

Benkert, Jean-Michel; Schmutzler, Armin

Working Paper

A theory of likes

Working Paper, No. 448

Provided in Cooperation with:

Department of Economics, University of Zurich

Suggested Citation: Benkert, Jean-Michel; Schmutzler, Armin (2026) : A theory of likes, Working Paper, No. 448, University of Zurich, Department of Economics, Zurich, <https://doi.org/10.5167/uzh-261876>

This Version is available at:

<https://hdl.handle.net/10419/342338>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



**University of
Zurich**^{UZH}

University of Zurich
Department of Economics

Working Paper Series
ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 448

A Theory of Likes

Jean-Michel Benkert and Armin Schmutzler

Revised version, June 2026

A Theory of Likes

Jean-Michel Benkert and Armin Schmutzler*

June 21, 2026

Abstract

We study binary, non-strategic ratings when consumer preferences reflect a combination of objective quality differences between products and idiosyncratic taste. A rating then decomposes into objective and subjective content. Rating systems that are extremely strict or extremely lenient only convey objective content, maximizing consumer welfare when the population is not biased toward a particular product. With a biased population, less extreme rating systems that convey subjective content can be valuable. In many environments, a sales-maximizing platform prefers lenient rating systems, creating a conflict with welfare maximization.

Keywords: recommendations, ratings, preference heterogeneity, information design, platform design, value maximization

JEL Classification: D02, D47, D83

*Benkert: Department of Economics, University of Bern, jean-michel.benkert@unibe.ch. Schmutzler: Department of Economics, University of Zurich, and CEPR, armin.schmutzler@econ.uzh.ch. This paper was previously circulated under the title “A Theory of Recommendations”. We are grateful for very helpful discussions with Dirk Bergemann, Piotr Dworczak, Johannes Johnen, Igor Letina, Simon Loertscher, Matt Mitchell, Marc Möller, Nick Netzer, Volker Nocke, Martin Peitz, Marcus Reisinger, Egor Starkov, Jakub Steiner and seminar participants in Bamberg, Bern, Fribourg, Lausanne, Zurich, at MaCCI (2025), SAET (2023), Swiss Society of Economics of Statistics (2024), Swiss IO Day (2024) and Verein für Socialpolitik (2024).

1 Introduction

Simple peer-to-peer ratings play an essential role for the diffusion of economic information in diverse domains, most notably in the digital economy. In some ways, this is surprising. For example, consider a consumer deciding whether to “like” a restaurant after a meal. Her experience may depend on at least two dimensions, such as food quality and service quality, each of which can be good or bad. Consumers agree that good food is better than bad food and good service is better than bad service, but they may differ in how much weight they place on each dimension. The consumer leaving the rating has little incentive to think hard about how her like will be interpreted, while the consumer receiving the rating may rely on it to avoid a poor match. Moreover, even a thoughtful consumer cannot convey all relevant nuances through a coarse binary signal. The example illustrates three core features of peer-to-peer rating: *Products and preferences are heterogeneous, stakes are asymmetric* for those who leave ratings and those who read them, and *ratings are inherently coarse*.

In this paper, we study how such ratings can nonetheless provide value. We first show how coarse rating systems can be designed in such a way that a like (dislike) conveys objectively good (bad) news about a product. We then argue that, from a consumer welfare perspective, it need not be optimal to make use of this option. The subjectivity of ratings is not always a bug to be eliminated, but can be the feature that makes a rating system valuable: If the population is known to be biased toward particular products, it may be desirable to design the system in such a way that consumers disagree about whether a like is good or bad news about a product. Consumers with majority taste will then understand that they should buy such a product, those with minority tastes understand that they should avoid it. Having identified the optimal design of the rating system from a consumer perspective, we then find that a sales-maximizing platform will often deviate from this prescription, relying too often on excessively lenient rating systems.

Our model captures the three core features of the example in a parsimonious way. An agent must decide between two *products*, which can come in different variants. There are

objectively good (bad) variants, for which all consumers agree that they are better (worse) than all alternatives. Moreover, there are two controversial variants with intermediate values, which consumers rank differently. Ex ante, consumers merely have distributional information on which variant they are facing. However, one of the products comes with a rating from a previous consumer: This *sender* mechanically gives a like for a product that is objectively good or at least sufficiently good from her perspective (with payoffs at least some exogenous rating standard R). The *receiver* of the signal knows the preference distribution and therefore how likely it is that his preferences are aligned with those of the sender. Finally, based on this information, he decides whether to be obedient, that is, behave according to the prescription of the rating.

A transparent cut-off condition determines whether a receiver is *obedient* or *disobedient*. Importantly, the condition is the same for likes and dislikes. The key insight is that ratings contain both objective and subjective content. From a (dis-)like, a receiver learns that the product is not objectively bad (good), generating a payoff effect that pushes every receiver toward obedience. A like usually also raises the probability that, if the product is controversial, it is of the variant preferred by the majority. For consumers with majority tastes, these two pieces of information are both good news, so that they are obedient. Consumers with minority tastes face opposing effects and are obedient only when the objective effect outweighs the subjective one, and disobedient otherwise.

Our main contributions concern the design of rating systems. Platforms often try to influence how critical consumer ratings are, for example through guidelines or instructions.¹ We capture this by assuming that the platform controls the rating standard R above which a like is given. Our first contribution is that we characterize the value of rating systems for consumers and identify the value-maximizing choice of R . When the population is not biased toward any of the controversial products on average, the subjective effect of ratings cancels out in the aggregate, and extreme rating standards are optimal: Likes should be reserved for objectively good products, or dislikes for bad ones.

¹For example, Uber employs a five-star system but informs riders that “5 stars means the ride went without issues,” while “4 stars or less” prompt requests for feedback (Uber, 2024). Similarly, Airbnb communicates that a 5-star rating means “everything was as expected” (Hospitable, 2024).

When the population is biased, ratings that convey subjective content due to an interior rating standard may be strictly preferable. Intuitively, ratings can provide value in two ways. First, they can push consumers toward objectively good products and help them avoid objectively bad ones. Second, they can help consumers obtain a product variant that fits their tastes, even when it is not perfect. These two objectives need not be conflicting: Consumers with majority tastes will benefit in both ways from being obedient. Instead, for consumers with minority tastes, being obedient reduces the chance of getting the preferred controversial product variant. Being disobedient reduces the chance of getting the objectively good rather than the objectively bad variant. Either way, the rating system creates value for these consumers, even if they do the opposite of what they are told.

Our second main contribution concerns the potential conflict between consumer welfare and sales maximization by a platform that benefits from transactions of the rated product, for instance through transaction fees or advertising revenue tied to traffic. We find that a sales-maximizing platform usually favors a low R , issuing likes for everything but objectively bad products. By contrast, the value-maximizing R can be high or interior.

We provide several further results. First, we introduce prices and show that the firm's pricing decision affects not only demand but also the informativeness of the rating. Second, we analyze the case where receivers observe multiple independent ratings. Perhaps surprisingly, access to more signals does not necessarily increase value or sales. Third, we allow the sender and receiver populations to differ, showing that our main results remain robust while new design considerations arise regarding the composition of the sender pool. We close by examining how polarization and the prevalence of controversial products shape the value of the rating system.

1.1 Related Literature

Our paper relates most directly to work distinguishing vertical quality from horizontal fit in online reviews. Sun (2012) argues that variance of ratings does not reflect pure

noise but conveys information about taste differences. Chen, Hitt, Hong, and Wu (2021) theoretically and empirically separate uncertainty about quality from uncertainty about fit, finding that online ratings resolve the former more effectively than the latter. Closest in spirit is Lafky and Ng (2024), who show experimentally that raters evaluate products through their own preferences and consumers interpret ratings accordingly. Bohren, Hakimov, and Santos-Pinto (2025) provide complementary experimental evidence, showing that ratings raise welfare under vertical but not under horizontal differentiation, while segmenting ratings by consumer type restores the welfare gains.²

In a broad sense, our static inference problem belongs to the literature on observational learning.³ A more closely related literature deals specifically with learning from online reviews. For instance, Acemoglu, Makhdoumi, Malekian, and Ozdaglar (2022) show that consumers with strong prior preferences are more likely to purchase and review, so the pool of recommenders is endogenously selected and receivers must account for this bias. Che and Hörner (2018) study the tension between informativeness and exploration in recommender systems. We study a single binary signal in isolation, which makes the rating standard the sole design instrument, enabling us to characterize precisely how it trades off objective against subjective content.

Our design problem is naturally framed as one of constrained information design: The platform cannot choose an arbitrary signal structure (Kamenica and Gentzkow, 2011), only the rating standard at which senders convert their experience into a like. This can be justified using results on the optimality of coarse disclosure in certification and rating design, including partial information revelation (Rayo and Segal, 2010), monopolistic pass/fail certification (Lizzeri, 1999), and optimal partition of quality into rating tiers (Hopenhayn and Saeedi, 2026). Unlike this prior work, our signal is generated mechanically from the experiences of heterogeneous senders, and the central trade-off is not more

²Bohren et al. (2025) isolate the pure vertical and pure horizontal cases, while we study environments in which both dimensions coexist within a single product space.

³However, this literature, e.g. Banerjee (1992) Bikhchandani, Hirshleifer, and Welch (1992) and Smith and Sørensen (2000), focuses on dynamics, and it emphasizes learning from the decisions of others rather than ratings.

versus less information but objective versus subjective content.⁴

Our analysis of the tradeoff between value and sales maximization connects to the literature on platform economics and rating bias, where platform incentives often diverge from consumer welfare. The most closely related papers in this area also work with heterogeneous tastes. Shi and Raghu (2020) study *which* product a platform should feature, whereas we study *how demanding* the standard for a positive recommendation should be. Peitz and Sobolev (2025) show that a profit-maximizing firm may optimally recommend products even to poorly matched consumers. A literature on reputation and recommendation bias (Bourreau and Gaudin, 2022; Tadelis, 2016) documents how feedback systems can be distorted by strategic behavior.⁵ By contrast, the wedge between platforms and consumers that we identify does not rest on fake reviews or inflated individual messages. Ratings are mechanically truthful, and the distortion arises because a sales-maximizing platform chooses a more permissive rating standard than a value-maximizing one.

Section 2 introduces the model. Section 3 characterizes receiver behavior. Section 4 studies the value of rating systems, first characterizing the sources of value and then analyzing the optimal design of the rating standard. Section 5 turns to the parallel question of sales maximization, where a platform seeks to maximize transaction volume rather than consumer welfare. Section 6 presents further results. Section 7 concludes. All proofs are relegated to the appendix.

2 Model

Products and Payoffs We consider a consumer’s choice between two products. Each product can be of one of four variants: g, b, c_1 and c_2 . The probability of each variant is

⁴Strategic senders can produce coarse messages endogenously, see, e.g., Chakraborty and Harbaugh (2010) and Smirnov and Starkov (2025). Our senders are non-strategic, so coarseness is imposed by the binary rating rule rather than chosen.

⁵For instance, Bolton, Greiner, and Ockenfels (2013) show how the threat of retaliation in reciprocal feedback systems leads buyers to withhold negative feedback.

independently drawn from the same distribution $\mathbf{q} = (q_g, q_1, q_2, q_b)$ where

$$\Pr(g) = q_g > 0, \Pr(b) = q_b > 0, \Pr(c_1) = q_1 > 0, \Pr(c_2) = q_2 > 0.$$

Consumers cannot observe the product variant before consumption, but they know the distribution. Consumer types are distributed on $[-1/2, 1/2]$ according to F , which is absolutely continuous with full support. All consumers obtain a payoff of 1 from consuming variant g and a payoff of 0 from consuming variant b . The payoff from consuming variant c_1 is $1/2 + i$, the payoff from variant c_2 is $1/2 - i$. Therefore, consumers agree that product variants g and b are (objectively) good and bad, leading to the maximal (minimal) payoff. By contrast, the payoff from c_1 and c_2 depends on type i , so that the assessment of these variants is subjective. We thus refer to them as *controversial*. Types $i > 0$ prefer c_1 to c_2 , and vice versa for $i < 0$, while type $i = 0$ is indifferent between the two variants.

We use the following self-explanatory terminology:

Definition 1 *The prevalence of controversial products is given as $Q := \frac{q_1 + q_2}{2}$. The odds of a good product are $\sigma := q_g / q_b$.*

A Motivating Example To motivate the payoff structure, one might assume that a product variant has two quality dimensions $d \in \{1, 2\}$ that are relevant to consumers. The quality level in each dimension d can take values $\omega_d \in \{0, 1\}$. All consumers agree that high quality in any dimension is desirable, but they put weight $\frac{1}{2} + i$ on dimension 1 and weight $\frac{1}{2} - i$ on dimension 2. Defining c_i as a product variant with quality 1 in dimension i and quality 0 in dimension $j \neq i$, the above payoff structure obtains.

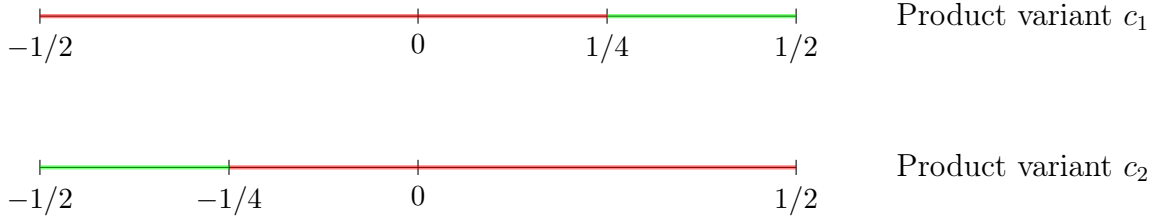
Ratings and Updating Before choosing products, a consumer receives a rating for one of the two products from a previous consumer. This *sender* (she) and the *receiver* (he) are both drawn from the distribution F . Senders and receivers who consume the same product face the same variant and differ only in their type-dependent evaluation of controversial products. The receiver chooses between the product with a rating r , which

can be a like ($r = L$) or a dislike ($r = D$), and the alternative without ($r = 0$).

To eliminate any scope for strategic sender behavior, we assume that a sender mechanically gives a like if she obtained a payoff of at least $R \in (0, 1)$ and a dislike otherwise.⁶ Intuitively, a like conveys the following information to the receiver: The probability of the bad product variant falls to zero, the likelihood of the good variant relative to the controversial variants rises, and the odds between the controversial variants change.

For the objectively good and bad variants g and b , there is *sender agreement*: All types like the former and dislike the latter. For the controversial variants, sender behavior is type-dependent: After consuming c_1 , a sender of type i gives a like if and only if $i \geq R - 1/2$, whereas types $i \leq 1/2 - R$ give a like after consuming c_2 ; see Figure 1. A rating has *objective content* on which all senders agree, and *subjective content* on which they do not. The rating standard determines the mix between objective and subjective content: For interior R , sender disagreement on controversial products is preserved and ratings carry nontrivial subjective content. As $R \rightarrow 0$ or $R \rightarrow 1$, sender disagreement vanishes and ratings become purely objective in content.

FIGURE 1. ILLUSTRATION OF SENDER BEHAVIOR



Notes: The figure illustrates what types send a like (green) and a dislike (red) for the two controversial product variants when $R = 3/4$.

Let $\phi_1(R) := 1 - F(R - 1/2)$ and $\phi_2(R) := F(1/2 - R)$ denote the probability of a like conditional on the product variant being c_1 or c_2 , respectively. The probability that a randomly chosen sender gives a like or dislike, respectively, is

$$\pi^L(R) := q_g + q_1\phi_1(R) + q_2\phi_2(R),$$

⁶We exclude the rating standard $R = 0$ as this would render the rating uninformative. For symmetry, we also exclude $R = 1$, but all results would essentially go through if we allowed $R = 1$.

$$\pi^D(R) := 1 - \pi^L(R) = q_1(1 - \phi_1(R)) + q_2(1 - \phi_2(R)) + q_b.$$

In words, the probability of receiving a like from a randomly chosen sender is given by the sum of the probability that a product is objectively good and the probability that it is subjectively good enough from a randomly chosen sender's perspective. We assume that receivers are Bayesian. In Appendix A.1.1, we present the receiver's posterior $\mathbf{p}^r(R) := (p_g^r(R), p_1^r(R), p_2^r(R), p_b^r(R))$ after a like $r = L$ or a dislike $r = D$.⁷

Finally, we use the following summary terminology:

Definition 2 $\mathcal{E} := (q, F)$ defines a *decision environment*. A *rating system* $\mathcal{R}$ consists of a decision environment $\mathcal{E}$ and a rating standard $R \in (0, 1)$.

Throughout, the decision environment $\mathcal{E}$ and the rating standard R are common knowledge among all agents.

3 Receiver Behavior

The receiver must decide whether to follow the prescription of the rating, that is, whether to purchase the rated product after a like and avoid it after a dislike. We now characterize the receiver's optimal choice, depending on his type i and the rating system $\mathcal{R}$.

The expected payoffs from buying the product with a rating and the alternative without are respectively given as

$$\begin{aligned} U^r(i) &:= p_g^r + (1/2 + i)p_1^r + (1/2 - i)p_2^r, \quad r \in \{L, D\}; \\ U^0(i) &:= q_g + (1/2 + i)q_1 + (1/2 - i)q_2. \end{aligned}$$

Receiver i buys the product rated with a like if and only if $U^L(i) \geq U^0(i)$ or equivalently

$$\frac{p_g^L - q_g}{2} - \frac{p_b^L - q_b}{2} + i[(p_1^L - q_1) - (p_2^L - q_2)] \geq 0. \quad (1)$$

⁷To simplify notation, we will often drop the dependence of these posteriors and other variables on R .

