity at which the optimal binary review system ceases to be symmetric. 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 horrible" (negative τ^*).

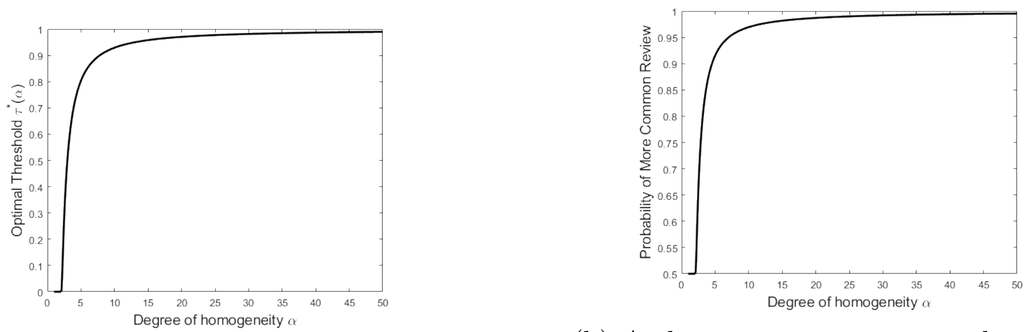
Proposition 3 shows comparative statics for $\tau^*(\alpha)$: both $\tau^*(\alpha)$ and $F_\alpha(\tau^*(\alpha))$ are increasing in α. Importantly, this means that the probability of the more common review increases in homogeneity, *regardless of the state*. Consider $r_{\tau^*(\alpha)}$, the optimal binary review with a positive threshold, for large α. Even when $\theta = H$, the majority of reviews are negative. The share of negative reviews increases to 1 as $\alpha \to \infty$. This suggests that review systems where the vast majority of reviewers leave the same review are optimal when reviewers are sufficiently homogeneous. These results are summarized in Fig. 1. In particular, even for relatively small degrees of homogeneity, the optimal review system is very asymmetric.

Proposition 3. *Consider the non-negative optimal threshold $\tau^*(\alpha)$. It has the following properties:*

[11] The next section shows that introducing (arbitrarily small) asymmetry in f_α breaks this multiplicity.

(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* $\alpha \geq 2$*, with* $\lim_{\alpha\to\infty} F_\alpha(\tau^*(\alpha)) = 1$.



(a) As homogeneity increases, the optimal review becomes more asymmetric.

(b) As homogeneity increases, the probability of the more common review increases to 1.

Figure 1: Properties of the optimal binary review system, for different degrees of homogeneity α.

4.2 Asymmetric Heterogeneity

The multiplicity of optimal reviews disappears when asymmetry in positive and negative taste homogeneity is introduced. When heterogeneity is asymmetric, the optimal binary review system's threshold is on the side of the distribution that features larger taste homogeneity. Let α^* be the value for which $f_\alpha(0)$ is maximized.[12]

Proposition 4. *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 $\alpha^* \leq \alpha_L, \alpha_H$*. 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.*

Fig. 2 highlights the intuition behind Proposition 4. Consider a positive threshold (dotted line). At the negative threshold that yields the same value of the density

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

(dashed line), 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 feasible positive candidate, so that the optimal threshold is negative.[13]

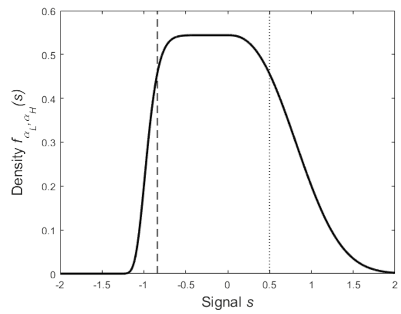


Figure 2: Asymmetric distribution of taste shocks f_{α_L,α_H}, for $\alpha_L = 10, \alpha_H = 2.5$.

Apart from breaking multiplicity, this is an important setting because asymmetric taste homogeneity is common. For instance, consider ride-sharing services. In this setting, reviewers agree on bad experiences: no reviewer wants to be late or get into an accident. On the other hand, consumers disagree on what makes a good experience: some (not all) people like having a discussion with their driver; some (not all) like music to be played.[14] Proposition 4 justifies why many platforms exhibit many more positive reviews than negative reviews (Hu, Zhang, and Pavlou 2009). If reviewers agree on negative experiences more than positive experiences, the platform *should* isolate extremely negative experiences. Summarizing in some settings the optimal review is positively skewed, even when the product is of low quality.

