CONTENT BIAS AND INFORMATION COMPRESSION*

Jinge Liu

November 12, 2024

Abstract

I model content formation for an information intermediary who selectively reports and omits information to fit her content length. I show that even when the sender and the receiver share the same preference and prior belief, the sender may still create two forms of content bias: audience appeal (the literal meaning of content looking more confirmatory than the underlying reality) and sensationalism (the literal meaning of content looking more extreme than the underlying reality, regardless of the latter being confirmatory or contradictory), in order to transmit information efficiently. These biases are transparent to the rational receiver and improve welfare. Taking this model to its asymptotic limit, I show that the content (measured by a sentiment-type variable) is a tractable and smooth function of a fundamental variable that is conditionally Gaussian (a common shock property in learning) and a set of contextual parameters capturing the economic environment. My model addresses the challenge to micro-found sentiment analysis and has numerous implications related to media slant and the analysis of non-content data such as product ratings. *JEL Codes: C40, D82, D83, D90*

*This paper is adapted from a chapter of my Ph.D. dissertation for the Fuqua School of Business, Duke University. I thank my Ph.D. committee co-chairs Simon Gervais and S. "Vish" Viswanathan, as well as committee members Arjada Bardhi and Felipe Varas, for their invaluable advice and support. I also thank seminar and conference participants at the Asia Meetings of the Econometric Society at Shenzhen and Tokyo, City University of Hong Kong, Duke Finance Student Theory Workshop, Hong Kong Joint Finance Research Workshop, Stony Brook International Conference on Game Theory, and Paris Workshop on Games, Decisions, and Language.

I INTRODUCTION

Reports created by information intermediaries often exhibit bias relative to the underlying reality. For instance, in certain circumstances, even factually accurate content produced by trusted news outlets may be skewed. This raises an intriguing question: What is the relationship between reported information, the underlying reality, and the economic context in which reports are produced? To reliably interpret real-world communication content (e.g., in business and politics) and make use of content data in economic research, we need a thorough understanding of this relationship.

In this paper, I micro-found content bias by presenting a model of content formation in which bias arises as a consequence of information selection. The model captures the situation of an information intermediary who holds abundant information, but faces physical constraints determining how much of it she can report. This situation is common, for example, in the news industry, where media outlets may fit information to newspaper size, broadcast time, or webpage size. (As *Seinfeld* put it, "It's amazing that the amount of news that happens in the world every day always just exactly fits the newspaper.") Such an intermediary must report selectively, omitting some of the information she holds.

To model this situation, I consider a sender who must choose a certain number of news pieces (out of a larger collection that she has available, called the *scenario*) to present to a decision-maker. Each news piece is binary, with, e.g., either positive or negative realization. The sender's goal is to maximize the decision-maker's utility. My main findings are as follows:

- In equilibrium, the selected content may exhibit two distinct types of bias, both are well-documented stylized facts: (i) audience appeal and (ii) sensationalism. The degree of each depends on the economic context. These biases are due solely to the sender's selection or omission of information; the sender never lies. The biases are apparent, in the sense that the rational decision-maker is not misled but correctly interprets the biases as a way of communicating more effectively. Rather than hurting welfare, the biases actually help maximize it.

- Asymptotically, as both the total number of news pieces available (the complexity of the scenario) and the sender's reporting capacity (the complexity of the content) become large, the model becomes tractable and smooth. The *fundamental* K, which summarizes the scenario, is asymptotically conditionally Gaussian, a common property of shock specifications in economic models with learning. The proportion of positive news in the content, which resembles the *sentiment* measure in empirical content analysis, has slope in K given by a probability density function proportional to

$$\lambda_F(K)^{\frac{1}{6}}\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}},$$

  where λ_F and λ_h come from higher-order curvatures of the fundamental distribution and the utility, respectively, and $\tilde{a}$ is the decision-maker's hypothetical action assuming he has perfect information. Contextual parameters affect these terms.

The term "bias" refers to the disparity between the selected content and the underlying scenario. My first main finding is that bias occurs even though the sender works for the decision-maker. This is because the sender's job is not just to relay information but to help the decision-maker maximize his utility. Because she does not have enough reporting capacity to provide perfectly precise content, she cannot uniquely label every possible scenario for the decision-maker. She must therefore ask: Which scenarios should be pooled and assigned the same label? What content should be used to label each pool?

The answers to these questions show how bias arises. For the first question, the sender chooses a pooling to minimize utility loss from information compression. This predicts a cutoff structure for the optimal pooling. The cutoff locations depend on each scenario's *newsworthiness*, that is, its relevance to the decision-maker's utility, which is determined by its conditional probability and sensitivity of the best possible payoff on each true state, as well as the elasticities of all these terms. The more newsworthy scenarios will be pooled less aggressively, because the sender will enjoy greater utility improvements if he can distinguish between them.

The answer to the second question is related to the structure of the content (or label) space. I show that the sender can always label the utility-maximizing pools of scenarios with content meeting two natural criteria: *honesty* and *self-consistency*. The way to do so is to assign adjacent labels to adjacent pools, in increasing order. I call this strategy the *content-generating information structure*. If we view it as a mapping from the scenario space to the content space, then the variations in pooling aggressiveness described in the previous paragraph create nonlinearity in the mapping. This nonlinearity manifests as content bias.

Specifically, two types of bias arise: audience appeal and sensationalism. Audience appeal means the content favors the version of reality preferred or believed by the decision-maker; for example, if he benefits when the state of nature is positive or has a prior belief that favors the positive state, then the proportion of positive news is higher in the content than in the underlying scenario. Sensationalism means the content is more extreme (less balanced between positive and negative) than the scenario. In my model, these biases are due to information selection; they reflect the fact that the sender is more willing to devote her limited reporting capacity to distinguishing between more newsworthy scenarios that are contradictory or moderate for maximizing utility. Thus, unlike the broader literature, this paper regards content bias as welfare-maximizing, given the sender's communication constraints.