The receiver follows the prescription of dislike ratings if and only if $U^0(i) \geq U^D(i)$. A rating's informational impact on the receiver decomposes into an objective and a subjective payoff effect. To make this precise:

Definition 3 For a rating $r \in \{D, L\}$, let Δ_O^r and $-i\Delta_S^r$ denote its objective and subjective payoff effects, respectively, where

$$\Delta_O^r := \frac{p_g^r - q_g}{2} - \frac{p_b^r - q_b}{2},$$

$$\Delta_S^r := (p_2^r - q_2) - (p_1^r - q_1).$$

If $\Delta_O^r > 0$, the rating r has shifted probability from the objectively bad to the objectively good variant, a payoff gain shared by all types. Δ_S^r measures the analogous shift between the controversial variants: from c_1 to c_2 when positive, and conversely when negative. Whether this shift is good news depends on the receiver: The sign of the subjective payoff effect, $-i\Delta_S^r$, depends on the signs of i and Δ_S^r , so that a shift toward c_2 is welcome only for types who prefer c_2 to c_1 . Applying this notation to condition (1), we obtain:

Proposition 1 For a given rating system $\mathcal{R}$, a receiver of type i buys the rated product after a like and does not buy it after a dislike if and only if

$$\Delta_O^L \geq i\Delta_S^L. \tag{2}$$

Intuitively, a receiver follows a prescription only if he expects the rating to come from a sender whose preferences are sufficiently aligned with his. Condition (2) formalizes the resulting trade-off via the objective and subjective payoff effects. Importantly, this condition is the same for likes and dislikes, which motivates the following terminology:

Definition 4

- (i) A receiver is **obedient** if he buys the rated product after a like but not after a dislike, **disobedient** if he buys it after a dislike but not after a like.

(ii) Receivers are **universally obedient** in decision environment $\mathcal{E}$ if all receivers are obedient for every $R \in (0, 1)$.

Denoting the type who is indifferent between being obedient or disobedient as

$$\tilde{i} := \frac{\Delta_O^L}{\Delta_S^L} \text{ if } \Delta_S^L \neq 0,$$

Proposition 1 immediately implies the following result:

Corollary 1 *Fix a rating system $\mathcal{R}$.*

- (i) *If $|\Delta_S^L| \leq 2\Delta_O^L$, then all $i \in [-1/2, 1/2]$ are obedient.*
- (ii) *If $\Delta_S^L < -2\Delta_O^L$, then all $i \geq \tilde{i}$ are obedient.*
- (iii) *If $\Delta_S^L > 2\Delta_O^L$, then all $i \leq \tilde{i}$ are obedient.*

Intuitively, in case (i), the objective effect is so large that it dominates even the subjective effect of the most extreme receivers. As Δ_S^L and hence the subjective effect increases, receivers who are not sufficiently aligned with the population at large will be disobedient; see cases (ii) and (iii). The interplay between objective and subjective effect leads to a stark result for extreme rating standards:

Proposition 2 *Suppose $R \rightarrow 1$ or $R \rightarrow 0$. Then, all types are obedient irrespective of the decision environment $\mathcal{E}$.*

In the limits, ratings have purely objective content. For instance, as $R \rightarrow 1$, a like guarantees the product is objectively good, while a dislike implies it is not.

4 Value of Rating Systems

This section takes a welfare perspective, studying the value of rating systems for receivers and deriving the value-maximizing choice of the rating standard.

4.1 Characterization of the Value

Without a rating system, all product variants are identical ex ante, with an expected payoff based on the prior distribution of product variants:

$$V^0(i) := U^0(i). \quad (3)$$

If one of the products comes with a rating, an obedient receiver i chooses this product (rather than the alternative) only if it comes with a like. The expected payoff is thus

$$V^+(i) := \pi^L U^L(i) + (1 - \pi^L) U^0(i). \quad (4)$$

The first term on the right-hand side comes from likes, the second one from dislikes.

A disobedient receiver will buy a product that comes with a dislike. Otherwise, he will buy the alternative. Hence, his expected payoff is

$$V^-(i) := \pi^L U^0(i) + (1 - \pi^L) U^D(i). \quad (5)$$

A like does not create value for disobedient receivers, who choose the alternative without a rating. Instead, a dislike, which is given with probability $(1 - \pi^L)$, influences their behavior: Rather than buying an alternative based on priors, updating leads them to choose the product for which they received the explicit recommendation not to buy.

We now define the value of a rating system.⁸

Definition 5 $V(\mathcal{R})$, the *value of a rating system* $\mathcal{R} = (\mathcal{E}, R)$, is the increase in expected payoffs resulting from the existence of a rating standard R in a decision environment $\mathcal{E}$, where the expectation is taken over all pairs of independently drawn senders and receivers.

Using Definition 5 and Corollary 1, we can characterize $V(\mathcal{R})$ as follows:

⁸This definition is in line with the definition of a valuable signal in Kamenica and Gentzkow (2011).

- (i) $\int_{-1/2}^{1/2} (V^+(i) - V^0(i)) dF(i)$ if $|\Delta_S^L| \leq 2\Delta_O^L$;
- (ii) $\int_{-1/2}^{\tilde{i}} V^-(i) dF(i) + \int_{\tilde{i}}^{1/2} V^+(i) dF(i) - \int_{-1/2}^{1/2} V^0(i) dF(i)$ if $\Delta_S^L < -2\Delta_O^L$;
- (iii) $\int_{-1/2}^{\tilde{i}} V^+(i) dF(i) + \int_{\tilde{i}}^{1/2} V^-(i) dF(i) - \int_{-1/2}^{1/2} V^0(i) dF(i)$ if $\Delta_S^L > 2\Delta_O^L$.

All receivers are obedient when $|\Delta_S^L| \leq 2\Delta_O^L$, as the objective effect dominates (case (i)). Hence, by equation (4), the rating system creates value only for a receiver who observes a like, as a dislike leads to the same expected payoff as without a rating system. Parts (ii) and (iii) refer to cases where the subjective effect of the messages matters more. Then, not all receivers are obedient and, as argued above, the system can create value for receivers that are obedient as well as those who are not.

We restate the value $V(\mathcal{R})$ to better understand how it reflects the role of the objective and subjective payoff effects of ratings:⁹

Proposition 3 *The value $V(R)$ of the rating system $\mathcal{R}$ is:*

- (i) $\pi^L [\Delta_O^L - \Delta_S^L \mathbb{E}[i]]$ if $|\Delta_S^L| \leq 2\Delta_O^L$;
- (ii) $(1 - \pi^L) F(\tilde{i}) [\Delta_O^D - \Delta_S^D \mathbb{E}[i \mid i \leq \tilde{i}]] + \pi^L (1 - F(\tilde{i})) [\Delta_O^L - \Delta_S^L \mathbb{E}[i \mid i \geq \tilde{i}]]$ if $\Delta_S^L < -2\Delta_O^L$;
- (iii) $\pi^L F(\tilde{i}) [\Delta_O^L - \Delta_S^L \mathbb{E}[i \mid i \leq \tilde{i}]] + (1 - \pi^L) (1 - F(\tilde{i})) [\Delta_O^D - \Delta_S^D \mathbb{E}[i \mid i \geq \tilde{i}]]$ if $\Delta_S^L > 2\Delta_O^L$.

The positive payoff effect of the rating system in case (i) consists of the type-independent objective effect Δ_O^L and the expected subjective effect $-\Delta_S^L \mathbb{E}[i]$. As Δ_S^L increases in absolute terms, some receivers become disobedient. For instance, when $\Delta_S^L > 2\Delta_O^L$, a like shifts beliefs toward the controversial product variant that is less preferred by types $i > 0$, who may thus become disobedient. Only the fraction $F(\tilde{i})$ of the population whose preferences are sufficiently aligned with the sender's ($i < \tilde{i}$) still buy. For them, the subjective effect is either positive or not sufficiently negative to dominate the positive objective effect. The average subjective payoff effect among the obedient types is $-\Delta_S^L \mathbb{E}[i \mid i \leq \tilde{i}]$. As argued above, the fraction $1 - F(\tilde{i})$ of disobedient types may benefit from the rating. They infer from a dislike that the product is likely to be of the controversial variant they personally prefer and thus buy despite or rather because of the negative signal.¹⁰

⁹As we focus on R as the design parameter, we often write $V(R)$ instead of $V(\mathcal{R})$.

¹⁰This is reflected in the term $(1 - \pi^L)(1 - F(\tilde{i})) [\Delta_O^D - \Delta_S^D \mathbb{E}[i \mid i \geq \tilde{i}]]$ in Proposition 3(iii).

4.2 Design

The design problem is to choose a rating standard $R \in (0, 1)$.¹¹ Extreme standards yield sender agreement and thus purely objective content, while interior standards preserve disagreement and admit subjective content. We organize the analysis along this dichotomy. We first show that, when the population is symmetric, the aggregate contribution of the subjective effect cancels out, so design is governed by the objective effect alone and extreme standards are generically optimal. We next show that, when the population is asymmetric, the subjective effect no longer cancels out in the aggregate, and balancing it against the objective effect yields an interior optimum under simple general conditions on the decision environment. Finally, we present a tractable parametric example in which both regimes appear and the optimal rating standard can be characterized in closed form.

4.2.1 Symmetric populations

We begin with symmetric populations, so that preference heterogeneity corresponds to pure noise on average.

Assumption 1 $F(-i) = 1 - F(i)$ for all $i \in [-1/2, 1/2]$.

Under symmetry, the share $1 - F(R - 1/2)$ of senders who give a like after consuming c_1 is the same as the share $F(1/2 - R)$ who give a like after consuming c_2 , and we define $\beta := F(1/2 - R) = 1 - F(R - 1/2)$. Symmetry implies that the subjective effect of ratings cancels out in the aggregate. From Proposition 3, we thus obtain:

Lemma 1 *For any rating system satisfying Assumption 1, all types are obedient and $V(R) = \pi^L \Delta_O^L$.*

The optimal rating standard then depends only on σ , the odds of a good product:

Proposition 4 *For any rating system satisfying Assumption 1, $V(R)$ is*

¹¹Throughout, we formulate the design problem on the open interval $R \in (0, 1)$. Statements favoring extreme standards should therefore be understood as boundary statements, i.e. as limits or maxima of the continuous extension at $R = 0$ or $R = 1$.

- (i) decreasing in R if $\sigma < 1$;
- (ii) increasing in R if $\sigma > 1$;
- (iii) independent of R if $\sigma = 1$.

Intuitively, since the subjective effect of ratings cancels out in the aggregate, the designer can focus on the objective effect. To this end, he either chooses a very low rating standard close to 0 that reveals whether a product is objectively bad or not, or a very high rating standard close to 1 that reveals whether a product is objectively good or not. Which choice is better depends on how important these two pieces of information are: When objectively bad goods are common ($\sigma < 1$), it is important to avoid them, so an extremely low rating standard is appropriate; when objectively good products are common ($\sigma > 1$), it is important to identify them through extremely high rating standards.

4.2.2 Asymmetric populations

We now turn to asymmetric populations. To capture biased populations in a tractable way, we introduce the following weak assumption.

Assumption 2 *F has mean $\mu := \mathbb{E}[i] \neq 0$. Moreover, its density f extends continuously to $[-\frac{1}{2}, \frac{1}{2}]$ and satisfies $0 < f(\pm\frac{1}{2}) < \infty$.*

We use a shorthand notation for the net payoff of the average receiver from consuming each controversial variant rather than taking the outside option:

$$\begin{aligned} A_1 &:= \left(\frac{1}{2} + \mu\right) - U^0(\mu), \\ A_2 &:= \left(\frac{1}{2} - \mu\right) - U^0(\mu). \end{aligned}$$

Here, $\frac{1}{2} + \mu$ and $\frac{1}{2} - \mu$ are the gross payoffs of type μ from variants c_1 and c_2 , respectively, while $U^0(\mu)$ is her expected payoff from the unrated alternative. We obtain a sufficient condition for an interior rating standard to be optimal:

Proposition 5 *For any rating system satisfying Assumption 2, if*

$$q_1 A_1 f(-\tfrac{1}{2}) + q_2 A_2 f(\tfrac{1}{2}) < 0 < q_1 A_1 f(\tfrac{1}{2}) + q_2 A_2 f(-\tfrac{1}{2}) \quad (6)$$

then there exists an optimal $R^ \in (0, 1)$.*

Condition (6) is clearly inconsistent with Assumption 1. Moreover, it requires that A_1 and A_2 have different signs. W.l.o.g., we assume that $A_1 > 0 > A_2$, meaning that variant c_1 is valuable for the average receiver relative to the outside option, whereas variant c_2 is not. Under this interpretation, condition (6) has a transparent meaning. The first inequality guarantees that rating standards near $R = 0$ are dominated by slightly more stringent standards. Starting from an extremely lenient standard, a small increase in R removes only likes from senders who have strong preferences for one of the controversial products and thus hardly value the other one. For variant c_1 , these likes come from types near $-1/2$; for variant c_2 , they come from types near $1/2$. The inequality states that these weakest marginal likes have negative average value: They are too heavily tilted toward recommendations for the controversial variant that is bad from a majority perspective.¹² Hence, an extremely lenient standard is too permissive. By similar arguments, the second inequality guarantees that extremely stringent rating standards are not desirable.

4.2.3 A tractable example

The general results of Sections 4.2.1 and 4.2.2 establish the qualitative dichotomy between symmetric and asymmetric populations. To illustrate the general mechanism from above, we now provide an explicit closed-form solution for a tractable parametric specification.¹³

Example 1 (i) $F(i) = (i + 1/2)^a$ for $a > 0$; (ii) $q_1 = q_2 = Q$.

¹²The fact that senders near $1/2$ switch from like to dislike after observing variant c_2 is desirable from the perspective of the average receiver, the corresponding switch for senders near $-1/2$ after observing variant c_1 is not. The left inequality in (6) guarantees that the beneficial effect of eliminating undesirable likes dominates the adverse effect of eliminating desirable likes.

¹³Example 1 is complementary to Assumption 2 rather than nested in it: the power distributions $F(i) = (i + 1/2)^a$ have boundary densities that vanish or diverge when $a \neq 1$. A tractable example of a distribution that satisfies Assumption 2 is given by tilted-uniform distributions.

The case $a = 1$ reflects a symmetric, unbiased population. By contrast, when $a \neq 1$, the population is biased toward one product variant or the other.¹⁴

Our characterization focuses on decision environments where the prevalence of controversial products is so low that receivers are universally obedient. In Appendix A.2.5, we show that this holds if $Q \leq \bar{Q}(a, \sigma)$ for a cutoff $\bar{Q}(a, \sigma) \in (0, 1/2]$ that is unique given a and σ , the other features of the decision environment. The condition is frequently satisfied.¹⁵ Moreover, we show in Lemma A.4 that this condition is easiest to satisfy in the unbiased benchmark $a = 1$ and becomes harder to satisfy as the population becomes more biased. We now describe the relation between the rating standard R and the value, focusing on the asymmetric case.

Proposition 6 *Suppose the decision environment $\mathcal{E}$ satisfies the conditions in Example 1 as well as $Q \leq \bar{Q}(a, \sigma)$ and $a \neq 1$.*

(i) *If $Q \geq \min \left\{ \frac{a}{a+1}, \frac{1}{a+1} \right\}$, then $V(R)$ is maximized at a unique $R^* \in (0, 1)$.*

(ii) *If $Q < \min \left\{ \frac{a}{a+1}, \frac{1}{a+1} \right\}$, then*

(a) *$V(R)$ is decreasing in R if and only if $\sigma \leq \min \left\{ \frac{a-Q(a+1)}{1-Q(a+1)}, \frac{1-Q(a+1)}{a-Q(a+1)} \right\}$;*

(b) *$V(R)$ is increasing in R if and only if $\sigma \geq \max \left\{ \frac{a-Q(a+1)}{1-Q(a+1)}, \frac{1-Q(a+1)}{a-Q(a+1)} \right\}$;*

(c) *$V(R)$ is maximized at a unique $R^* \in (0, 1)$ otherwise.*

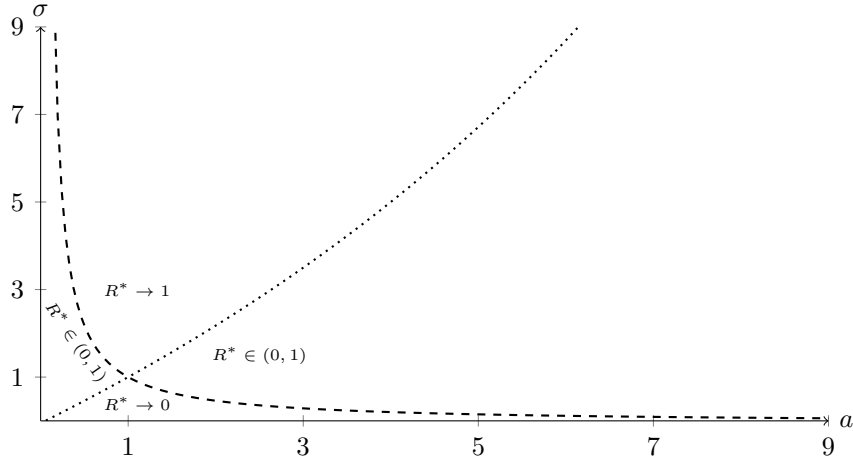
Proposition 6 first identifies $\min \left\{ \frac{a}{a+1}, \frac{1}{a+1} \right\}$ as a cutoff on the share of controversial products, above which an interior optimum exists. Below that cutoff, extreme rating standards are optimal only when the odds of a good product are sufficiently unbalanced. Figure 2 illustrates how the occurrence of interior solutions depends on the parameters: the further a is away from 1, the larger is the interval of σ values for which an interior solution exists.