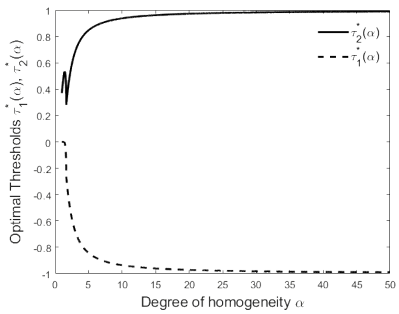
4.3 Beyond Binary Reviews

Similar qualitative results apply to optimal finite review systems with k components, where $k > 2$. Fig. 3 exhibits the optimal three-bin system for varying degrees of homogeneity α. In particular, the optimal asymmetry of the system is increasing in the degree of homogeneity. An important difference from the binary system is that both extreme positive and negative signals can be isolated. When homogeneity is large, the platform "throws out" many signals. The middle review provides no

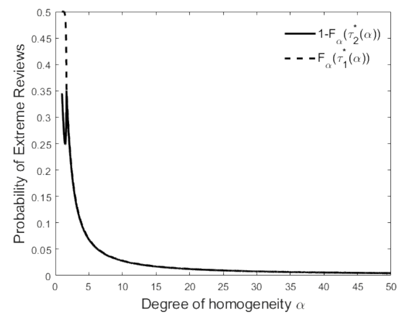
[13] An asymmetry is asymmetry in the variance of positive and negative taste shocks. If the variance of signals for positive taste shocks was larger than for negative taste shocks, then by a similar logic the optimal threshold is negative.

[14] For search goods, larger negative homogeneity is expected: most reviewers are in agreement about whether a good doesn't meet expectations, but may disagree about how it over-performs.

information to the platform (since it is equally likely in each state). Its presence is important, however, because it allows the platform to skim uninformative reviews in order to isolate very informative positive and negative signals.



(a) The optimal thresholds become more extreme as homogeneity grows.



(b) The probability of the highest and lowest review shrinks as homogeneity grows.

Figure 3: Properties of the optimal three-bin review system, for different degrees of homogeneity α.

5 Extensions

5.1 Signal-Dependent Reporting Rates

In practice, a reviewer's experience is likely to affect the probability that she submits a review. For example, a reviewer with an extreme signal is more likely to leave a review.[15] In some contexts reviewers may report positive and negative experiences at different rates. In this section I show how to incorporate signal-dependent reporting rates into my framework. While the qualitative insights from previous sections remain, several new forces emerge.

Suppose now, in addition to the cost of submitting the review $c(r)$, there is a benefit $b(s)$ of reporting signal s (with $b(s) \leq c(r)$). This reflects an expressive component to submitting reviews (Lafky 2014). A reviewer now receives utility $w_i + b(s) - c(r)$ from submitting review r when her signal is s. In this setting, the probability that a reviewer leaves a review is a function of both the review system and her signal: $p(r, s) := \mathbb{P}(W_i \geq c(r) - b(s))$.

[15] 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.

As before, the difference in the distribution of reviews across the two states determines the review system's performance. Define the distribution $\gamma_L^{(p,r)}$ of reviews in state L (implicitly dependent on σ) as

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

Similarly define the distribution $\gamma_H^{(p,r)}$ of reviews in state H. As before, these measures determine the rate at which the platform learns the state. Importantly, $p(r,s)$ is no longer separable from these measures. Define $\rho(\sigma; r, p)$ as the analog of $p_r \nu(\sigma; r)$:

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

There are two significant differences between ρ and $p_r \nu(r)$. First, different signals receive different weight because $p(r, \cdot)$ varies with the signal realization. Second, "null reviews" (reviews that are not submitted) may now be informative of the state. The probability that a signal is not submitted when the state is L is $\gamma_L^{(1-p,r)}(\mathbb{R})$. If this is larger than $\gamma_H^{(1-p,r)}(\mathbb{R})$, then a null signal is indicative of state L. Despite these changes, Theorem 1 generalizes to this setting, where $\rho(\sigma; r, p)$ replaces $p_r \nu(\sigma; r)$.