¹⁴The condition $q_1 = q_2$ ensures that all subjective content in ratings is driven by the asymmetry of the type distribution F , not by prior differences in how common the controversial variants are.

¹⁵If $a = 1$, $\bar{Q}(1, \sigma) = 1/2$, so universal obedience arises for every admissible Q . For $a \neq 1$, the cutoff is still quite large (for example, $\bar{Q}(2, 1) = 1/3$). At $\sigma = 1$ and $a > 1$ the closed form $\bar{Q}(a, 1) = 1/[4(1 - 2^{-a})]$ obtains (Corollary A.2). For general σ , $\bar{Q}(a, \sigma)$ is also computable from the primitives.

In some cases, the restriction $Q \leq \bar{Q}(a, \sigma)$ is not needed. In Example 1, if $\sigma = 1$, so that $q_g = q_b$, the design role of the objective channel is shut down. The rating standard matters only through the subjective content of the rating. One can show that this content is maximized at $R = 1/2$, where senders like precisely the controversial variant they personally prefer. Thus the midpoint standard makes ratings maximally informative about the controversial dimension.¹⁶

FIGURE 2. ILLUSTRATION OF PROPOSITION 6



Notes: This figure illustrates the parameter regimes of Proposition 6 for $Q = 0.05$. The dotted and dashed curves correspond to $\sigma = (a - Q(a + 1))/(1 - Q(a + 1))$ and its reciprocal, respectively. An interior solution arises between the two curves. Otherwise, extreme rating standards are optimal.

5 Sales Maximization

We now consider a platform that maximizes sales of the product with a rating rather than receiver welfare. It will turn out that there are parameter regions where the two objectives are aligned and others where they are in stark conflict.

5.1 Characterization

Using Corollary 1, the sales for the rated product $D(R)$ can be expressed analogously to the value of the rating system in Proposition 3 as

¹⁶This remains true even when $Q > \bar{Q}(a, 1)$ and some receivers are disobedient, as disobedience does not affect the amount of subjective content conveyed by the rating.

- (i) $\pi^L(R)$ if $|\Delta_S^L| \leq 2\Delta_O^L$;
- (ii) $\pi^L(R)F(\tilde{i}(R)) + (1 - \pi^L(R))(1 - F(\tilde{i}(R)))$ if $\Delta_S^L > 2\Delta_O^L$;
- (iii) $\pi^L(R)(1 - F(\tilde{i}(R))) + (1 - \pi^L(R))F(\tilde{i}(R))$ if $\Delta_S^L < -2\Delta_O^L$,

where we have made explicit that the indifferent receiver $\tilde{i}$ and the probability of a like π^L are functions of the rating standard R . Similar to the expression for the value in Proposition 3, the relation between Δ_S^L and Δ_O^L determines which types are obedient. In case (i), all types are obedient so that sales are given by $\pi^L(R)$, the probability of a like. In cases (ii) and (iii), sales come from both obedient and disobedient types. Importantly, disobedient types buy a product rated with a dislike, thus contributing to sales.

The structural parallel between $D(R)$ above and the value expression in Proposition 3 is instructive. The determinants of both expressions are the same (π^L , Δ_O^L , Δ_S^L and $\tilde{i}$). The key difference lies in the aggregation across receivers. Value weights each rating by the resulting payoff effect for the receiver, while demand simply counts purchases. When all types are obedient, sales equal π^L , the frequency of likes, regardless of their objective effect on receivers. This difference drives the potential conflict between value and sales maximization.

5.2 Design

Proceeding as in Section 4.2, we first show that, for symmetric populations, sales are monotone in the rating standard. Then, we establish an analogous result that applies to asymmetric populations as well.

Proposition 7 *For a decision environment $\mathcal{E}$ satisfying Assumption 1, demand $D(R) = \pi^L(R)$ is decreasing in R .*

The simple form of $D(R)$ is implied by Lemma 1, which states that receivers are universally obedient under Assumption 1. This implies the result because the probability of likes is decreasing in R .¹⁷

¹⁷See Lemma A.2 in the appendix.

Comparison with Proposition 4 highlights the potential for conflict between value and sales maximization. When $\sigma < 1$, value is also decreasing in R , so the two objectives are aligned and both call for a low standard. When $\sigma > 1$, however, value is increasing in R while sales remain decreasing, so that the objectives are diametrically opposed: Sales maximization favors a low standard to make likes frequent, while value maximization favors a high standard to make likes informative.

Like value maximization, sales maximization becomes more complex for asymmetric population distributions. Nonetheless, we establish a sufficient condition for sales maximization to call for low rating standards, mirroring the result in the symmetric case.

Proposition 8 *Consider a decision environment $\mathcal{E}$ with $q_g \geq q_b$. Then, the continuous extension of $D(R)$ to $[0, 1]$ is maximized at $R = 0$.*

The result holds without assumptions on the distribution F , but requires $q_g \geq q_b$. Intuitively, sales come from obedient types who received a like and disobedient types who received a dislike. Lowering R enlarges the like channel and shrinks the dislike channel in any decision environment. The assumption $q_g \geq q_b$ ensures the first effect dominates, making $R = 0$ optimal. This contrasts sharply with value maximization. In Example 1 with $q_g > q_b$, the condition for value to be decreasing in R in Proposition 6 is never satisfied: Since good products are common, likes are frequently given for low R , and the emphasis shifts from frequency to informativeness. Increasing R filters out controversial products, making likes more selective and thus more valuable.

When $q_g < q_b$, the analytical argument for Proposition 8 no longer extends. One sufficient condition under which demand is nonetheless highest for low rating standards is *universal obedience*: Then demand coincides with the probability of a like, so lowering R weakly increases sales regardless of whether q_g is larger or smaller than q_b . Thus, the failure of Proposition 8 in the unrestricted model must rely on some receivers being *disobedient*. This is not merely a limitation of the proof: without further restrictions on F , one can construct absolutely continuous full-support type distributions for which the continuous extension of $D(R)$ is not maximized at $R = 0$. The reason is that, when bad products are

sufficiently common, dislikes are frequent. If the type distribution is sufficiently biased, the subjective effect may dominate the objective effect for many receivers. Demand can then be driven by disobedient receivers rather than by the frequency of likes.

This possibility appears not to arise in the tractable specification of Example 1, in which $q_g < q_b$ corresponds to $\sigma < 1$. An extensive numerical search found no violations of monotonicity, so that sales remain decreasing in R . Thus, while Proposition 8 is false in the unrestricted model without $q_g \geq q_b$, the low-standard conclusion appears robust within Example 1.

6 Further Results

6.1 Optimal Pricing

We now introduce prices to highlight that, when ratings are based on net utility, the price of a product affects not only its sales but also the value of a rating. This connects to a literature on how prices and ratings interact.¹⁸ Our purpose is not to study full-fledged pricing competition, but to isolate, in the simplest setup, how the firm’s pricing choice feeds back into the value of the rating.

We consider a symmetric population and assume for simplicity that types are distributed uniformly.¹⁹ To simplify the exposition further, we assume that the controversial variants are equally likely and set $Q = q_1 = q_2$. Receivers choose between buying the product with a rating at price $P \geq 0$ and buying the product without a rating at a price normalized to be $P_0 = 0$. The rating standard $R \in (0, 1)$ is fixed (e.g., by a platform), and the firm selling the product with a rating chooses P to maximize profit $\Pi(P; R) = P D(P; R)$, where $D(P; R)$ denotes demand at price P given standard R . Ratings are based on *net* utility. For instance, after consuming product variant c_1 , a

¹⁸Johnen and Ng (2024) show that firms can strategically use pricing to harvest favorable ratings, Peitz and Sobolev (2025) document rating inflation, and Carnehl, Stenzel, and Schmidt (2024) study dynamic pricing when prices directly shape the distribution of reviews.

¹⁹Assuming the uniform distribution simplifies the analysis and admits closed-form solutions. Economically, the key assumption is symmetry as in Assumption 1 to ensure that the two controversial variants generate likes with the same probability at a given rating standard.

sender gives a like if and only if $1/2 + i - P \geq R$, resulting in an effective rating standard $T := R + P$. We work with the continuous extension of the rating rule to $T = 1$ and can therefore consider $T \in (0, 1]$, equivalently prices $P \in [0, 1 - R]$, without loss.

The receiver buys the product with a like if and only if

$$U^L(R + P) - P \geq U^0 \Leftrightarrow P \leq \Delta_O^L(R + P).$$

Since $\Delta_S^r = 0$, all types consider a like objectively good news and the dislike objectively bad news. After a dislike rating, the product with a rating is dominated by the outside option.²⁰ Moreover, either all types buy the product with a like at price P or none do. Therefore, sales at price P take the simple form

$$D(P; R) = \pi^L(R + P) \cdot \mathbf{1}\{P \leq \Delta_O^L(R + P)\}, \quad (7)$$

and the firm's profits are $\Pi(P; R) = P \cdot D(P; R)$. Intuitively, at a price of zero, all consumers who observe a like buy the rated product. By increasing the effective rating standard $T = R + P$ and thus decreasing $\pi^L(R + P)$, a higher price makes a like less likely. However, conditional on observing a like, the rating becomes more valuable because it is harder to obtain. This increases the receiver's willingness to pay after a like, so that $\Delta_O^L(R + P)$ is increasing in P .²¹ In regions where the latter effect dominates, sales are increasing in P . They remain positive as long as $P \leq \Delta_O^L(R + P)$.

Define $P^u(R)$ as the unrestricted maximizer of the objective $P \pi^L(R + P)$, ignoring the demand constraint $P \leq \Delta_O^L(R + P)$. Denote the upper bound of the feasible price set as $\bar{P}(R) := \max \{P \in [0, 1 - R] : P \leq \Delta_O^L(R + P)\}$. We then obtain:

Proposition 9 *Consider a decision environment $\mathcal{E}$ such that the type distribution F is uniform, $q_1 = q_2$ and $q_g > Q$.²² Then, for a fixed rating standard $R \in (0, 1)$, the firm's*

²⁰Formally, $U^D(R + P) \leq U^0$ and $P \geq 0$ imply $U^D(R + P) - P \leq U^0$.

²¹This mechanism is related to the rating inflation studied in Peitz and Sobolev (2025) and the price-rating dynamics in Johnen and Ng (2024), where firms strategically manipulate the informativeness of ratings through pricing.

²²The assumption $q_1 = q_2$ eliminates the *subjective* content of ratings and delivers the simple “all-

profit-maximizing price is $P^*(R) = \min \{ P^u(R), \bar{P}(R) \}$.

The above result also sheds light on the conflict between value and sales maximization. A higher price raises the effective rating standard $T = R + P$ above R . By Proposition 4, when (objectively) good products are common, this upward shift moves T in a value-enhancing direction: a higher positive price makes the rating more selective and thereby more valuable. Conversely, when good products are scarce, a higher price pushes T away from the value-maximizing level. In this case, the firm's profit motive undermines the value of the rating system.

6.2 Multiple Ratings

We now show that additional ratings fail to add value when controversial variants are equally likely, or when the outside option is so attractive or unattractive that the single-rating optimum already induces the same decision. Formally, we consider a rating system where the receiver observes an arbitrary large number of ratings (b likes and d dislikes, each independently drawn from a population of senders as in the benchmark model).²³ To highlight the limits of informational gains from additional ratings, we focus on the case of *infinite learning*, where the receiver observes an unbounded number of independent ratings.

Proposition 10 *In the model with multiple ratings, suppose the decision environment $\mathcal{E}$ satisfies Assumption 1 holds. Let $\lambda := q_1/q_2$. Then, infinite learning does not increase the value of the rating system relative to the optimally designed single-rating benchmark if and only if one of the following holds:*

$$(i) \lambda = 1; \quad (ii) \lambda > 1 \text{ and } \sigma \notin \left(\frac{1}{\lambda}, \lambda \right); \quad (iii) \lambda < 1 \text{ and } \sigma \notin \left(\lambda, \frac{1}{\lambda} \right).$$

or-nothing" demand in (7). If instead $q_1 \neq q_2$, then a like typically also shifts beliefs *between* the two controversial variants. Because receiver types value these two variants differently, the same rating is more attractive for some types than for others. Demand after L then becomes *type-dependent*, and potentially only a subset of consumers are obedient at a given price.

²³We derive the posteriors for general (b, d) in Appendix A.4.2.

Intuitively, a single like is sufficient to rule out the objectively bad variant, and a single dislike rules out the objectively good variant. Hence, when the receiver obtains a mix of both, he infers that the product must be one of the two controversial variants. With infinitely many ratings, the receiver can perfectly infer whether the product is objectively good, bad, or controversial. As in the case with only one rating, receivers buy objectively good products and reject bad ones. Thus, any potential gain must come from improved decisions about controversial products. However, controversial products always generate mixed ratings. If the odds of good products are low, the outside option is unattractive, and the receiver may prefer a controversial product even if it is only an imperfect match. If the odds of good products are high, the receiver is more willing to choose the outside option. In either extreme, the decision under infinite learning coincides with what could be achieved by a single rating from an optimally designed rating system. Only when the uncontroversial variants are similarly likely and the likelihoods of the controversial variants are not too similar does additional information add value.

Finally, note that, under Assumption 1, providing more ratings does not help a sales-maximizing platform either. To achieve this goal, it suffices to choose a single rating standard close to 0, which leads all senders to issue likes when the product is not bad. Since mixed ratings can only reduce the likelihood of purchasing controversial products, adding them weakly reduces sales.²⁴

6.3 Distinct Sender and Receiver Distributions

Sender and receiver distributions may be distinct. For instance, as early adopters, senders may differ systematically from the broader population.²⁵ Our earlier results extend naturally to such a setting. Let the sender be drawn from F and the receiver from G . If

²⁴The finding that additional ratings need not add value is related to Ifrach, Maglaras, Scarsini, and Zseleva (2019), who develop a Bayesian model of sequential learning from binary like/dislike reviews and show that a seller’s optimal dynamic pricing can cause learning to fail with positive probability. Both findings illustrate that more signals need not raise welfare, though through different mechanisms.

²⁵Acemoglu et al. (2022) study how such endogenous selection into reviewing shapes the speed of Bayesian learning. Li and Hitt (2008) document how early buyers with systematically different preferences generate self-selection bias in reviews even when ratings are mechanically truthful, framed around the same objective-quality versus subjective-fit decomposition we use.

F satisfies Assumption 1, the results from Sections 4 and 5 remain valid. For example, generalizing Proposition 4, Proposition A.1 in the appendix shows that, when F is symmetric, all receiver types are obedient, $V(R)$ remains monotone in R and σ continues to play the same role as before in determining the sign of the slope. Similarly, the potential conflict between value and sales objectives persists.

Beyond such robustness results, differing sender and receiver distributions give rise to new design considerations. In many cases, platforms may have partial control over the sender population or the information receivers have about it. For instance, Amazon’s phrasing “customers who bought this item also bought...” implicitly informs the receiver about who the senders are.

First, though it may seem that making the sender distribution more precise would raise the value of the system, this is not always true. A more concentrated sender pool makes ratings less dependent on idiosyncratic taste and therefore strengthens their objective content while reducing subjective content. Whether this raises or lowers value depends on which source of information is more important in the environment.²⁶

Second, suppose that the sender and receiver distributions are fixed, but that the designer can provide the receivers with a signal about the sender’s type. Assume for simplicity that the signal fully reveals the type. If the sender is revealed to be $i = 0$, all types are obedient, as her rating only contains objective content. In contrast, if the signal reveals a sender of type $i \neq 0$, the receiver may learn about controversial products, too.²⁷ Thus, having a biased sender is not necessarily a bad thing.

Third, suppose sender and receiver types are correlated, so that ratings are effectively “tailored” to the receiver. Then the relevant sender distribution depends on the receiver’s type. If these conditional sender distributions are symmetric, the subjective component of a rating vanishes for every receiver. Hence all receivers are obedient, ratings are interpreted purely through their objective content, and a sales-maximizing platform again

²⁶In particular, adapting the argument behind Proposition A.1 shows that the effect depends on the prevalence of controversial products and on the extent to which receiver preferences are tilted toward one controversial variant or the other.

²⁷For instance, if $i = 1/2$ and $R \in (0, 1)$, a like reveals that the product is either of variant g or c_1 .

prefers a rating close to $R = 0$.

These three cases share a common thread: The sender’s characteristics shape not only how often likes are issued but also what they reveal. A biased sender, for instance, can be more informative about controversial products precisely because her ratings carry subjective content. The design of the sender pool thus offers a complementary lever to the rating standard for balancing objective and subjective content. Bohren et al. (2025) show experimentally that this lever can be relevant, where welfare in horizontally differentiated markets is restored by segmenting the rating system by consumer type.²⁸

6.4 Polarization and the Prevalence of Controversial Products

Invoking symmetry (Assumption 1), we now ask how the value of the rating system responds to two features of the environment other than the rating standard: the degree of preference polarization and the prevalence of controversial product variants.

First consider polarization (or preference heterogeneity). Recall that $\beta = F(1/2 - R)$ is the fraction of the population willing to recommend a controversial product at standard R . Under Assumption 1, β summarizes how often controversial products generate positive ratings. A higher β raises the probability of likes, but weakens their objective content, since more positive ratings stem from controversial rather than uncontroversially good products. Whether greater preference heterogeneity raises or lowers the value of the rating system depends on σ : When uncontroversially good products are common, highly informative likes increase the value of the system, so a higher β is harmful; when uncontroversially bad products are common, the value comes from ruling out bad products through more frequent likes, so a higher β is beneficial (see Appendix A.4.4).

Next, we turn to the prevalence of controversial product variants. A natural conjecture is that a higher share of controversial products always reduces the value of the rating system, because it dilutes the objective content of recommendations. This is indeed true when good and bad products are similarly likely. More generally, however, the effect

²⁸Moreover, when the platform can influence the sender population, the choice of senders may itself depend on the objective: A value-maximizing platform may prefer a different sender composition than a sales-maximizing one, adding a new dimension to the value–sales conflict.

of Q is non-monotone. Under Assumption 1, the value of the rating system can be written as $V(R) = \pi^L \Delta_O^L$, reflecting that Q has two opposing effects: Likes become more common but less likely to reflect uncontroversially good products. When σ is far from one, the frequency effect can dominate the loss in objective informativeness, so the value of the rating system may be maximized at an interior share of controversial products (Corollary A.3 in Appendix A.4.4).

7 Conclusion

We analyze the design of coarse rating systems, focusing on the binary case of likes and dislikes.²⁹ Our approach relies on the distinction between objective and subjective content of a rating. The payoff effect of a rating on a receiver has an objective part, shared by all types, and a subjective part, with sign depending on how closely her preferences are aligned with the sender's. A receiver is obedient if the objective effect outweighs the subjective one for her type. This decomposition shapes optimal design. When the population is sufficiently biased toward one of the controversial variants, a value-maximizing system preserves some subjective content at an interior standard, enabling inference about controversial products; otherwise, an extreme standard that filters out all subjective content is optimal.

The conflict between value and sales maximization is a central theme of our paper. When good products are scarce, both objectives call for low standards. When good products are common, this is usually still true for sales maximization but not for value maximization. Our result on pricing reinforces this conflict. When good products are common, the firm's pricing enhances the value of the rating system by charging a high price, but when they are low, pricing undermines the value of the system.

Although disobedient receivers play an important conceptual role in the model, they

²⁹One can show that the monotonicity result of Proposition 4 extends to richer rating structures. With three rating levels—like, neutral, dislike—defined by two rating standards $R_1 \leq R_2$, value remains monotone in the relevant standard under Assumption 1: decreasing in R_1 if $\sigma < 1$ and increasing in R_2 if $\sigma > 1$. Intuitively, the neutral rating reveals that the product is controversial but carries no subjective content under symmetry, so the design logic reduces to pushing one of the extreme standards to its boundary.

appear only sparingly in the main design results. The reason is that optimal standards often induce obedience by all receiver types. Under sales maximization, this is often the case because the platform prefers extremely lenient standards, fostering obedience. Under value maximization, universal obedience always obtains for symmetric populations, and it can arise even under asymmetric populations when the subjective effect of ratings remains sufficiently weak. Thus, disobedient receivers are central for understanding how ratings can create value, but they are less often a feature of the optimally designed system itself.

There are several avenues for future research. For instance, the current model maintains classical assumptions such as risk neutrality and Bayesian updating. Relaxing these assumptions could affect which receiver types are obedient, with potential implications for the design of the system. On a related note, it would be interesting to use experimental evidence to see whether our model captures key features of receiver behavior and in which ways it fails to do so.

A Formal Details and Proofs

A.1 Receiver Behavior

A.1.1 Posterior Distributions

The receiver's posterior $\mathbf{p}^L(R) := (p_g^L(R), p_1^L(R), p_2^L(R), p_b^L(R))$ after a like is

$$\begin{aligned} p_g^L(R) &:= \frac{q_g}{q_g + q_1\phi_1(R) + q_2\phi_2(R)}, & p_2^L(R) &:= \frac{q_2\phi_2(R)}{q_g + q_1\phi_1(R) + q_2\phi_2(R)}, \\ p_1^L(R) &:= \frac{q_1\phi_1(R)}{q_g + q_1\phi_1(R) + q_2\phi_2(R)}, & p_b^L(R) &:= 0. \end{aligned}$$

A dislike gives rise to $\mathbf{p}^D(R) := (p_g^D(R), p_1^D(R), p_2^D(R), p_b^D(R))$, where

$$\begin{aligned} p_g^D(R) &:= 0, & p_2^D(R) &:= \frac{q_2(1 - \phi_2(R))}{q_1(1 - \phi_1(R)) + q_2(1 - \phi_2(R)) + q_b}, \\ p_1^D(R) &:= \frac{q_1(1 - \phi_1(R))}{q_1(1 - \phi_1(R)) + q_2(1 - \phi_2(R)) + q_b}, & p_b^D(R) &:= \frac{q_b}{q_1(1 - \phi_1(R)) + q_2(1 - \phi_2(R)) + q_b}. \end{aligned}$$

A.1.2 Proof of Proposition 1

Let $\Delta^L(i) \equiv U^L(i) - U^0(i)$ be the gain from buying after a like and $\Delta^D(i) \equiv U^0(i) - U^D(i)$ the gain from not buying after a dislike. The receiver buys after a like (does not buy after a dislike) if and only if $\Delta^L(i) \geq 0$ (resp. $\Delta^D(i) \geq 0$).

We first show the equivalence of these conditions. For every state $s \in \{g, 1, 2, b\}$, Bayes' rule implies $\pi^L(p_s^L - q_s) + \pi^D(p_s^D - q_s) = 0$. Multiplying by the state-contingent payoffs $\{1, 1/2 + i, 1/2 - i, 0\}$ and summing over the states yields $\pi^L \Delta^L(i) = \pi^D \Delta^D(i)$. Because $\pi^L, \pi^D > 0$, $\Delta^L(i)$ and $\Delta^D(i)$ have the same sign, the desired equivalence holds.

It remains to show that this is equivalent to $\Delta_O^L \geq i\Delta_S^L$. By condition (1), the receiver buys after a like if and only if

$$(p_g^L - q_g) + \frac{p_1^L - q_1}{2} + \frac{p_2^L - q_2}{2} \geq i[(p_2^L - q_2) - (p_1^L - q_1)],$$

which is exactly $\Delta_O^L \geq i\Delta_S^L$ by Definition 3.

A.1.3 Proof of Proposition 2

First, observe that $\lim_{R \rightarrow 1} p_1^L = \lim_{R \rightarrow 1} p_2^L = 0$ and thus $\lim_{R \rightarrow 1} p_g^L = 1$. Hence, in the case $R \rightarrow 1$ the obedience condition (1) simplifies to

$$1 - q_g - \frac{q_1 + q_2}{2} \geq i(q_1 - q_2). \quad (8)$$

Equation (8) always holds for $i \geq 0$ and $q_1 \geq q_2$. Further, for $q_1 > q_2$ the condition is most stringent for $i = 1/2$. In this case, it is equivalent to $1 \geq q_1 + q_g$, which always holds. The case $i < 0$ is analogous. In the case $R \rightarrow 0$ we get $\lim_{R \rightarrow 0} p_s^L = q_s/(1 - q_b)$ for $s \in \{g, 1, 2\}$ and thus the obedience condition simplifies to

$$\begin{aligned} \frac{q_g q_b}{1 - q_b} + \frac{q_1 q_b}{2(1 - q_b)} + \frac{q_2 q_b}{2(1 - q_b)} &\geq i \left[\frac{q_2 q_b}{1 - q_b} - \frac{q_1 q_b}{1 - q_b} \right] \\ \Leftrightarrow 2q_g + q_1 + q_2 &\geq 2i[q_2 - q_1]. \end{aligned}$$

For $i \geq 0$, the last condition always holds if $q_1 \geq q_2$. Further, for $q_2 > q_1$ the condition is most stringent for $i = 1/2$. In this case, it is equivalent to $q_g + q_1 \geq 0$, which is always satisfied. The case $i < 0$ is analogous.

A.2 Value of Rating Systems

A.2.1 Proof of Proposition 3

(i) If $|\Delta_S^L| \leq 2\Delta_O^L$, inserting (3) and (4) yields

$$\begin{aligned} V(R) &= \int_{-1/2}^{1/2} (V^+(i) - V^0(i)) dF(i) \\ &= \pi^L \int_{-1/2}^{1/2} (p_g^L + (1/2 + i)p_1^L + (1/2 - i)p_2^L) - (q_g + (1/2 + i)q_1 + (1/2 - i)q_2) dF(i) \\ &= \pi^L \int_{-1/2}^{1/2} (p_g^L - q_g) + \frac{p_1^L - q_1}{2} + \frac{p_2^L - q_2}{2} dF(i) - \pi^L \int_{-1/2}^{1/2} i[(p_2^L - q_2) - (p_1^L - q_1)] dF(i) \\ &= \pi^L \int_{-1/2}^{1/2} \underbrace{\frac{p_g^L - q_g}{2} - \frac{p_b^L - q_b}{2}}_{=\Delta_O^L} dF(i) - \pi^L \int_{-1/2}^{1/2} i \underbrace{[(p_2^L - q_2) - (p_1^L - q_1)]}_{=\Delta_S^L} dF(i) \end{aligned}$$

$$= \pi^L [\Delta_O^L - \Delta_S^L \mathbb{E}[i]]$$

(ii) If $\Delta_S^L < -2\Delta_O^L$, inserting (3), (4) and (5) yields

$$\begin{aligned} V(R) &= \int_{-1/2}^{\tilde{i}} V^-(i) dF(i) + \int_{\tilde{i}}^{1/2} V^+(i) dF(i) - \int_{-1/2}^{1/2} V^0(i) dF(i) \\ &= \int_{-1/2}^{\tilde{i}} \pi^L \left(q_g + q_1 \left(\frac{1}{2} + i \right) + q_2 \left(\frac{1}{2} - i \right) \right) + (1 - \pi^L) \left(p_g^D + p_1^D \left(\frac{1}{2} + i \right) + p_2^D \left(\frac{1}{2} - i \right) \right) dF(i) \\ &\quad + \int_{\tilde{i}}^{1/2} \pi^L \left(p_g^L + p_1^L \left(\frac{1}{2} + i \right) + p_2^L \left(\frac{1}{2} - i \right) \right) + (1 - \pi^L) \left(q_g + q_1 \left(\frac{1}{2} + i \right) + q_2 \left(\frac{1}{2} - i \right) \right) dF(i) \\ &\quad - \int_{-1/2}^{\tilde{i}} q_g + q_1 \left(\frac{1}{2} + i \right) + q_2 \left(\frac{1}{2} - i \right) dF(i) - \int_{\tilde{i}}^{1/2} q_g + q_1 \left(\frac{1}{2} + i \right) + q_2 \left(\frac{1}{2} - i \right) dF(i) \\ &= (1 - \pi^L) \int_{-1/2}^{\tilde{i}} \underbrace{\frac{p_g^D - q_g}{2} - \frac{p_b^D - q_b}{2}}_{=\Delta_O^D} - i \underbrace{[(p_2^D - q_2) - (p_1^D - q_1)]}_{=\Delta_S^D} dF(i) \\ &\quad + \pi^L \int_{\tilde{i}}^{1/2} \underbrace{\frac{p_g^L - q_g}{2} - \frac{p_b^L - q_b}{2}}_{=\Delta_O^L} - i \underbrace{[(p_2^L - q_2) - (p_1^L - q_1)]}_{=\Delta_S^L} dF(i) \\ &= (1 - \pi^L) F(\tilde{i}) [\Delta_O^D - \Delta_S^D \mathbb{E}[i \mid i \leq \tilde{i}]] + \pi^L (1 - F(\tilde{i})) [\Delta_O^L - \Delta_S^L \mathbb{E}[i \mid i \geq \tilde{i}]]. \end{aligned}$$

(iii) If $\Delta_S^L > 2\Delta_O^L$, then $V(R) = \int_{-1/2}^{\tilde{i}} V^+(i) dF(i) + \int_{\tilde{i}}^{1/2} V^-(i) dF(i) - \int_{-1/2}^{1/2} V^0(i) dF(i)$.

Inserting (3), (4) and (5) and proceeding analogously as in case (ii) gives the expression.

A.2.2 Proof of Lemma 1

According to Corollary 1, all types are obedient if

$$\min\left\{\Delta_O^L + \frac{\Delta_S^L}{2}, \Delta_O^L - \frac{\Delta_S^L}{2}\right\} \geq 0. \quad (9)$$

To see that this condition is fulfilled, first observe that

$$\begin{aligned} \Delta_O^L + \frac{\Delta_S^L}{2} &= p_g^L - q_g + p_2^L - q_2 \\ &= \frac{q_g}{q_g + \beta(q_1 + q_2)} - q_g + \frac{\beta q_2}{q_g + \beta(q_1 + q_2)} - q_2 \\ &= \frac{\beta(q_2 - (q_g + q_2)(q_1 + q_2)) + q_g(1 - q_g - q_2)}{q_g + \beta(q_1 + q_2)} \end{aligned}$$

This term is positive if $q_2 - (q_g + q_2)(q_1 + q_2) \geq 0$. If $q_2 - (q_g + q_2)(q_1 + q_2) < 0$, then

$$\begin{aligned} & \beta(q_2 - (q_g + q_2)(q_1 + q_2)) + q_g(1 - q_g - q_2) \\ & \geq (q_2 - (q_g + q_2)(q_1 + q_2)) + q_g(1 - q_g - q_2) = (q_g + q_2)(1 - q_1 - q_2 - q_g) \geq 0. \end{aligned}$$

Thus, $\Delta_O^L + \frac{\Delta_S^L}{2} \geq 0$ even in this case. Proceeding analogously, one obtains $\Delta_O^L - \frac{\Delta_S^L}{2} = p_g^L - q_g + p_1^L - q_1 \geq 0$. Corollary 1 thus implies that all types are obedient. Assumption 1 implies $\mathbb{E}[i] = 0$, so that the result follows from Proposition 3.

A.2.3 Proof of Proposition 4

We first restate the probabilities of likes, their objective effect and the value of the rating system in terms of σ , β and Q :

Lemma A.1 *Suppose Assumption 1 holds. Then,*

$$\begin{aligned} \pi^L &= \frac{(1 - 2Q)\sigma}{1 + \sigma} + 2Q\beta; \\ \Delta_O^L &= \frac{\beta Q(\sigma + 1) - 2Q\sigma + \sigma}{2\beta Q(\sigma + 1) - 2Q\sigma + \sigma} - \frac{(1 - 2Q)\sigma}{1 + \sigma} - Q; \\ V(\beta) &= \pi^L \Delta_O^L = (1 - 2Q) \frac{\sigma + Q(\beta - \sigma + \sigma^2(1 - \beta))}{(\sigma + 1)^2}. \end{aligned}$$

Proof. The result on π^L follows from inserting $q_g = (1 - 2Q)\sigma/(1 + \sigma)$, $\phi_1(R) = \phi_2(R) = \beta$ and $q_1 + q_2 = 2Q$ into $\pi^L = q_g + q_1\phi_1(R) + q_2\phi_2(R)$. Next, equation (3) gives

$$\Delta_O^L = \frac{\beta Q(\sigma + 1) - 2Q\sigma + \sigma}{2\beta Q(\sigma + 1) - 2Q\sigma + \sigma} - \frac{(1 - 2Q)\sigma}{1 + \sigma} - Q \quad (10)$$

The result on $V(\beta)$ then follows immediately from Proposition 3. ■

The following result shows that countervailing effects of β on the probability of a like and the objective value Δ_O^L of accepting it result in a non-trivial effect on $V(\beta) = \pi^L \Delta_O^L$.

Lemma A.2 *Suppose Assumption 1 holds. Then,*

- (i) π^L is strictly increasing and Δ_O^L strictly decreasing in β ;

(ii) $\frac{\partial^2 V}{\partial \beta \partial \sigma} < 0$ and $V'(\beta)$ has the same sign as $1 - \sigma$.

Proof. (i) Note that $\pi^L = (1 - 2Q)\sigma/(1 + \sigma) + 2Q\beta$, so that π^L is strictly increasing. Moreover, $\frac{\partial \Delta_O^L}{\partial \beta} = -\frac{(Q(1-2Q)\sigma(1+\sigma))}{(\sigma-2Q\sigma+2\beta Q(1+\sigma))^2} < 0$. (ii) Lemma A.1 implies $V'(\beta) = \frac{Q(1-2Q)(1-\sigma)}{1+\sigma}$, and hence $\frac{\partial^2 V}{\partial \beta \partial \sigma} = -\frac{2Q(1-2Q)}{(1+\sigma)^2} < 0$, so that $V'(\beta)$ is negative for $\sigma > 1$, positive for $\sigma < 1$ and zero for $\sigma = 1$. ■

Then, using $\beta = F(1/2 - R)$, Lemma A.2 directly implies Proposition 4.³⁰

A.2.4 Proof of Proposition 5

The proof follows from Steps 1–6, which we now state and prove. We will repeatedly use the formulation that *a statement holds near the extremes*, meaning that for any decision environment $\mathcal{E}$, there exists $\delta(\mathcal{E}) > 0$ such that the statement holds for $R < \delta(\mathcal{E})$ and $R > 1 - \delta(\mathcal{E})$.