Lemma 3. *Fix two review systems r, r' such that $\rho(\sigma; r, p) > \rho(\sigma; r', p)$. For any finite action set $\mathcal{A}$ and utility function u such that the platform's decision problem is not trivial ($\operatorname{argmax}_{a\in\mathcal{A}} u(a, L) \cap \operatorname{argmax}_{a\in\mathcal{A}} u(a, H) = \emptyset$), there exists an $\overline{N}$ such that for all $N \geq \overline{N}$, such that $u^*(r, N) > u^*(r', N)$.*

Without additional structure on the probabilities of review p, little more can be said about the comparison between review systems. With arbitrary reporting rates p, the underlying distribution of signals that different review systems draw from differ. This means that anything can happen. In order to gain structure, I restrict attention to reporting rates p that are separable between the impact of the review function and the signal realization.

Definition 4. 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 $q : \mathbb{R} \to [0, 1]$.

If the willingness to review W_i are distributed exponentially, then the resulting propensity to review is separable (following a normalization so that $q(s) \leq 1$).[16] To understand the new forces that emerge when reporting rates are signal-dependent, I now develop an analogue of Theorem 2 in the case that propensities to review 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 $q \cdot f$ is not probability distribution (unless $q(s) = 1$ almost everywhere). Similarly, given a review system r, for each review $\tilde{r} \in \mathcal{R}_r$, define

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

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

Theorem 3. *Assume that f is log-concave and satisfies Assumption 1 and $p(r, s) = p_r q(s)$. Let $\ell_f := \frac{f'}{f}$ denote the linear score of f. The generalization of relative learning for a finite review system r is*

$$\frac{p_{r_f}}{p_r} \cdot \lim_{\sigma \to \infty} \frac{\rho(\sigma; r, p_r \cdot q)}{\rho(\sigma; r_f, p_{r_f} \cdot q)} = \frac{\sum_{\tilde{r} \in \mathcal{R}_r} \frac{E(\ell_f; \tilde{r})^2}{E(1; \tilde{r})} + \frac{p_r E(\ell_f)^2}{1 - p_r E(1)}}{E\left(\ell_f^2\right) + \frac{p_{r_f} E(\ell_f)^2}{1 - p_{r_f} E(1)}} \in [0, 1]. \tag{2}$$

The optimal k component review is a k-threshold system whose thresholds maximize (2).

The left-hand side of (2) is the generalization of $\kappa(\infty, r)$. In the case that $q \equiv 1$, Theorem 3 generalizes Lemma 2 to non-threshold finite reviews. The sum in the

[16] Importantly, q depends on the reviewer's signal and *not* their taste shock ϵ_i. This latter case would simply reduce to the previous setting.

numerator and the first term of the denominator are the direct generalizations of the numerator and denominator from Lemma 2.

The impact of null reviews is the second term of the numerator and denominator. Importantly, the information contained in null reviews depends on the rate at which a review is submitted. To illustrate, consider Fig. 4. The top two panels show the distribution of submitted reviews. As p_r increases, these distributions are scaled. However, for null reviews (shown in the bottom panels), this is not the case. As 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 asymmetry due to q). Hence, null reviews become less informative as p_r decreases. In the case that $p_r \approx 0$, null reviews are uninformative. In this setting, signal-independent reporting rates approximate the more general model apart from a change in measure.

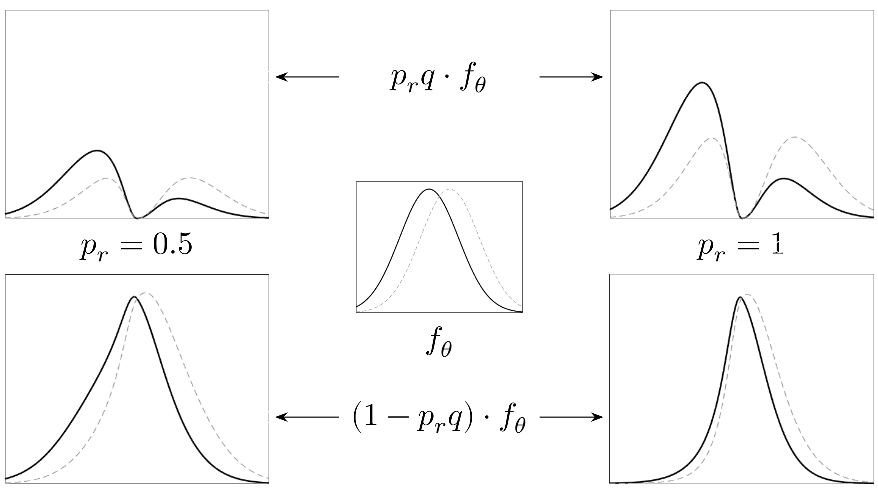


Figure 4: Distributions of reviews for $f_\theta(s)$ normal, with $\frac{\mu}{\sigma} = 0.25$, and $q(s) = 1 - \frac{1}{1+2(x^2+4\mathbb{1}\{x\leq 0\}x^2)}$ and for $p_r = 0.5$ (left), 1 (right). The top panels show the distribution of submitted reviews, and the bottom panels show the distribution of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