Step 1: *Near the extremes*, $V(R) = \pi^L(R)(\Delta_O^L(R) - \mu \Delta_S^L(R))$.

Proof of Step 1: For $R \rightarrow 0$ or $R \rightarrow 1$, all types have a strict preference for being obedient. Since $\Delta_O^L(R)$ and $\Delta_S^L(R)$ are continuous in R , for any decision environment $\mathcal{E}$, there exists $\delta(\mathcal{E}) > 0$ such that all types are obedient for all $R \in (0, \delta(\mathcal{E})) \cup (1 - \delta(\mathcal{E}), 1)$. Hence, for such R Proposition 3(i) applies, and Step 1 follows.

Step 2: *Near the extremes*,

$$V(R) = q_1\left(\frac{1}{2} + \mu\right)\phi_1(R) + q_2\left(\frac{1}{2} - \mu\right)\phi_2(R) - [q_g + \frac{q_1+q_2}{2} - \mu(q_2 - q_1)] \pi^L(R) + C,$$

where the term C is independent of R .

Proof of Step 2: Recalling the posteriors after a like (see page 28) and applying Definition 3, we obtain

$$\begin{aligned} \pi^L \Delta_O^L &= \left(q_g - \pi^L q_g\right) + \frac{q_1 \phi_1 - \pi^L q_1}{2} + \frac{q_2 \phi_2 - \pi^L q_2}{2}, \\ \pi^L \Delta_S^L &= \left(q_2 \phi_2 - \pi^L q_2\right) - \left(q_1 \phi_1 - \pi^L q_1\right) = (q_2 \phi_2 - q_1 \phi_1) - \pi^L (q_2 - q_1). \end{aligned}$$

Inserting these terms into the expression from Step 1 and rearranging gives the claim.

³⁰Result (iii) follows because $V(\beta)$ is independent of β in this case.

Step 3: Near the extremes, $V'(R) = q_1 A_1 \phi'_1(R) + q_2 A_2 \phi'_2(R)$.

Proof of Step 3: Differentiating the expression obtained in Step 2 gives

$$\begin{aligned} V'(R) &= q_1 \left(\frac{1}{2} + \mu \right) \phi'_1(R) + q_2 \left(\frac{1}{2} - \mu \right) \phi'_2(R) \\ &\quad - \left[q_g + \frac{q_1 + q_2}{2} - \mu(q_2 - q_1) \right] (q_1 \phi'_1(R) + q_2 \phi'_2(R)). \end{aligned}$$

The claim follows from this expression using

$$\begin{aligned} \frac{1}{2} + \mu - \left[q_g + \frac{q_1 + q_2}{2} - \mu(q_2 - q_1) \right] &= \frac{q_b - q_g}{2} + \mu(q_2 - q_1) + \mu = A_1, \\ \frac{1}{2} - \mu - \left[q_g + \frac{q_1 + q_2}{2} - \mu(q_2 - q_1) \right] &= \frac{q_b - q_g}{2} + \mu(q_2 - q_1) - \mu = A_2. \end{aligned}$$

Step 4: $V'(0+) = -q_1 A_1 f_- - q_2 A_2 f_+$ and $V'(1-) = -q_1 A_1 f_+ - q_2 A_2 f_-$.

Proof of Step 4: Since $\phi'_1(R) = -f(R - \frac{1}{2})$ and $\phi'_2(R) = -f(\frac{1}{2} - R)$,

$$V'(R) = -q_1 A_1 f(R - \frac{1}{2}) - q_2 A_2 f(\frac{1}{2} - R)$$

near the extremes. Let $f_- := f(-\frac{1}{2})$ and $f_+ := f(\frac{1}{2})$. Taking limits in the expression for $V'(R)$ gives the result.

Step 5: Condition (6) is equivalent to $V'(0+) > 0$ and $V'(1-) < 0$.

Proof of Step 5: Since Assumption 2 implies $f(\frac{1}{2}) > 0$, we have

$$V'(0+) > 0 \iff q_1 A_1 f(-\frac{1}{2}) + q_2 A_2 f(\frac{1}{2}) < 0,$$

and

$$V'(1-) < 0 \iff q_1 A_1 f(\frac{1}{2}) + q_2 A_2 f(-\frac{1}{2}) > 0.$$

Thus condition (6) is equivalent to $V'(0+) > 0$ and $V'(1-) < 0$.

Step 6: $V(R)$ has an interior maximum $R^* \in (0, 1)$.

Proof of Step 6: According to Step 5, condition (6) implies $V'(0+) > 0$ and $V'(1-) < 0$.

Hence, for any decision environment $\mathcal{E}$, there exists $\delta(\mathcal{E}) > 0$ such that V is strictly increasing on $(0, \delta(\mathcal{E}))$ and strictly decreasing on $(1 - \delta(\mathcal{E}), 1)$. Thus the supremum of V over $(0, 1)$ equals the supremum of V over some compact interval $[\delta(\mathcal{E}), 1 - \delta(\mathcal{E})] \subset (0, 1)$.

Since V is continuous in R ,³¹ the extreme value theorem implies that $V(R)$ attains its maximum on $[\delta(\mathcal{E}), 1 - \delta(\mathcal{E})]$, and hence on $(0, 1)$, at some $R^* \in (0, 1)$.

A.2.5 Universal Obedience in Example 1

Recall that, for the economic analysis, we maintain the assumptions that $R \in (0, 1)$ and $q_s > 0$ for s , implying $Q < 1/2$. For continuity and convexity arguments, we occasionally extend functions of R and Q to the closures $[0, 1]$ and $[0, 1/2]$.

The following two lemmas supply the exact cutoff $\bar{Q}(a, \sigma)$ invoked informally before Proposition 6, clarifying under which conditions the requirement that all receivers are obedient is satisfied.³² Lemma A.3 establishes existence, uniqueness, and an explicit characterization for $a > 1$. Lemma A.4 describes how the cutoff varies with a .

Lemma A.3 *Suppose we are in Example 1. Then, for every $(a, \sigma) \in \mathcal{R}_+ \times \mathcal{R}_+$, there exists a unique cutoff $\bar{Q}(a, \sigma) \in (0, 1/2]$ such that receivers are universally obedient for (Q, a, σ) if and only if $Q \leq \bar{Q}(a, \sigma)$. Moreover:*

(i) *For $a = 1$ and every $(Q, \sigma) \in (0, 1/2) \times \mathcal{R}_+$, receivers are universally obedient for (Q, a, σ) .*

(ii) *For $a > 1$ and $(Q, \sigma) \in (0, 1/2) \times \mathcal{R}_+$, receivers are universally obedient for (Q, a, σ) if and only if the following holds for all $R \in (0, 1)$:*

$$N(R; Q, a, \sigma) := q_g q_b - Q^2 + Q(q_b + Q)(1 - R)^a + Q(q_g + Q)R^a \geq 0.$$

(iii) *For $a < 1$, receivers are universally obedient if and only if the following holds for all $R \in (0, 1)$:*

$$\tilde{N}(R; Q, a, \sigma) := q_g q_b - Q^2 + Q(q_b + Q)(1 - R^a) + Q(q_g + Q)(1 - (1 - R)^a) \geq 0.$$

Proof. In Steps 1-7, we establish the lemma for $a \geq 1$. Steps 1 and 2 hold for all a . In Example 1, $F(i) = (i + 1/2)^a$ and $q_1 = q_2 = Q$, so $\phi_1(R) = 1 - R^a$ and $\phi_2(R) = (1 - R)^a$. To simplify notation, we set $s_a(R) := \phi_1(R) + \phi_2(R)$ and $b_a(R) := \phi_2(R) - \phi_1(R) = R^a + (1 - R)^a - 1$,

³¹Since payoffs are bounded and, for every type i , the receiver's ex ante payoff under standard R is a continuous function of R , the value function $V(R)$ is continuous by the dominated convergence theorem.

³²Whenever we write $\bar{Q}(a, \sigma) = 1/2$, this should be understood as a boundary statement: universal acceptance then holds for every admissible $Q < 1/2$.

Step 1. *We have*

$$\Delta_S^L(R) = \frac{Q b_a(R)}{\pi^L(R)}, \quad \Delta_O^L(R) = \frac{q_g + \frac{Q}{2} s_a(R)}{\pi^L(R)} - (q_g + Q).$$

Proof of Step 1: In Example 1, the probability of like is $\pi^L(R) = q_g + Q s_a(R) > 0$. The posteriors after a like are $P_g^L = q_g/\pi^L$, $p_1^L = Q\phi_1/\pi^L$, and $p_2^L = Q\phi_2/\pi^L$, so Step 1 follows from by Definition 3.

Step 2. (i) For $a > 1$, $\Delta_S^L(R) < 0$. (ii) For $a = 1$, $\Delta_S^L \equiv 0$, so that all types are universally obedient. (iii) For $a < 1$, $\Delta_S^L(R) > 0$.

Proof of Step 2: For $a > 1$, strict convexity of $x \mapsto x^a$ on $[0, 1]$ gives $b_a(R) < 0$ on $(0, 1)$, hence $\Delta_S^L(R) < 0$. For $a = 1$, $b_a \equiv 0$, so $\Delta_S^L \equiv 0$ and Lemma 1 implies that receivers are universally obedient; in particular $\bar{Q}(1, \sigma) = 1/2$, proving (ii). For $a < 1$, strict concavity of $x \mapsto x^a$ gives $b_a(R) > 0$, hence $\Delta_S^L(R) > 0$.

Step 3. For $a > 1$, the objective rating effect $\Delta_O^L(R) > 0$ is strictly positive for all $R \in [0, 1]$.

Proof of Step 3: Multiplying $\Delta_O^L(R)$ by $\pi^L(R)$, substituting $\pi^L = q_g + Q s_a$ and rearranging yields $\pi^L(R) \Delta_O^L(R) = q_g q_b + \frac{Q}{2} (s_a(R) q_b + (2 - s_a(R)) q_g)$. Since $0 \leq s_a(R) \leq 2$ and $q_g, q_b > 0$ for $Q < 1/2$, both summands are nonnegative and the first is strictly positive. Hence $\Delta_O^L(R) > 0$ for all $R \in (0, 1)$, and by continuity the same inequality extends to the boundary points $R = 0$ and $R = 1$.

Step 4. For (Q, a, σ) with $a > 1$, receivers are universally obedient if and only if $N(R; Q, a, \sigma) \geq 0$ for all $R \in (0, 1)$.

Proof of Step 4: By Corollary 1, all receiver types are obedient for the standard R if and only if $|\Delta_S^L(R)| \leq 2\Delta_O^L(R)$. Since $\Delta_S^L < 0$ for $a > 1$, this is equivalent to $\Delta_O^L(R) + \frac{1}{2}\Delta_S^L(R) \geq 0$. Multiplying by $\pi^L(R) > 0$ and substituting the closed forms from Step 1 gives

$$\pi^L(R) (\Delta_O^L(R) + \frac{1}{2}\Delta_S^L(R)) = q_g + \frac{Q}{2} (s_a(R) + b_a(R)) - (q_g + Q)\pi^L(R) \geq 0$$

Using $s_a(R) + b_a(R) = 2(1-R)^a$ and $\pi^L(R) = q_g + Qs_a(R) = q_g + Q(1-R^a + (1-R)^a)$, this rearranges to $\pi^L(R)(\Delta_O^L(R) + \frac{1}{2}\Delta_S^L(R)) = N(R; Q, a, \sigma) \geq 0$. Therefore, for (Q, a, σ) receivers are universally obedient if and only if $N(R; Q, a, \sigma) \geq 0$ for all $R \in (0, 1)$. Because N extends continuously to $[0, 1]$, this is equivalent to $N(R; Q, a, \sigma) \geq 0$ for all $R \in [0, 1]$.

Step 5. For all $(a, \sigma) \in (1, \infty) \times \mathcal{R}_+$, there exists $\varepsilon(a, \sigma) > 0$ such that all receivers are obedient for all (a, σ, Q) such that $Q \leq \varepsilon(a, \sigma)$ and all $R \in (0, 1)$.

Proof of Step 5: Since $|b_a(R)|$ is continuous on the compact interval $[0, 1]$, $C_a := \max_{R \in [0, 1]} |b_a(R)| < \infty$. Using $\pi^L(R) \geq q_g$ from Step 1, $|\Delta_S^L(R)| \leq QC_a/q_g$ for every R . From Step 3, $\pi^L \Delta_O^L \geq q_g q_b$, and $\pi^L(R) \leq q_g + 2Q$ since $s_a \leq 2$, so $\Delta_O^L(R) \geq q_g q_b / (q_g + 2Q)$. As $Q \downarrow 0$,

$$\frac{QC_a}{q_g} \rightarrow 0, \quad \frac{2q_g q_b}{q_g + 2Q} \rightarrow \frac{2}{1 + \sigma} > 0.$$

Hence there exists $\varepsilon(a, \sigma) > 0$ such that $|\Delta_S^L(R)| < 2\Delta_O^L(R)$ uniformly in $R \in [0, 1]$ for every $Q \leq \varepsilon(a, \sigma)$. By Corollary 1, receivers are universally obedient for every such Q, a, σ .

Step 6. Define $\mathcal{S}(a, \sigma) := \{Q \in [0, 1/2) \mid \text{all } i \text{ are universally obedient at } Q(a, \sigma)\}$ as the obedience set. Then, for $a > 1$, either $\mathcal{S}(a, \sigma) = [0, \bar{Q}(a, \sigma)]$ for some $\bar{Q}(a, \sigma) \in (0, 1/2)$ or $\mathcal{S}(a, \sigma) = [0, 1/2)$.

Proof of Step 6

$\mathcal{S}(a, \sigma)$ is nonempty: More generally, by Step 5, $[0, \varepsilon(a, \sigma)] \subseteq \mathcal{S}(a, \sigma)$.

$\mathcal{S}(a, \sigma)$ is downward closed. By Corollary 1, all receivers are obedient for the standard R if and only if $\tilde{i}(R, Q) \leq -1/2$. Writing $\delta := (\sigma - 1)/(\sigma + 1) \in (-1, 1)$, direct differentiation at fixed σ (taking the dependence of $q_g = (1 - 2Q)\sigma/(1 + \sigma)$ and $q_b = (1 - 2Q)/(1 + \sigma)$ on Q into account) yields

$$\frac{\partial \tilde{i}(R, Q)}{\partial Q} = - \frac{Q^2 \delta (R^a - (1 - R)^a + \delta) + \frac{1 - \delta^2}{4}}{Q^2 b_a(R)}. \quad (11)$$

Letting $x := R^a - (1 - R)^a \in [-1, 1]$, the bound $\delta(x + \delta) \geq -|\delta|(1 - |\delta|)$ gives

$$Q^2 \delta(x + \delta) + \frac{1 - \delta^2}{4} \geq (1 - |\delta|) \left(\frac{1 + |\delta|}{4} - Q^2 |\delta| \right) > 0$$

for $Q < 1/2$, so that the numerator in (11) is strictly positive. For $a > 1$, Step 2 gives $b_a(R) < 0$ on $(0, 1)$, so $\partial \tilde{i}(R, Q)/\partial Q > 0$: $\tilde{i}(R, Q)$ is strictly increasing in Q . Hence, if $\tilde{i}(R, Q) \leq -1/2$ at some Q , the same inequality holds at every $Q' \leq Q$; equivalently, $\mathcal{S}(a, \sigma)$ is downward closed.

$\mathcal{S}(a, \sigma)$ is relatively closed in $[0, 1/2)$: Define $G(R, Q) := 2\Delta_O^L(R, Q) - |\Delta_S^L(R, Q)|$. As G is continuous on $[0, 1] \times [0, 1/2)$, $m(Q) := \min_{R \in [0, 1]} G(R, Q)$ exists for every $q \in [0, 1/2)$ by compactness and is continuous in Q . Hence $\mathcal{S}(a, \sigma) = \{Q \in [0, 1/2) : m(Q) \geq 0\}$ is relatively closed in $[0, 1/2)$.

Conclusion. Combining non-emptiness, downward closedness and relative closedness, Step 6 follows. In both cases, $\bar{Q}(a, \sigma) := \sup \mathcal{S}(a, \sigma) \in (0, 1/2]$.

Step 7. $\bar{Q}(a, \sigma) < 1/2$ for $a > 1$.

Proof of Step 7: To rule out the possibility that the cutoff reaches the boundary $1/2$, we evaluate the continuous extension of N at the boundary point $Q = 1/2$. At $Q = 1/2$, $q_g = q_b = 0$, so $N(R; 1/2, a, \sigma) = -\frac{1}{4} + \frac{1}{4}(R^a + (1 - R)^a)$. By strict convexity of $x \mapsto x^a$ for $a > 1$, $R^a + (1 - R)^a < 1$ on $(0, 1)$, hence $N(R; 1/2, a, \sigma) < 0$ for all $R \in (0, 1)$. Now fix any $R_0 \in (0, 1)$. Since N is continuous in (R, Q) , there exists $\delta > 0$ such that $N(R_0; Q, a, \sigma) < 0$ for every $Q \in (1/2 - \delta, 1/2)$. By Step 4, universal obedience fails at every such Q , so $\bar{Q}(a, \sigma) \leq 1/2 - \delta < 1/2$.

Putting together Steps 1-7 implies the result for $a > 1$.

Step 8. The proposition also holds for $a < 1$.

Sketch of proof: The case $a < 1$ is analogous: The main difference is that the sign of Δ_S^L changes by Step 2. Therefore, the obedience condition $|\Delta_S^L(R)| \leq 2\Delta_O^L(R)$ becomes $\Delta_O^L - \frac{1}{2}\Delta_S^L \geq 0$, and the resulting polynomial is again strictly convex on $[0, 1]$ with strictly positive boundary values, so the same chain of arguments delivers a unique cutoff $\bar{Q}(a, \sigma) < 1/2$. ■

The universal obedience conditions in Lemma A.3 are formulated as requirements that,

for all $R \in (0, 1)$, $N(R; Q, a, \sigma) \geq 0$ and $\tilde{N}(R; Q, a, \sigma) \geq 0$ must hold, respectively. We now show that it is sufficient that the requirements hold at specific values of R .

Corollary A.1 *Suppose Example 1 holds.*

- (i) *For $a > 1$ and $(Q, \sigma) \in (0, 1/2) \times \mathcal{R}_+$, receivers are universally obedient for (Q, a, σ) if and only if $N(R_{\min}(Q, a, \sigma); Q, a, \sigma) \geq 0$, where*

$$R_{\min}(Q, a, \sigma) := \frac{\left((q_b + Q)/(q_g + Q)\right)^{1/(a-1)}}{1 + \left((q_b + Q)/(q_g + Q)\right)^{1/(a-1)}}.$$

- (ii) *For $a < 1$ and $(Q, \sigma) \in (0, 1/2) \times \mathcal{R}_+$, receivers are universally obedient for (Q, a, σ) if and only if $\tilde{N}(\tilde{R}_{\min}(Q, a, \sigma); Q, a, \sigma) \geq 0$, where*

$$\tilde{R}_{\min}(Q, a, \sigma) := \frac{1}{1 + \left((q_b + Q)/(q_g + Q)\right)^{1/(a-1)}}.$$

Proof. (i) Simple calculations show that

$$N''(R) = a(a-1)Q[(q_b + Q)(1-R)^{a-2} + (q_g + Q)R^{a-2}] > 0 \quad \text{for } R \in (0, 1).$$

Thus, $N(\cdot; Q, a, \sigma)$ is strictly convex on $[0, 1]$. Direct evaluation gives $N(0; Q, a, \sigma) = q_b(q_g + Q) > 0$ and $N(1; Q, a, \sigma) = q_g(q_b + Q) > 0$. Setting $N'(R) = 0$ yields $(q_g + Q)R_{\min}^{a-1} = (q_b + Q)(1 - R_{\min})^{a-1}$, so $R_{\min}$ is the expression in the statement. Since N is strictly convex with positive boundary values, $N \geq 0$ on $[0, 1]$ if and only if $N(R_{\min}; Q, a, \sigma) \geq 0$. The result thus follows from Lemma A.3.

- (ii) For $a < 1$, direct differentiation gives

$$\tilde{N}''(R) = aQ(1-a)[(q_b + Q)R^{a-2} + (q_g + Q)(1-R)^{a-2}] > 0 \quad \text{for } R \in (0, 1),$$

so $\tilde{N}(\cdot; Q, a, \sigma)$ is strictly convex on $[0, 1]$. Direct evaluation gives $\tilde{N}(0; Q, a, \sigma) = q_b(q_g + Q) > 0$ and $\tilde{N}(1; Q, a, \sigma) = q_g(q_b + Q) > 0$. Setting $\tilde{N}'(R) = 0$ yields $(q_g + Q)(1 - \tilde{R}_{\min})^{a-1} = (q_b + Q)\tilde{R}_{\min}^{a-1}$, so $\tilde{R}_{\min}$ is the expression in the statement. Strict convexity with

positive boundary values again implies $\tilde{N} \geq 0$ on $[0, 1]$ if and only if $\tilde{N}(\tilde{R}_{\min}; Q, a, \sigma) \geq 0$. The result thus follows from Lemma A.3. ■

A particularly simple case arises when $\sigma = 1$.

Corollary A.2 *For $\sigma = 1$, receivers are universally obedient at (Q, a, σ) if and only if $Q \leq \bar{Q}(a, 1)$, where*

$$\bar{Q}(a, 1) = \begin{cases} 2^{a-2}, & 0 < a < 1, \\ 1/2, & a = 1, \\ 1/[4(1 - 2^{-a})], & a > 1. \end{cases}$$

Proof. The result follows from Corollary A.1 after noting that at $\sigma = 1$ we have $R_{\min} = \tilde{R}_{\min} = 1/2$. ■

To illustrate the result, note that $\bar{Q}(2, 1) = 1/3$, so that, if $a = 2$, obedience arises for quite high prevalence of controversial products.

Lemma A.4 *In Example 1, the cutoff $\bar{Q}(a, \sigma)$ from Lemma A.3 is strictly increasing in a on $(0, 1)$, attains its maximum $\bar{Q}(1, \sigma) = 1/2$ at $a = 1$, and is strictly decreasing in a on $(1, \infty)$.*

Proof. The value at $a = 1$ is Lemma A.3(i). We first prove strict monotonicity on $(1, \infty)$; we then give the corresponding argument on $(0, 1)$ using the polynomial $\tilde{N}$ from Lemma A.3(iii). Throughout, we use the proof of Step 1 in Lemma A.3 that, for each $a > 1$, the acceptance set $\mathcal{S}(a, \sigma) \subset [0, 1/2)$ is nonempty, downward closed, and relatively closed, and that $\bar{Q}(a, \sigma) < 1/2$.

Step 1: $N(R; Q, a, \sigma)$ is strictly decreasing in a for fixed $R \in (0, 1)$.

Proof of Step 1: By Lemma A.3(ii), for $a > 1$ receivers are universally obedient at (Q, a, σ) if and only if $N(R; Q, a, \sigma) \geq 0$ for all $R \in (0, 1)$; by continuity this is equivalent to the same condition on $[0, 1]$. Since q_g and q_b depend only on (Q, σ) , only the terms R^a

and $(1 - R)^a$ in N depend on a . Hence

$$\frac{\partial N(R; Q, a, \sigma)}{\partial a} = Q(q_b + Q)(1 - R)^a \ln(1 - R) + Q(q_g + Q)R^a \ln R < 0 \quad \text{for } R \in (0, 1),$$

because the prefactors are strictly positive while $\ln R < 0$ and $\ln(1 - R) < 0$.

Step 2: Suppose $a_2 > a_1 > 1$. Then $\bar{Q}(a_2, \sigma) \leq \bar{Q}(a_1, \sigma)$.

Proof of Step 2: Take $a_2 > a_1 > 1$ and any $Q \in \mathcal{S}(a_2, \sigma)$; by Lemma A.3, equivalently any admissible $Q \in [0, \bar{Q}(a_2, \sigma)]$. Then $N(R; Q, a_2, \sigma) \geq 0$ for all $R \in (0, 1)$. By Step 1, $N(R; Q, a_1, \sigma) > N(R; Q, a_2, \sigma) \geq 0$ for every $R \in (0, 1)$. At the auxiliary boundary points $R = 0$ and $R = 1$, the values $N(0; Q, a, \sigma) = q_b(q_g + Q)$ and $N(1; Q, a, \sigma) = q_g(q_b + Q)$ (computed in Step 6 of the proof of Lemma A.3) are independent of a and strictly positive, so $N(\cdot; Q, a_1, \sigma) \geq 0$ on $[0, 1]$. Hence $Q \in \mathcal{S}(a_1, \sigma)$, which gives $\bar{Q}(a_2, \sigma) \leq \bar{Q}(a_1, \sigma)$.

Step 3: Suppose $a_2 > a_1 > 1$. Then $\bar{Q}(a_2, \sigma) < \bar{Q}(a_1, \sigma)$.

Proof of Step 3: Suppose, toward a contradiction, that $\bar{Q}(a_1, \sigma) = \bar{Q}(a_2, \sigma) =: \bar{Q}$. By Lemma A.3(ii), $\bar{Q} < 1/2$, and by the relative closedness of $\mathcal{S}(a_2, \sigma)$ in the admissible domain $[0, 1/2)$ established in Step 7 of the previous proof, $\bar{Q} \in \mathcal{S}(a_2, \sigma)$. Hence $N(R; \bar{Q}, a_2, \sigma) \geq 0$ for all $R \in [0, 1]$. By Step 1, for every $R \in (0, 1)$, $N(R; \bar{Q}, a_1, \sigma) > N(R; \bar{Q}, a_2, \sigma) \geq 0$. At the auxiliary boundary points, $N(0; \bar{Q}, a_1, \sigma) = q_b(q_g + \bar{Q})$ and $N(1; \bar{Q}, a_1, \sigma) = q_g(q_b + \bar{Q})$, both of which are strictly positive since $\bar{Q} < 1/2$ implies $q_g, q_b > 0$. Therefore, $m := \min_{R \in [0, 1]} N(R; \bar{Q}, a_1, \sigma) > 0$. Choose any $\delta > 0$ with $\bar{Q} + \delta < 1/2$. The function N is uniformly continuous on the compact set $[0, 1] \times [\bar{Q}, \bar{Q} + \delta]$ (with a_1 and σ fixed), so there exists $\varepsilon \in (0, \delta]$ such that

$$|N(R; Q, a_1, \sigma) - N(R; \bar{Q}, a_1, \sigma)| < m \quad \text{for all } R \in [0, 1] \text{ and all } Q \in [\bar{Q}, \bar{Q} + \varepsilon].$$

For such Q , $N(R; Q, a_1, \sigma) > 0$ for all $R \in [0, 1]$, so receivers are universally obedient at (Q, a_1, σ) . This contradicts the definition of $\bar{Q}(a_1, \sigma)$ as $\sup \mathcal{S}(a_1, \sigma)$. We conclude $\bar{Q}(a_2, \sigma) < \bar{Q}(a_1, \sigma)$.

Step 4: $\bar{Q}(a, \sigma)$ is strictly increasing in a on $(0, 1)$.

Proof of Step 4: By Lemma A.3(iii), for $a < 1$ universal acceptance at (Q, a, σ) is equivalent to $\tilde{N}(R; Q, a, \sigma) \geq 0$ for all $R \in (0, 1)$, where

$$\tilde{N}(R; Q, a, \sigma) := q_g q_b - Q^2 + Q(q_b + Q)(1 - R^a) + Q(q_g + Q)(1 - (1 - R)^a).$$

For fixed $R \in (0, 1)$, only the terms R^a and $(1 - R)^a$ depend on a , so

$$\frac{\partial \tilde{N}(R; Q, a, \sigma)}{\partial a} = -Q(q_b + Q)R^a \ln R - Q(q_g + Q)(1 - R)^a \ln(1 - R) > 0,$$

since $\ln R < 0$ and $\ln(1 - R) < 0$. Hence, if $0 < a_1 < a_2 < 1$ and $Q \in \mathcal{S}(a_1, \sigma)$, then $\tilde{N}(R; Q, a_1, \sigma) \geq 0$ for all $R \in (0, 1)$, and Step 1's strict inequality yields $\tilde{N}(R; Q, a_2, \sigma) > \tilde{N}(R; Q, a_1, \sigma) \geq 0$. The boundary values $\tilde{N}(0; Q, a, \sigma) = q_b(q_g + Q)$ and $\tilde{N}(1; Q, a, \sigma) = q_g(q_b + Q)$ are independent of a and strictly positive, so $\tilde{N}(\cdot; Q, a_2, \sigma) \geq 0$ on $[0, 1]$, giving $Q \in \mathcal{S}(a_2, \sigma)$. Thus $\bar{Q}(a_1, \sigma) \leq \bar{Q}(a_2, \sigma)$. The strict-inequality argument from Step 3 carries over verbatim with $\tilde{N}$ in place of N (using $\partial \tilde{N} / \partial a > 0$ instead of $\partial N / \partial a < 0$, and the relative closedness of $\mathcal{S}(a_1, \sigma)$). From this, Step 4 follows. ■

Thus the permissible prevalence of controversial products is largest at the symmetric benchmark $a = 1$ and shrinks as the population becomes more polarized in either direction.

A.2.6 Proof of Proposition 6

Throughout this proof, we use the identities $q_g = (1 - 2Q)\sigma / (1 + \sigma)$ and $q_b = (1 - 2Q) / (1 + \sigma)$. Since Proposition 6 assumes $Q \leq \bar{Q}(a, \sigma)$, Lemma A.3 implies that all receiver types are obedient for every $R \in (0, 1)$. Hence Proposition 3(i) applies throughout, yielding $V(R) = \pi^L[\Delta_O^L - \Delta_S^L \mathbb{E}[i]]$. In Example 1, $\phi_1(R) = 1 - R^a$, $\phi_2(R) = (1 - R)^a$, and $\mathbb{E}[i] = \frac{a-1}{2(a+1)}$, so that

$$\begin{aligned} V(R) = & \frac{(1 - 2Q)\sigma(Q(\sigma - 1) + 1)}{(\sigma + 1)^2} + \frac{Q}{a + 1}(1 - R^a) \left(a - (a + 1) \left(\frac{(1 - 2Q)\sigma}{1 + \sigma} + Q \right) \right) \\ & + \frac{Q}{a + 1}(1 - R)^a \left(1 - (a + 1) \left(\frac{(1 - 2Q)\sigma}{1 + \sigma} + Q \right) \right) \end{aligned}$$

$$= c_0 + \frac{Q}{a+1} [c_1(a)(1-R^a) + c_2(a)(1-R)^a]$$

with $c_0 = \frac{(1-2Q)\sigma(Q(\sigma-1)+1)}{(\sigma+1)^2}$, $c_1(a) = a + (a+1)\frac{Q(\sigma-1)-\sigma}{\sigma+1}$ and $c_2(a) = 1 + (1+a)\frac{Q(\sigma-1)-\sigma}{\sigma+1}$.

We thus obtain

$$\begin{aligned} V'(R) &= -\frac{Qa}{a+1} [R^{a-1}c_1(a) + (1-R)^{a-1}c_2(a)], \\ V''(R) &= -\frac{Qa(a-1)}{a+1} [R^{a-2}c_1(a) - (1-R)^{a-2}c_2(a)], \end{aligned}$$

Since $R^{a-1} > 0$ and $(1-R)^{a-1} > 0$ on $(0, 1)$ and $Qa/(a+1) > 0$, the sign of $V'(R)$ is determined by the signs of $c_1(a)$ and $c_2(a)$, which can be written as

$$\begin{aligned} c_1(a) &= \frac{\sigma[Q(a+1) - 1] + a - Q(a+1)}{\sigma+1}, \\ c_2(a) &= \frac{\sigma[Q(a+1) - a] + 1 - Q(a+1)}{\sigma+1}. \end{aligned}$$

Solving $c_1(a) = 0$ and $c_2(a) = 0$ gives $\sigma_1^* := \frac{a-Q(a+1)}{1-Q(a+1)}$ and $\sigma_2^* := \frac{1-Q(a+1)}{a-Q(a+1)}$. If $Q < \min\{a, 1\}/(a+1)$, then the numerators of $c_1(a)$ and $c_2(a)$ are strictly decreasing affine functions of σ with positive intercepts. Therefore

$$c_1(a), c_2(a) \geq 0 \iff \sigma \leq \min\{\sigma_1^*, \sigma_2^*\},$$

$$c_1(a), c_2(a) \leq 0 \iff \sigma \geq \max\{\sigma_1^*, \sigma_2^*\}.$$

Hence, if $Q < \min\{a, 1\}/(a+1)$, V is decreasing exactly if $\sigma \leq \min\{\sigma_1^*, \sigma_2^*\}$ and increasing exactly if $\sigma \geq \max\{\sigma_1^*, \sigma_2^*\}$.

If instead $Q \geq \min\{a, 1\}/(a+1)$, these two monotone cases cannot occur. To see this, first suppose $a > 1$, so that the cutoff equals $1/(a+1)$. Since $Q < 1/2 < a/(a+1)$, we have $Q(a+1) - 1 \geq 0$ and $a - Q(a+1) > 0$, and hence $c_1(a) > 0$ for every $\sigma > 0$. Moreover, $Q(a+1) - a < 0$ and $1 - Q(a+1) \leq 0$, so that $c_2(a) < 0$ for every $\sigma > 0$. Thus $c_1(a) > 0 > c_2(a)$.

Now suppose $a < 1$, so that the cutoff equals $a/(a+1)$. Since $Q < 1/2 < 1/(a+1)$, we have $Q(a+1) - 1 < 0$ and $a - Q(a+1) \leq 0$ and hence $c_1(a) < 0$ for every $\sigma > 0$.