5.2 Comparing Numbers of Reviews versus Probability of Reviewing

An alternative to comparing rates of reporting is to directly compare n_r reviews from system r with $n_{r'}$ reviews from system r'.[17] This comparison falls to consumers who must decide between multiple types of reviews on a website: should they read written

[17]This is the traditional approach taken in the literature (Chernoff 1952; Moscarini and Smith 2002; Mu et al. 2021).

reviews, or instead consider the 5-star rating? The consumer's preferences and the platform's preferences over review systems are related, they are not identical.

Consider two different review systems, r and r', with $\nu(r) > \nu(r')$ and $p_r < p_{r'}$. The platform must choose a review system before information is collected. This means that it must consider the possibility of not receiving any (or receiving very few) reviews. This happens with higher probability under review system r. However, on average, there are many reviews of type r submitted. For a consumer deciding between Np_r reviews from system r and $Np_{r'}$ reviews from system r', the risk of seeing few reviews is not a concern. This means that varying p_r has a larger effect on the platform's preferences over review systems than it does on the consumer's. When reviewers' information is imprecise, this difference disappears and the two orders agree. Appendix B formalizes the model and the differences, and shows these relationships.

5.3 Multi-Dimensional Information

In practice, even simple products like toasters have multiple dimensions of quality. While my results do not directly speak to this setting, they do suggest a naive approach. It is always possible for the platform to isolate different dimensions of product quality, and ask a simple question along each dimension. My results suggest that, because of the performance of binary systems in the one-dimensional case, even this naive approach performs well. An interesting question for future work is when it is optimal to isolate dimensions in this way, and when it is better to combine information across dimensions.

5.4 More than Two States

When there are more than two states of the world, my results can be extended in the standard way. First, partition states by the optimal action in that state. Then, the rate of learning for the platform is determined by the two states across elements of the partition that are hardest to distinguish. In the case of binary action environments (fire/keep on, recommend/do not recommend, or when flagging a problem), the platform's problem becomes one of differentiating two states, mirroring my specification. An important difference is that the optimal review system in this setting depends on the platform's decision problem, because its problem determines which two states are must be distinguished. In settings with a richer action space, more care must be taken. It is possible that a review system that distinguishes two elements of the partition well might poorly distinguishes a third. In this case, more complicated review systems may be preferred.

6 Conclusion

My results show how a principal optimally designs review systems if it has complete control over how reviewers submit reviews. I abstract from the implementation of review systems throughout my analysis. In practice, controlling how reviewers leave reviews is difficult. Common review systems, like 5-star reviews, are likely difficult to design because of reviewers' extensive experience with similar systems: reviewers are likely to default to their own, private, thresholds.[18] Binary systems, because of their simplicity, are easier to carefully design. In practice, many different binary reviews are used—"did you like the product?", "would you buy the product again?", "would you recommend the product to a friend/colleague?"—all of which induce different thresholds. If implementation of complex review systems is difficult, then there is an additional benefit to using simple review systems.

Whenever a principal elicits information from agents, it must consider the trade-off between the quality of information that it elicits and the burden it places on agents. If the principal asks for too much information, agents may choose to opt-out and not provide any information. My model characterizes this trade-off. I show that learning is sensitive to the design of review systems and that the principal must carefully consider reviewers' information. When the platform optimally elicits information even simple review systems perform well. However, the optimal review system depends on the information of reviews. If reviewers are sufficiently heterogeneous, the symmetric "good" or "bad" review systems performs well. On the other hand, if reviewers' tastes are homogeneous, this naive binary review system performs poorly. Instead, the optimal binary review system is asymmetric: the platform isolates informative signals in one review at the cost of the other review.

[18] This is a major mechanism driving the results of Botelho et al. (2025).

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 (cit. on pp. 4, 7).

Bean, A. G. and M. J. Roszkowski (1995): "The Long and Short of It: When Does Questionnaire Length Affect Response Rate?" *Marketing Research* 7(1), pp. 21–26 (cit. on p. 1).

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

Blackwell, D. (1951): "Comparison of Experiments". *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 93–102 (cit. on p. 5).

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 (cit. on pp. 5, 26).