Moreover, $Q(a+1) - a \geq 0$, and $1 - Q(a+1) > 0$, so $c_2(a) > 0$ for every $\sigma > 0$. Thus $c_1(a) < 0 < c_2(a)$. Therefore, at and above the cutoff $\min\{a, 1\}/(a+1)$, the coefficients have opposite signs for every $\sigma > 0$. Together with the case $Q < \min\{a, 1\}/(a+1)$ and

$$\min\{\sigma_1^*, \sigma_2^*\} < \sigma < \max\{\sigma_1^*, \sigma_2^*\},$$

this is exactly the region where $c_1(a)$ and $c_2(a)$ have opposite signs. Since

$$c_1(a) - c_2(a) \begin{cases} < 0 & \text{if } a < 1, \\ > 0 & \text{if } a > 1, \end{cases}$$

the mixed-sign case means $c_1(a) > 0 > c_2(a)$ when $a > 1$, and $c_1(a) < 0 < c_2(a)$ when $a < 1$. In either case,

$$V''(R) = -\frac{Qa(a-1)}{a+1} [R^{a-2}c_1(a) - (1-R)^{a-2}c_2(a)] < 0 \quad \text{for all } R \in (0, 1),$$

so V is strictly concave on $(0, 1)$. To conclude that the maximum is interior, we also check the derivatives at $R = 0$ and $R = 1$. If $a > 1$ and $c_1(a) > 0 > c_2(a)$, then

$$V'(0+) = -\frac{Qa}{a+1}c_2(a) > 0, \quad V'(1-) = -\frac{Qa}{a+1}c_1(a) < 0.$$

If $a < 1$ and $c_1(a) < 0 < c_2(a)$, then $R^{a-1}c_1(a) \rightarrow -\infty$ as $R \downarrow 0$ while $(1-R)^{a-1}c_2(a) \rightarrow c_2(a)$, so $V'(0+) > 0$; similarly, $(1-R)^{a-1}c_2(a) \rightarrow +\infty$ as $R \uparrow 1$, so $V'(1-) < 0$. Hence V increases near 0 and decreases near 1, and strict concavity implies a unique interior maximizer $R^* \in (0, 1)$.

Combining the cases $Q \geq \min\{a, 1\}/(a+1)$ and $Q < \min\{a, 1\}/(a+1)$ proves Proposition 6.

A.3 Sales Maximization

A.3.1 Proof of Proposition 8

We show $D(R) \leq D(0)$ for all $R > 0$. Let $\hat{F}(R) \in [0, 1]$ denote the fraction of receiver types that buy a product with a like at standard R . Then the claim is equivalent to

$$\pi^L(R)\hat{F}(R) + (1 - \pi^L(R))(1 - \hat{F}(R)) \leq q_g + q_1 + q_2. \quad (12)$$

Fix $R > 0$. Then the left-hand side of (12) equals

$$\pi^L(R)\hat{F}(R) + (1 - \pi^L(R))(1 - \hat{F}(R)) = (1 - \pi^L(R)) + (2\pi^L(R) - 1)\hat{F}(R).$$

As a function of $\hat{F}(R) \in [0, 1]$ this expression is affine, hence it is bounded above by its endpoint values:

$$\pi^L(R)\hat{F}(R) + (1 - \pi^L(R))(1 - \hat{F}(R)) \leq \max\{1 - \pi^L(R), \pi^L(R)\}.$$

Next, by definition $\pi^L(R) = q_g + q_1\phi_1(R) + q_2\phi_2(R)$ with $\phi_1, \phi_2 \in [0, 1]$, so

$$\pi^L(R) \leq q_g + q_1 + q_2 = D(0).$$

Moreover, since $\phi_1, \phi_2 \geq 0$, we also have $\pi^L(R) \geq q_g$. Under the assumption $q_g \geq q_b$ and using $q_g + q_1 + q_2 = 1 - q_b$, it follows that

$$1 - \pi^L(R) \leq 1 - q_g \leq 1 - q_b = q_g + q_1 + q_2 = D(0).$$

Therefore both $\pi^L(R)$ and $1 - \pi^L(R)$ are at most $D(0)$, implying

$$\max\{1 - \pi^L(R), \pi^L(R)\} \leq D(0).$$

Combining the inequalities yields (12), hence $D(R) \leq D(0)$ for all $R > 0$. Consequently, sales are maximized as $R \rightarrow 0$.

A.4 Further Results

A.4.1 Optimal Pricing

Proof of Proposition 9

The proof will follow from Steps 1-4 below.

Step 1. *The unique unrestricted profit maximizer is*

$$P^u(R) = \frac{q_g + 2Q(1 - R)}{4Q}.$$

Fix $R \in (0, 1)$ and consider prices $P \in [0, 1 - R]$. For a uniform type distribution and $q_1 = q_2 = Q$, a sender gives a like for each controversial variant with probability $1 - T = 1 - R - P$. Hence

$$\pi^L(R + P) = q_g + q_1(1 - R - P) + q_2(1 - R - P) = q_g + 2Q(1 - R - P).$$

Ignoring the demand constraint in (7), the firm's (unrestricted) profit is

$$\Pi^u(P; R) := P \pi^L(R + P) = P(q_g + 2Q(1 - R - P)).$$

This function is strictly concave in P (since $Q > 0$), and its first-order condition is

$$\frac{\partial \Pi^u(P; R)}{\partial P} = q_g + 2Q(1 - R - P) - 2QP = q_g + 2Q(1 - R) - 4QP = 0.$$

This first-order condition yields

$$P^u(R) = \frac{q_g + 2Q(1 - R)}{4Q}.$$

This is the unconstrained maximizer of the quadratic objective $\Pi^u(P; R)$. It need not lie in the admissible interval $[0, 1 - R]$, so at this stage it is not yet the solution to the constrained pricing problem.

Step 2. *The objective content of the rating system satisfies*

$$\Delta_O^L(R + P) = \frac{1}{2} - q_g - Q + \frac{q_g}{2(q_g + 2Q(1 - R - P))}.$$

Sales are positive if and only if the acceptance constraint in (7) holds, i.e. if and only if $P \leq \Delta_O^L(R + P)$. Under $q_1 = q_2 = Q$ and uniform types, $\Delta_S^L = 0$, so $\Delta_O^L(R + P) =$

$U^L(i) - U^0(i)$ is the common willingness to pay after a like for all types i . Moreover, the posterior probability of a good product after L is

$$p_g^L(R + P) = \frac{q_g}{q_g + 2Q(1 - R - P)},$$

and the posterior probabilities of the two controversial variants are equal. Plugging these posteriors into Definition 3(i) yields the result.

Step 3: *The acceptance gap $\gamma(P) := \Delta_O^L(R + P) - P$ is strictly decreasing on $[0, 1 - R]$.*

Differentiating γ gives

$$\gamma'(P) = \frac{q_g Q}{(q_g + 2Q(1 - R - P))^2} - 1.$$

For all $P \in [0, 1 - R]$ we have $q_g + 2Q(1 - R - P) \geq q_g$, hence

$$\frac{q_g Q}{(q_g + 2Q(1 - R - P))^2} \leq \frac{q_g Q}{q_g^2} = \frac{Q}{q_g} < 1,$$

where the strict inequality uses the assumption $q_g > Q$. Therefore $\gamma'(P) < 0$ on $[0, 1 - R]$.

Moreover, $\gamma(0) = \Delta_O^L(R) > 0$. To see this, note that

$$\Delta_O^L(R) = \frac{p_g^L(R) - q_g + q_b}{2}.$$

Since $R \in (0, 1)$ and $Q > 0$, we have $q_g + 2Q(1 - R) < q_g + 2Q = 1 - q_b < 1$, so $p_g^L(R) = q_g / (q_g + 2Q(1 - R)) > q_g$. Hence $p_g^L(R) - q_g + q_b > 0$, proving $\gamma(0) > 0$.

Therefore the feasible set

$$\left\{ P \in [0, 1 - R] : P \leq \Delta_O^L(R + P) \right\} = \left\{ P \in [0, 1 - R] : \gamma(P) \geq 0 \right\}$$

Since γ is continuous and strictly decreasing, this feasible set is a nonempty compact interval of the form $[0, \bar{P}(R)]$, where $\bar{P}(R) := \max\{P \in [0, 1 - R] : \gamma(P) \geq 0\}$. If $\gamma(1 - R) \leq 0$, then $\bar{P}(R)$ is the unique root of γ in $[0, 1 - R]$, characterized by $\gamma(\bar{P}(R)) = 0$, i.e. $\bar{P}(R) = \Delta_O^L(R + \bar{P}(R))$. If instead $\gamma(1 - R) > 0$, then $\bar{P}(R) = 1 - R$ and every price in $[0, 1 - R]$ is feasible.

Step 4: *The unique maximizer on $[0, \bar{P}(R)]$ is $P^*(R) = \min\{P^u(R), \bar{P}(R)\}$.*

From Step 3, the original problem reduces to maximizing the strictly concave quadratic $\Pi^u(P; R)$ over the interval $[0, \bar{P}(R)]$. Since $P^u(R)$ is the unconstrained maximizer from Step 1, the unique constrained maximizer is

$$P^*(R) = \begin{cases} P^u(R), & \text{if } P^u(R) \leq \bar{P}(R), \\ \bar{P}(R), & \text{if } P^u(R) > \bar{P}(R). \end{cases}$$

Equivalently,

$$P^*(R) = \min\{P^u(R), \bar{P}(R)\}.$$

Since $\bar{P}(R) \leq 1 - R$, this formula automatically respects the admissible price restriction $P \in [0, 1 - R]$.

A.4.2 Multiple ratings

We now derive some auxiliary results for the case of multiple ratings, before we prove Proposition 10.

After b likes and d dislikes, where $b + d > 0$, the posterior reads

$$\begin{aligned} p_g(b, d) &= \begin{cases} \frac{q_g}{q_g + q_1(\phi_1(R))^b + q_2(\phi_2(R))^b} & \text{if } d = 0 \\ 0 & \text{if } d > 0 \end{cases} \\ p_1(b, d) &= \frac{q_1(\phi_1(R))^b(1 - \phi_1(R))^d}{q_1(\phi_1(R))^b(1 - \phi_1(R))^d + q_2(\phi_2(R))^b(1 - \phi_2(R))^d} \\ p_2(b, d) &= \frac{q_2(\phi_2(R))^b(1 - \phi_2(R))^d}{q_1(\phi_1(R))^b(1 - \phi_1(R))^d + q_2(\phi_2(R))^b(1 - \phi_2(R))^d} \\ p_b(b, d) &= \begin{cases} \frac{q_b}{q_1(1 - \phi_1(R))^d + q_2(1 - \phi_2(R))^d + q_b} & \text{if } b = 0 \\ 0 & \text{if } b > 0 \end{cases} \end{aligned}$$

The formulas for p_1 and p_2 are stated for $b, d > 0$, where objectively good and bad products are ruled out. When $d = 0$ or $b = 0$, p_1 and p_2 follow from $p_g + p_1 + p_2 + p_b = 1$ and the ratio $p_1/p_2 = q_1\phi_1^b(1 - \phi_1)^d/[q_2\phi_2^b(1 - \phi_2)^d]$.

Lemma A.5 *Consider the model with multiple ratings. Suppose Assumption 1 holds and consider any standard $R \in (0, 1)$.*

(i) If the receiver obtains only likes (dislikes), then, as the number of ratings increases, her posterior belief converges to $p_g = 1$ ($p_b = 1$).

(ii) If the receiver obtains mixed ratings, then $p_1(b, d)/p_2(b, d) = q_1/q_2$ and $p_g(b, d) = p_b(b, d) = 0$ for any $b, d > 0$.

Proof. (i) Suppose the receiver gets only likes and that $b \rightarrow \infty$. Then, $\phi_1(R) = \phi_2(R) =: \phi$. Since F has full support and $R \in (0, 1)$, $\phi \in (0, 1)$, so that $\lim_{b \rightarrow \infty} (\phi_i(R))^b = 0$ for $i = 1, 2$. Hence, $\lim_{b \rightarrow \infty} p_g(b, 0) = 1$. The argument for dislikes is analogous.

(ii) For any $b, d > 0$, $p_g(b, d) = p_b(b, d) = 0$ obviously holds. Further, because $\phi_1(R) = \phi_2(R)$ by Assumption 1, $p_i(b, d) = \frac{q_i}{q_1 + q_2}$ for $i = 1, 2$, yielding statement (ii). ■

Proof of Proposition 10

The proof will follow from Steps 1-4 below. We use the notation $b(i) := \frac{q_1}{q_1 + q_2} \left(\frac{1}{2} + i \right) + \frac{q_2}{q_1 + q_2} \left(\frac{1}{2} - i \right) - U^0(i)$ for the receiver's gain from buying a controversial product after infinitely many ratings, relative to the outside option.

Step 1: *the value of the infinite-learning system is*

$$V_\infty = q_g(1 - q_g - Q) + 2Q \mathbb{E}[b(i)^+], \quad b(i)^+ := \max\{b(i), 0\}.$$

By Lemma A.5, infinite learning perfectly reveals objectively good and bad products, while mixed ratings identify a controversial product but not its variant. Hence, (i) good products are always bought, (ii) bad products are never bought, and (iii) controversial products are bought exactly by those types with $b(i) \geq 0$. Thus, the result holds.³³

Step 2: *The optimally designed single-rating system gives the value*

$$V_1^* = q_g(1 - q_g - Q) + 2Q \max\{\mathbb{E}[b(i)], 0\}.$$

Under Assumption 1, Proposition 4 implies that the optimal single-standard design is extreme:

³³The first term is the gain from correctly identifying objectively good products: a good product yields payoff 1, while the outside option yields $\mathbb{E}[U^0(i)] = q_g + Q$ under symmetric F , so the gain is $q_g(1 - q_g - Q)$. Objectively bad products contribute zero because they are identified and not bought.

- if $\sigma < 1$, the designer chooses $R \rightarrow 0$, so all controversial products are bought;
- if $\sigma > 1$, the designer chooses $R \rightarrow 1$, so no controversial product is bought;
- if $\sigma = 1$, all standards are optimal.

Hence the benchmark value is $V_1^* = q_g(1 - q_g - Q) + 2Q \max\{\mathbb{E}[b(i)], 0\}$.

Step 3: $V_\infty = V_1^* \iff b(i)$ has a constant sign almost surely.

Since $b = b^+ - b^-$ with $b^- := \max\{-b, 0\}$, we have $\mathbb{E}[b^+] \geq \max\{\mathbb{E}[b], 0\}$. Moreover,

- if $\mathbb{E}[b] \geq 0$, then $\mathbb{E}[b^+] - \mathbb{E}[b] = \mathbb{E}[b^-]$, so equality requires $b^- = 0$ a.s.;
- if $\mathbb{E}[b] \leq 0$, then $\mathbb{E}[b^+] = 0$ requires $b^+ = 0$ a.s.

Thus the claim holds.

Step 4: $b(i)$ has a constant sign almost surely if and only if one of the following holds:

$$(i) \lambda = 1; \quad (ii) \lambda > 1 \text{ and } \sigma \notin \left(\frac{1}{\lambda}, \lambda\right); \quad (iii) \lambda < 1 \text{ and } \sigma \notin \left(\lambda, \frac{1}{\lambda}\right).$$

Using $q_1 = \frac{2Q\lambda}{\lambda+1}$, $q_2 = \frac{2Q}{\lambda+1}$, and $q_g = \frac{(1-2Q)\sigma}{1+\sigma}$:

$$b(i) = \frac{(1-2Q)(\lambda-1)}{\lambda+1} (i - \tilde{i}_\infty), \quad \tilde{i}_\infty = \frac{\sigma-1}{2(1+\sigma)} \frac{\lambda+1}{\lambda-1}.$$

Case 1: $\lambda = 1$. Then $b(i)$ is constant in i , so it has a constant sign a.s. Hence $V_\infty = V_1^*$.

Case 2: $\lambda > 1$. Then $b(i)$ is strictly increasing in i . Because F has full support on $[-1/2, 1/2]$, $b(i)$ has a constant sign a.s. iff $\tilde{i}_\infty \leq -1/2$ or $\tilde{i}_\infty \geq 1/2$. The first inequality is equivalent to $\sigma \leq 1/\lambda$, and the second to $\sigma \geq \lambda$. Hence $V_\infty = V_1^* \iff \sigma \notin (1/\lambda, \lambda)$.

Case 3: $\lambda < 1$. Then $b(i)$ is strictly decreasing in i . By the same argument, $b(i)$ has a constant sign a.s. iff $\tilde{i}_\infty \geq 1/2$ or $\tilde{i}_\infty \leq -1/2$. Now $\lambda - 1 < 0$, so the direction of the inequalities reverses: $\tilde{i}_\infty \geq 1/2$ gives $\sigma \leq \lambda$, and $\tilde{i}_\infty \leq -1/2$ gives $\sigma \geq 1/\lambda$. Hence $V_\infty = V_1^* \iff \sigma \notin (\lambda, 1/\lambda)$.

At the boundary values (e.g., $\sigma = 1/\lambda$ with $\lambda > 1$), $\tilde{i}_\infty = -1/2$, so $b(-1/2) = 0$ and $b(i) > 0$ for $i > -1/2$. Under full support with continuous F , zero at a single point has

measure zero, so $b \geq 0$ a.s. The boundary belongs to the equality region, yielding the open intervals in the statement.

A.4.3 Distinct Sender and Receiver Distributions

We prove the following result:

Proposition A.1 *Consider the model with distinct preference distributions of senders and receivers. Suppose the sender distribution F satisfies Assumption 1. Then, all types are obedient. Further, $V(R) = \pi_F^L[\Delta_{O,F}^L - \Delta_{S,F}^L \mathbb{E}_G[i]]$ is*

- (i) *decreasing in R if $\sigma < \frac{Q+(q_1-q_2)\mathbb{E}_G[i]}{Q-(q_1-q_2)\mathbb{E}_G[i]}$,*
- (ii) *increasing in R if $\sigma > \frac{Q+(q_1-q_2)\mathbb{E}_G[i]}{Q-(q_1-q_2)\mathbb{E}_G[i]}$,*
- (iii) *constant in R if $\sigma = \frac{Q+(q_1-q_2)\mathbb{E}_G[i]}{Q-(q_1-q_2)\mathbb{E}_G[i]}$.*

Proof.

The value of the rating system can be written as

$$V(\mathcal{R}) = \begin{cases} \int_{-1/2}^{1/2} (V_{A,F}(i) - V_{0,F}(i)) dG(i) & \text{if } |\Delta_{S,F}^L| \leq 2\Delta_{O,F}^L \\ \int_{-1/2}^{\tilde{i}} V_{N,F}(i) dG(i) + \int_{\tilde{i}}^{1/2} V_{A,F}(i) dG(i) - \int_{-1/2}^{1/2} V_{0,F}(i) dG(i) & \text{if } \Delta_{S,F}^L < -2\Delta_{O,F}^L \\ \int_{-1/2}^{\tilde{i}} V_{A,F}(i) dG(i) + \int_{\tilde{i}}^{1/2} V_{N,F}(i) dG(i) - \int_{-1/2}^{1/2} V_{0,F}(i) dG(i) & \text{if } \Delta_{S,F}^L > 2\Delta_{O,F}^L \end{cases}$$

and the result from Proposition 3 translates accordingly so that we can rewrite $V(\mathcal{R})$ as

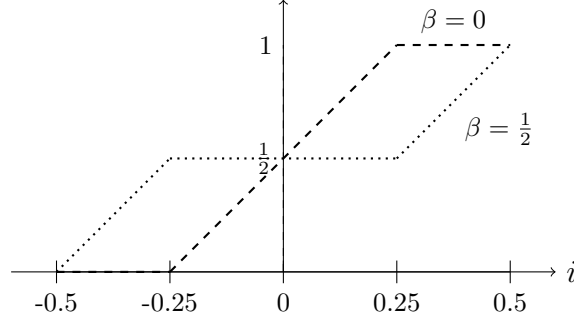
$$\begin{aligned} & \pi^L[\Delta_{O,F}^L - \Delta_{S,F}^L \mathbb{E}_G[i]] && \text{if } |\Delta_{S,F}^L| \leq 2\Delta_{O,F}^L \\ & (1 - \pi^L)G(\tilde{i}) [\Delta_{O,F}^D - \Delta_{S,F}^D \mathbb{E}_G[i \mid i \leq \tilde{i}]] + \pi^L(1 - G(\tilde{i})) [\Delta_{O,F}^L - \Delta_{S,F}^L \mathbb{E}_G[i \mid i \geq \tilde{i}]] && \text{if } \Delta_{S,F}^L < -2\Delta_{O,F}^L \\ & \pi^L G(\tilde{i}) [\Delta_{O,F}^L - \Delta_{S,F}^L \mathbb{E}_G[i \mid i \leq \tilde{i}]] + (1 - \pi^L)(1 - G(\tilde{i})) [\Delta_{O,F}^D - \Delta_{S,F}^D \mathbb{E}_G[i \mid i \geq \tilde{i}]] && \text{if } \Delta_{S,F}^L > 2\Delta_{O,F}^L. \end{aligned}$$

Since F satisfies Assumption 1, all receivers will be obedient so that we are in the case

$|\Delta_{S,F}^L| \leq 2\Delta_{O,F}^L$. Thus, we have

$$\begin{aligned} V(\beta) &= \pi^L[\Delta_{O,F}^L - \Delta_{S,F}^L \mathbb{E}_G[i]] \\ &= q_g + \beta Q - (q_g + 2\beta Q)(q_g + Q) + (q_2 - q_1)(q_g - \beta(1 - 2Q)) \mathbb{E}_G[i] \end{aligned}$$

FIGURE 3. POPULATION POLARIZATION



Notes: We have fixed $R = 3/4$. The dotted and dashed lines both correspond to symmetric type distributions F , where $\beta = 1/2$ for the dotted line and $\beta = 0$ for the dashed line. A move from $\beta = 0$ to $\beta = 1/2$ illustrates a symmetric shift in the type distribution that increases the share of types willing to recommend a controversial product at the fixed standard $R = 3/4$.

and therefore $V'(\beta) = Q - 2Q(q_g + Q) - (q_2 - q_1)(1 - 2Q)\mathbb{E}_G[i]$. Replacing $q_g = ((1 - 2Q)\sigma)/(\sigma + 1)$ and rearranging the inequalities $V'(\beta) > 0$ and $V'(\beta) < 0$ yields the statement in the proposition. ■

A.4.4 Polarization and the Prevalence of Controversial Products

Polarization. Under Assumption 1, polarization affects the rating system through $\beta = F(1/2 - R)$, the share of senders who recommend a controversial product at rating standard R . By Lemma A.2, the value of the rating system is strictly increasing in β when $\sigma < 1$, strictly decreasing in β when $\sigma > 1$, and independent of β when $\sigma = 1$.

Figure 3 illustrates such a change in β . The intuition is that a higher β makes likes more frequent but less objectively informative. To see this, suppose $\sigma < 1$ and $R > 1/2$. As shown in Figure 1, when $R > 1/2$, only sufficiently extreme types issue likes for controversial products. Hence a higher β means that likes are more likely to stem from controversial rather than objectively good products. This raises the probability of likes while lowering their objective value. When good products are rare, the frequency effect dominates; when good products are common, the informativeness effect dominates.

The impact of preference heterogeneity on sales is straightforward. Under Assumption 1, all types are obedient so sales equal the probability of a like and are therefore increasing in β . Thus, shifts that increase β also increase sales.

Prevalence of Controversial Products. The following lemma gives the comparative statics of V in Q :

Lemma A.6 *Suppose Assumption 1 holds.*

(i) *If $3\sigma - \beta - \sigma^2 + \sigma^2\beta > 0$, then the value of the rating system is decreasing in Q . In particular, this is the case when $\sigma = 1$.*

(ii) *If $3\sigma - \beta - \sigma^2 + \sigma^2\beta < 0$, then $Q^* := \frac{3\sigma - \beta - \sigma^2 + \sigma^2\beta}{4\sigma - 4\beta - 4\sigma^2 + 4\sigma^2\beta} \in (0, \frac{1}{2})$ maximizes $V(R)$.*

Proof.

We first take the derivative of V with respect to Q to obtain

$$\frac{\partial V}{\partial Q} = \frac{1}{(\sigma + 1)^2} (\beta - 3\sigma + 4Q\sigma - 4Q\beta + \sigma^2 - 4Q\sigma^2 - \sigma^2\beta + 4Q\sigma^2\beta)$$

Grouping the Q -terms, this can be written in affine form as

$$\frac{\partial V}{\partial Q} = \frac{-(3\sigma - \beta - \sigma^2 + \sigma^2\beta) + 4Q(\sigma - \beta - \sigma^2 + \sigma^2\beta)}{(\sigma + 1)^2},$$

which is linear in Q .

Proof of (i). Suppose $3\sigma - \beta - \sigma^2 + \sigma^2\beta > 0$. Straightforward calculations show that $\partial V/\partial Q$ is negative for $Q = 0$ and negative for $Q = 1/2$. Because it is affine in Q , it is thus negative throughout $(0, 1/2)$. Hence V is decreasing in Q , proving (i). In particular, at $\sigma = 1$, $3\sigma - \beta - \sigma^2 + \sigma^2\beta = 2$ for every $\beta \in [0, 1]$, so (i) applies.

Proof of (ii). Suppose $3\sigma - \beta - \sigma^2 + \sigma^2\beta < 0$. Setting $\partial V/\partial Q$ equal to zero and solving for the candidate solution we obtain

$$Q^* = \frac{3\sigma - \beta - \sigma^2 + \sigma^2\beta}{4\sigma - 4\beta - 4\sigma^2 + 4\sigma^2\beta}$$

Next, the SOC reads $4\sigma - 4\beta - 4\sigma^2 + 4\sigma^2\beta < 0$, where the l.h.s. of the SOC is the denominator of Q^* . Using the identity $4\sigma - 4\beta - 4\sigma^2 + 4\sigma^2\beta = 4(3\sigma - \beta - \sigma^2 + \sigma^2\beta) - 8\sigma$, the condition $3\sigma - \beta - \sigma^2 + \sigma^2\beta < 0$ (together with $\sigma > 0$) implies the SOC holds. Hence,

the candidate solution is positive. It remains to show that $Q^* < \frac{1}{2}$. The requirement $\frac{3\sigma - \beta - \sigma^2 + \sigma^2\beta}{4\sigma - 4\beta - 4\sigma^2 + 4\sigma^2\beta} < 1/2$ becomes

$$2\sigma - 2\beta - 2\sigma^2 + 2\sigma^2\beta < 3\sigma - \beta - \sigma^2 + \sigma^2\beta,$$

because the denominator of the ratio is negative. Equivalently, $(\sigma + 1)(\sigma(1 - \beta) + \beta) > 0$, which always holds for $\beta \leq 1$. ■

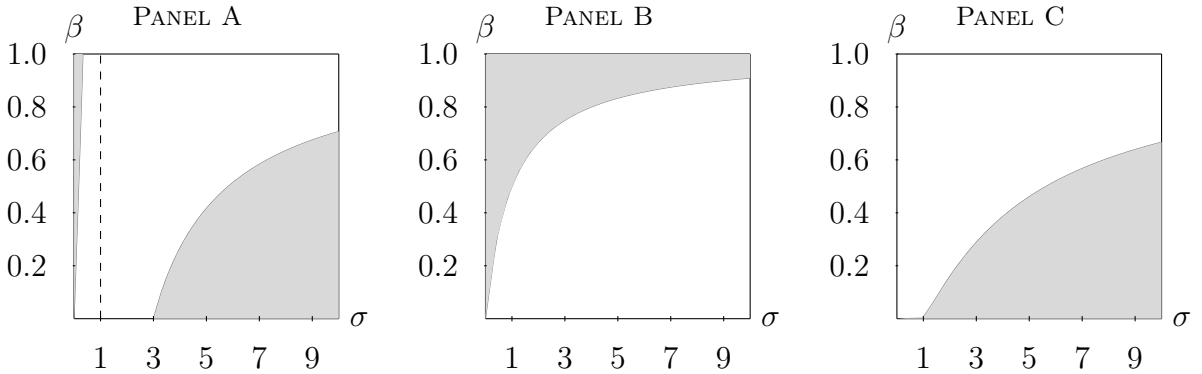
As an immediate consequence we obtain:

Corollary A.3 *Suppose Assumption 1 holds. Then for every $\beta \in (0, 1)$, there exist $\underline{\sigma}(\beta) \in (0, 1)$ and $\bar{\sigma}(\beta) > 1$ such that*

(i) *the value of the rating system is decreasing in Q on $(\underline{\sigma}(\beta), \bar{\sigma}(\beta))$.*

(ii) *For $\sigma < \underline{\sigma}$ and $\sigma > \bar{\sigma}$, the value of the rating system is maximized for some $Q^* > 0$.*

FIGURE 4. ILLUSTRATION OF COROLLARY A.3



Notes: In Panel A, the shaded area indicates the parameter space for which an interior solution for Q arises by Corollary A.3. In Panel B, the shaded area indicates where π^L is increasing in Q . In Panel C, the shaded area indicates where Δ_O^L is increasing in Q , illustrated for $Q = 0.1$.

Figure 4 illustrates the result. In the white area in Panel A, $V(Q)$ is decreasing in Q ; in the shaded area, an interior maximizer exists. When objectively good and bad products are similarly likely (σ is close to 1), a higher prevalence of controversial products indeed reduces $V(Q)$. For very high or low values of σ , however, this need not be true. This surprising result reflects the multiplicative structure of $V(Q)$. Under Assumption 1, $V(Q)$ is the product of the buying probability π^L and the objective value of the rating per buyer,

Δ_O^L . The effect of Q on π^L is negative only if σ is sufficiently large (see Panel B). Intuitively, as Q increases, objectively good products are less likely, reducing the probability of likes. However, there are now more likes stemming from controversial products, as these have become more likely. β and σ determine which of these effects dominates. When β is high, the greater likelihood that controversial products receive likes increases the total probability of likes, strengthening the positive effect. When σ is high, the lower share of objectively good products dominates, strengthening the negative effect. Similarly, one can show that the effect of Q on Δ_O^L becomes negative only for small σ (see Panel C), reflecting the increased probability that a product with a like is controversial. ³⁴

³⁴There is also a positive effect of higher Q , reflecting a change in the value of the outside option. The values of σ and β determine which effect dominates.

References

- ACEMOGLU, D., A. MAKHDOUMI, A. MALEKIAN, AND A. OZDAGLAR (2022): “Learning from reviews: The selection effect and the speed of learning,” *Econometrica*, 90, 2857–2899.
- BANERJEE, A. V. (1992): “A Simple Model of Herd Behavior,” *The Quarterly Journal of Economics*, 107, 797–817.
- BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *Journal of Political Economy*, 100, 992–1026.
- BOHREN, N., R. HAKIMOV, AND L. SANTOS-PINTO (2025): “Beyond the Stars: Exploring the Welfare Effects of Ratings in Differentiated Markets,” Working paper, University of Lausanne; AEA Registry AEARCTR-0011386.
- BOLTON, G., B. GREINER, AND A. OCKENFELS (2013): “Engineering trust: reciprocity in the production of reputation information,” *Management Science*, 59, 265–285.
- BOURREAU, M. AND G. GAUDIN (2022): “Streaming Platform and Strategic Recommendation Bias,” *Journal of Economics and Management Strategy*, 31, 25–47.
- CARNEHL, C., A. STENZEL, AND P. SCHMIDT (2024): “Pricing for the Stars: Dynamic Pricing in the Presence of Rating Systems,” *Management Science*, 70, 1755–1772.
- CHAKRABORTY, A. AND R. HARBAUGH (2010): “Persuasion by Cheap Talk,” *American Economic Review*, 100, 2361–2382.
- CHE, Y.-K. AND J. HÖRNER (2018): “Recommender systems as mechanisms for social learning,” *The Quarterly Journal of Economics*, 133, 871–925.
- CHEN, P.-Y., L. M. HITT, Y. HONG, AND S. WU (2021): “Measuring Product Type and Purchase Uncertainty with Online Product Ratings: A Theoretical Model and Empirical Application,” *Information Systems Research*, 32, 1470–1489.

- HOPENHAYN, H. AND M. SAEEDI (2026): “Optimal Simple Ratings,” *The Journal of Industrial Economics*.
- HOSPITABLE (2024): “Understanding Airbnb Ratings: How They Work and How to Improve Them,” Accessed: March 12, 2025.
- IFRACH, B., C. MAGLARAS, M. SCARSINI, AND A. ZSELEVA (2019): “Bayesian Social Learning from Consumer Reviews,” *Operations Research*, 67, 1209–1221.
- JOHNEN, J. AND R. NG (2024): “Harvesting Ratings,” Working Paper, CORE/LIDAM, UCLouvain and University of Mannheim.
- KAMENICA, E. AND M. GENTZKOW (2011): “Bayesian Persuasion,” *American Economic Review*, 101, 2590–2615.
- LAFKY, J. AND R. NG (2024): “Ratings with Heterogeneous Preferences,” Discussion Paper No. 594, CRC TR 224, University of Bonn and University of Mannheim.
- LI, X. AND L. M. HITT (2008): “Self-Selection and Information Role of Online Product Reviews,” *Information Systems Research*, 19, 456–474.
- LIZZERI, A. (1999): “Information Revelation and Certification Intermediaries,” *The RAND Journal of Economics*, 30, 214–231.
- PEITZ, M. AND A. SOBOLEV (2025): “Inflated Recommendations,” *The RAND Journal of Economics*.
- RAYO, L. AND I. SEGAL (2010): “Optimal Information Disclosure,” *Journal of Political Economy*, 118, 906–948.
- SHI, Z. M. AND T. S. RAGHU (2020): “An Economic Analysis of Product Recommendation in the Presence of Quality and Taste-Match Heterogeneity,” *Information Systems Research*, 31, 399–411.
- SMIRNOV, A. AND E. STARKOV (2025): “Designing Social Learning,” *European Economic Review*, 178.

- SMITH, L. AND P. SØRENSEN (2000): “Pathological Outcomes of Observational Learning,” *Econometrica*, 68, 371–398.
- SUN, M. (2012): “How Does the Variance of Product Ratings Matter?” *Management Science*, 58, 696–707.
- TADELIS, S. (2016): “Reputation and feedback systems in online platform markets,” *Annual Review of Economics*, 8, 321–340.
- UBER (2024): “Understanding Driver Ratings,” Accessed: March 12, 2025.