Che, Y.-K. and J. Hörner (2018): "Recommender systems as mechanisms for social learning". *The Quarterly Journal of Economics* 133(2), pp. 871–925 (cit. on pp. 4, 7).

Che, Y.-K., K. Kim, and W. Zhong (2024): "Statistical Discrimination in Ratings-Guided Markets". *Working Paper* (cit. on p. 4).

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* (cit. on p. 4).

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 (cit. on pp. 5, 10, 24, 57).

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 (cit. on p. 5).

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 (cit. on p. 5).

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 (cit. on p. 4).

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 (cit. on p. 4).

Frick, M., R. Iijima, and Y. Ishii (2023): “Learning Efficiency of Multiagent Information Structures”. *Journal of Political Economy* 131(12), pp. 3377–3414 (cit. on p. 5).

— (2024a): “Monitoring with rich data”. *arXiv preprint arXiv:2312.16789* (cit. on p. 5).

— (2024b): “Welfare comparisons for biased learning”. *American Economic Review* 114(6), pp. 1612–1649 (cit. on p. 5).

Garg, N. and R. Johari (2019): “Designing optimal binary rating systems”. *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1930–1939 (cit. on p. 4).

Hann-Caruthers, W., V. V. Martynov, and O. Tamuz (2018): “The speed of sequential asymptotic learning”. *Journal of Economic Theory* 173, pp. 383–409 (cit. on p. 5).

Harel, M., E. Mossel, P. Strack, and O. Tamuz (2021): “Rational groupthink”. *The Quarterly Journal of Economics* 136(1), pp. 621–668 (cit. on p. 5).

Hu, N., P. A. Pavlou, and J. Zhang (2017): “On self-selection biases in online product reviews”. *MIS quarterly* 41(2), pp. 449–475 (cit. on pp. 4, 8).

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 (cit. on pp. 3, 4, 20).

Ifrach, B., C. Maglaras, M. Scarsini, and A. Zseleva (2019): “Bayesian social learning from consumer reviews”. *Operations Research* 67(5), pp. 1209–1221 (cit. on p. 3).

Lafky, J. (2014): “Why do people rate? Theory and evidence on online ratings”. *Games and Economic Behavior* 87, pp. 554–570 (cit. on pp. 8, 21).

Magnani, M. (2020): “The economic and behavioral consequences of online user reviews”. *Journal of Economic Surveys* 34(2), pp. 263–292 (cit. on pp. 5, 16).

Moscarini, G. and L. Smith (2002): "The law of large demand for information". *Econometrica* 70(6), pp. 2351–2366 (cit. on pp. 5, 24).

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 (cit. on pp. 5, 24, 57).

Nelson, P. (1970): "Information and consumer behavior". *Journal of Political Economy* 78(2), pp. 311–329 (cit. on p. 16).

Oprea, R. (2020): "What makes a rule complex?" *American economic review* 110(12), pp. 3913–3951 (cit. on p. 6).

Puri, I. (2022): "Simplicity and risk". *The Journal of Finance* (cit. on p. 6).

Rosenberg, D. and N. Vieille (2019): "On the efficiency of social learning". *Econometrica* 87(6), pp. 2141–2168 (cit. on p. 5).

Sadzik, T. and E. Stacchetti (2015): "Agency models with frequent actions". *Econometrica* 83(1), pp. 193–237 (cit. on p. 14).

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 (cit. on pp. 5, 10, 29).

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 (cit. on p. 6).

A Proofs

All proofs for results in the main text are proved in this appendix.

A.1 Proofs for Section 3

In this appendix I provide the proof of the major results 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 $\mathcal{R}_r \cup \{\emptyset\}$ representing the distribution of reviews in state L. That is,

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

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,

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

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

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

Hence,

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

Concluding,

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

which is what we set out to show. $\square$

The proof of Lemma 1 requires additional mechanics. I first prove a similar result in posterior space, and then show that the result applies in signal space when f is log-concave.

Let $\overline{f}(s) = \frac{1}{2}(f_L(s) + f_H(s))$ 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 system is an expectation that depends on the posterior and the ex ante distribution:

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

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 $\overline{f}$.[19] That is,

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

The learning efficiency of the full review system can then be written as

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

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 almost all signals $s, s' \in \mathbb{R}$ such that $\pi(s) = \pi(s')$, $r(s) = r(s')$. Notice that if γ_π has no atoms, then it is without loss to restrict attention to consistent review systems.

The representation of $\nu(r_f)$ in terms of the distribution of posteriors extends to arbitrary 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.

[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.

Lemma A.1. Let r be any consistent review system. Then,

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

Proof. I show this in the case that r is a finite review system, and when γ^r has full support on $\mathcal{R}_r$, 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 $\nu(r)$ can be written as as

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

If $\pi(s)$ is injective, then by straightforward computation it is easy to show that $\frac{1}{1-\pi}f_H(s) = 2\gamma_\pi(d\pi(s))$ and $f_H(s) = 2(1-\pi(s))\gamma_\pi(d\pi(s))$. In general, as long as r is consistent, even if $\pi(s)$ is not injective, then for each λ

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

From here, the conclusion is reached by minimizing over λ. □

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 is from the pre-evaluation of $2\pi^{\lambda}(1-\pi)^{1-\lambda}$. Since, for any λ, this function is concave, information loss is exactly due to Jensen's inequality.

If r is a consistent review system with k components, it is a partition of the space $[0,1]$ into k sections. Let $\{\Pi_1, ..., \Pi_k\}$ be the partition of the interval. Then $\nu(r)$ can be rewritten as

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

In order to show that threshold rules are optimal in signal space, I show that they are optimal in threshold space.

Definition A.1. A (posterior) review system r is called a *k-threshold* review system if $r^{-1}(\tilde{r})$ is convex (except perhaps on a set of measure zero) for each $\tilde{r} \in \mathcal{R}_r$.

Optimal review systems are necessarily threshold rules in posterior space.

Lemma A.2. Let r be an optimal (posterior) review system with k reviews. Then r is k-threshold review system.

Proof. I show that if r is a finite review with k reviews that is not such a threshold rule, then there exists a review with k reviews 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 reviews 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 reviews. 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

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

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 $\nu(r) < \nu(r')$. $\square$

Proof of Lemma 1. If f is log-concave, then for any level of noise $\frac{f_H}{f_L}$ is monotonically increasing. Consider the optimal posterior review system r_π. The optimal review system in signal space is given by $r_\pi \circ \pi(s)$. Since π is monotonic (because $\frac{f_H}{f_L}$ is increasing), and r_π is a k-threshold system by Lemma A.2, it must be the case that $r_\pi \circ \pi$ is a threshold system. $\square$

Proof of Lemma 2. The proof follows from an application to Theorem 3, to 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. $\square$

Proof of Theorem 2. This follows from an immediate application of Lemmas 1 and 2. $\square$

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 A.2. 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 A.3. 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. 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 $\kappa_{f_\alpha}(\infty; r_\tau)$ by $\kappa(\alpha;\tau)$. This is just done to condense notation: κ is a function of α and the threshold τ of the review system.

Proof of Proposition 1. First, a simple computation shows that $\kappa(\alpha;0) = \frac{1}{\Gamma(\frac{1}{\alpha})\Gamma(2-\frac{1}{\alpha})}$. 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 $\left(\frac{1}{\kappa(\alpha;0)}\right)'$ denote $\frac{\partial}{\partial\alpha}\frac{1}{\kappa(\alpha;0)}$. A straightforward computation yields that

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

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

It is left to show that $\lim_{\alpha\to\infty}\kappa(\alpha;0) = 0$. 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

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

That is, $\lim_{\alpha\to\infty} \alpha\frac{\kappa'(\alpha;0)}{\kappa(\alpha;0)} = -1$. This means that $\log(\kappa(\alpha;0))' \to -\infty$, and hence $\kappa(\alpha;0) \to 0$. This shows that $\kappa(\alpha;0)$ converges to zero, completing the proof. $\square$

Proof of Proposition 2. As noted, I restrict attention to $\tau \in [0,\infty)$ for the proof. The analogous proofs apply for $\tau \leq 0$.

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

$$g_\alpha(\tau) := \frac{f_\alpha(\tau)^2}{F_\alpha(\tau)(1-F_\alpha(\tau))} \tag{3}$$

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 A.4. 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. The derivative of (3) 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

$$2f'(\tau) = f(\tau)^2\left(\frac{1}{F(\tau)} - \frac{1}{1-F(\tau)}\right). \tag{4}$$

For any $\alpha > 1$, 0 is a solution to (4), 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. $\square$

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 A.5. The following are true:

(i) For $\alpha > 2$, there is exactly one value of $\tau \in (0, \infty)$ that satisfies (4) and it is a local maximum of $g_\alpha(\tau)$; and

(ii) For $\alpha \leq 2$, no value of $\tau \in (0, \infty)$ satisfies (4).

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

Case 1: $\alpha > 2$.
Straightforward algebra allows (4) to be rewritten as

$$\frac{2f'(\tau)}{1 - 2F(\tau)} = \frac{f^2(\tau)}{F(\tau)(1 - F(\tau))}. \tag{5}$$

Notice that the right-hand side of (5) 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

$$2f''(\tau)(1 - 2F(\tau)) + 4f'(\tau)f(\tau) \tag{6}$$

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 A.4), $g(\tau)$ 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.[20] 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 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

[20] 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.

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. $\square$

Now that I have showed that τ^* exists and is unique, it is sufficient to look at the solution to the first order condition (4) 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 A.5.

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

$$
-2\tau^{\alpha-1} + e^{-\tau^\alpha} \left[\frac{1}{\Gamma(1/\alpha, \tau^\alpha)} - \frac{1}{\Gamma(1/\alpha) + \gamma(1/\alpha, \tau^\alpha)} \right]. \tag{7}
$$

That is, $g'_\alpha(\tau) > 0$ exactly when (7) 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 also returns a value that is less than this limit. This shows the final claim, and completes the proof of Proposition 2. □

Proof of Proposition 3. In order to prove Proposition 3(i), I prove the following, stronger, lemma.

Lemma A.6. $\tau^*(\alpha)^\alpha$ is strictly increasing in α for $\alpha \geq 2$.

Proof. 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

$$\frac{\partial}{\partial\alpha}\left[-2\frac{x}{x^{1/\alpha}}e^x+\left[\frac{1}{\Gamma(1/\alpha,x)}-\frac{1}{\Gamma(1/\alpha)+\gamma(1/\alpha,x)}\right]\right] \tag{8}$$

must be determined at the optimal value of $x^*(\alpha)=\tau^*(\alpha)^\alpha$. Notice that (8) 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 does not influence sign of this object). This just follows from the envelope theorem. So, we must show that (8) 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 (8) is the opposite of the sign of

$$b_\beta(x) := 2\log(x)\frac{x}{x^\beta}e^x$$

$$-\frac{\int_x^\infty y^{\beta-1}\log(y)e^{-y}dy}{\Gamma(\beta,x)^2}+\frac{\int_0^\infty y^{\beta-1}\log(y)e^{-y}dy+\int_0^x y^{\beta-1}\log(y)e^{-y}dy}{(\Gamma(\beta)+\gamma(\beta,x))^2}.$$

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

$$2\log(x)\frac{x}{x^\beta}e^x-\frac{\int_x^\infty \log(y)d\lambda_\beta(y)}{\Gamma(\beta,x)^2}+\frac{\int_0^\infty \log(y)d\lambda_\beta(y)+\int_0^x \log(y)d\lambda_\beta(y)}{(\Gamma(\beta)+\gamma(\beta,x))^2}. \tag{9}$$

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 (8), 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 (9) at $x^*(\beta)$ (I write x instead of $x^*(\beta)$ for conciseness) as

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

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 (9) 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 3(i) follows as an immediate corollary of this lemma, since $\tau^*(\alpha) < 1$ for all α.

Proof of Proposition 3(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 A.6. 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 A.6 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

$$\frac{f_\alpha(\tau^*(\alpha))^2}{F_\alpha(\tau^*(\alpha))(1 - F_\alpha(\tau^*(\alpha)))} \tag{10}$$

is increasing in α, then the claim holds in this case. This is because, for (10) 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 is that

$$\frac{\left(\frac{\alpha}{\Gamma(1/\alpha)}\right)^2}{(1 + P(1/\alpha, \tau))(1 - P(1/\alpha, \tau))} \tag{11}$$

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 (10) at $\tau^*(\alpha)$ for some α. If (11) 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.

I now show that (11) is indeed increasing in α. Perform the switch to $\beta := 1/\alpha$, as the in proof of Lemma A.6. 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 A.6. 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

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

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

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

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 4. 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.[21] Explicitly, given a $\tau > 0$, we have that $-\tau^{\alpha/\alpha'} < -\tau$ (for $\tau < 1$, which is again be optimal here) has 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

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

[21] 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 does not hold for $\tau = 0$).

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

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

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

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

$$-\frac{1}{\beta} < \frac{\int_x^{\infty} \log(y) d\lambda_{\beta}(y)}{\int_x^{\infty} d\lambda_{\beta}(y)} \tag{12}$$

where the $d\lambda_{\beta}(y)$ notation is the same as in Lemma A.6. Notice first that the right-hand side of (12) 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 increases. 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 (12) holds for $x^* := 0.001$, then the (sub)claim will be proved. From here, several straightforward computations prove the claim:

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

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

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

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 $\beta \in \left(0, \frac{1}{\alpha^*}\right]$ it is the case that (12) 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

Proof of Lemma 3. This is a straightforward application of Theorem 1. □

Proof of Theorem 3. The statement holds for general p, except for the result that the value is contained in $[0, 1]$. The object of interest is

$$\lim_{\sigma \to \infty} \frac{\rho(\sigma; r, p)}{\rho(\sigma; r_f, p)}.$$

In order to simplify notation, I consider an equivalent limit in terms of μ. When signals are normalized, $f_L(s) = f\left(s + \frac{\mu}{\sigma}\right)$. This means that taking the limit as $\sigma \to \infty$ holding μ fixed is equivalent to taking the limit as $\mu \to 0$ while holding σ fixed. The notation is easier holding σ fixed, and so I consider everything in terms of the limit as $\mu \to 0$. Since the fixed value of σ is irrelevant, I set $\sigma = 1$.

First, define

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

Let $\rho(\mu; r_f, p, \lambda)$ denote the value of $\rho(\mu; r_f, p)$ evaluated at a fixed (but not necessarily optimal) λ. That is, $\rho(\mu; r_f, p) = \max_{\lambda\in[0,1]} \rho(\mu; r_f, p, \lambda)$. It can easily be seen that $\rho(\mu; r_f, p), \rho(\mu; r, p) \to 0$. Hence, l'Hopital's rule must be applied. As is shown below, it is also the case that $\frac{\partial}{\partial\mu}\rho(\mu; r_f, p), \frac{\partial}{\partial\mu}\rho(\mu; r_f, p) \to 0$. Hence, both the first and second derivative of the numerator and denominator must be computed.

Step 1: Denominator r_f. The regularity assumptions on f mean that, for a fixed λ, this is given by

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

Notice that because of the envelope theorem, the derivative with respect to λ is equal to zero. Consider the limit as $\mu \to 0$. First, $f(s,\mu,\lambda) \to f(s)$. Hence, the limit is equal to

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

This is equal to zero since $\ell_f(s) f(s) = f'(s)$, and $\int f'(s) ds = 0$ (because the expectation of the linear score function is zero). This means that the second derivative must be considered.[22] Again, the derivative with respect to λ is zero. This means that λ may be fixed, and the optimal value of λ derived at the end of the computation.

To simplify the null review notation, let $H_+(\mu, n) := \int (1 - p(r_f, s)) f(s+\mu) ds$

[22] This is reliant on the numerator converging to zero at the same rate. I show this below.

(where n refers to "null"), and similarly $H_-(\mu,n)$. In a similar vein, let $H'_+(\mu,n) := \int (1-p(r_f,s))f'(s+\mu)ds$. Similarly define $H'_-(\mu,n)$ (and $H''_\pm(\mu,n)$). Computing,

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

The limit is given by

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

As the anti-derivative of $f''(s)$ is $f'(s)$, it is the case that $\int f''(s)ds = 0$. Similarly $\int f'(s)ds = 0$, and $\int f(s)ds = 1$. This shows that this value is equal to the denominator in the statement of the proposition when $\lambda = \frac{1}{2}$. As $\lambda = \frac{1}{2}$ maximizes $\lambda(1-\lambda)$, it is the maximizer of this object, and so is the optimal λ.

This completes the derivations for the denominator. I now turn attention to the numerator. The computations are similar to those for null reviews above.

Step 2: Numerator r. The terms in the numerator are of the following form. Define

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

Similarly, for null reviews define

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

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

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

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

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

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 $p(r,s)$ and not $1-p(r,s)$, 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

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

As in the case of full reviews, the variation in λ does not contribute and so λ may be considered point-wise. In particular, this means that the second derivative must be considered. Computing,

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

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 the variation in λ does not contribute. This means that

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

All that is left is to collect terms and notice again tjat $\int f''(s)ds = 0$:

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

Observe that $\lambda = \frac{1}{2}$ is again the maximizer. This finishes the computation for the numerator. Finally, putting this result together with the computations from the denominator above proves the claim. $\square$

B 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 Lemma 2 applies again.

Lemma B.1. 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,

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

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

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 B.1. 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. $\square$

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 values 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.