For a full structural analysis, I pass to an asymptotic version of the model. This brings several advantages. First, the asymptotic model is practically relevant: As previously noted, asymptotically, the fundamental is a conditionally Gaussian signal, and such signals are common as shock specifications in economic models with learning. Similarly, the content measure—the proportion of positive signals in the content—is one often used in empirical studies, especially in sentiment analysis. Second, going asymptotic makes the model more tractable, facilitating interpretation and analysis by opening up the black box of the content measure. In particular, I show that for many common utility functions, the sender's optimal reporting policy takes a simple form. I find that, asymptotically, sensationalism is inevitable and audience appeal is common. When the marginal utility of extreme bets is very high and becomes a dominant force, an alarmist or reverse-appeal bias is also possible.

The extent of bias depends on the economic context, including parameters such as the decision-maker's prior belief, the payoff-relevance of each state, the informativeness of the signals comprising the scenario, and the shape of the utility curve. These parameters determine the relative importance of each scenario and thus affect the sender's reporting policy. In particular, the prior and payoff-relevance affect the degree of audience appeal in the content, while increasing informativeness leads to more sensationalism.

An important feature of my model is the distinction between the literal meaning and the implicit true meaning of the content. That is, concretely, the content consists of a list of facts. However, the rational decision-maker, who is aware of the economic context, does not take these facts at face value. Rather, accounting for the sender's reporting policy, he discerns the biases in the content and infers the underlying meaning. This distinction has a practical implication: Researchers analyzing content data, who stand outside the economic context in which the data were generated, should consider the context in order to avoid specification errors.

The model has useful applications in many settings. As discussed, it gives a novel rationalization for slant in the news media. It also provides micro-foundations for sentiment analysis, a first in the literature. Furthermore, because it conceptualizes content formation as the compression of complex information into simple reports, it has implications beyond textual content. For example, it can be applied to micro-found the analysis of product star ratings or student exam scores.

Literature. To the best of my knowledge, this paper is the first to identify physical communication capacity as a source of content bias and the first to tractably micro-found content data in context for potential empirical application. In addition, it provides a novel theoretical framework for communication with limited capacity.

My work is related to the literature on media bias as a demand-side phenomenon, which attributes bias to the desire to attract an audience. Seminal papers in this area include those of Mullainathan and Shleifer (2005), who consider media competing for a heterogeneous audience, and Gentzkow and Shapiro (2006), who consider the sender's reputation. Both of these papers require some heterogeneity in beliefs or preferences. By contrast, I establish the occurrence of bias without assuming heterogeneity. My paper also connects to work on communication games with limited attention and bias; see, for instance, Che and Mierendorff (2019) and Perego and Yuksel (2022).

In methodology, this paper is related to the literature on communication with limited capacity, which is often termed "limited attention" if the capacity limitations are attributed to the receiver. In common with works on Bayesian persuasion (Kamenica and Gentzkow 2011) with limited attention (e.g. Gentzkow and Kamenica 2014 and Bloedel and Segal 2021), it discusses optimal compression and attention allocation; however, it departs from those works by incorporating a practically motivated capacity limit in place of an information-theoretic one (see Cover and Thomas 2006 and, e.g., Sims 2003).

This paper models labels. Doing so is crucial for empirical relevance, as real-world content data are all labels. This feature is novel in the literature on communication games with commitment (Kamenica and Gentzkow 2011, Bergemann and Morris 2019), which focuses on posteriors and abstracts away from how labels look. The paper is also connected to the literature on partial disclosure of information or hard evidence (e.g., Milgrom 1981, Milgrom and Roberts 1986, and Dye 1985).

The modeling of economic foundations is novel in the extensive and growing empirical literature on content analysis (see the survey of Gentzkow, Kelly, and Taddy 2019). My approach to modeling content particularly speaks to studies that use textual frequency measures to investigate tendencies related to two competing extremes, such as economic boom and bust phases, left-wing and right-wing politics, or stability and instability. Examples include Antweiler and Frank (2004), Tetlock (2007), Tetlock, Saar-Tsechansky, and Macskassy (2008), and Loughran and McDonald (2011), which use frequencies of linguistic tokens, and Gentzkow and Shapiro (2010) and Baker, Bloom, and Davis (2016), which use frequencies of articles or covered events. This paper provides a method of parameterizing a model for content data to extract information from such textual measures, which addresses a longstanding challenge in the study of content data in context.

The rest of the paper proceeds as follows. In Section II I introduce the baseline model and illustrate how biases arise. In Section III I take the baseline model to its asymptotic limit and examine the solution. In Section IV I discuss the model's implications for media bias and content analysis. In Section V I extend the model to analyze consumer ratings. Section VI concludes.

II THE BASELINE MODEL

II.A Agents

A sender (*she*) reports to a decision-maker (*he*) information about the binary state of nature $\theta \in \{0,1\}$. The decision-maker has prior belief $\Pr(\theta = 1) = \pi \in (0,1)$ and places a bet $a \in [0,1]$ on the true state, with payoff

$$u(a;\theta) = u_\theta h(1 - |a-\theta|) = \begin{cases} u_1 h(a) & \text{if } \theta = 1, \\ u_0 h(1-a) & \text{if } \theta = 0, \end{cases} \tag{1}$$

where $u_1, u_0 > 0$ are payoff-relevance parameters for the two states, and $h(\cdot)$ is an auxiliary function defined on $[0,1]$ that captures the closeness between the true state and the bet. I make the following assumption about $h(\cdot)$.

Assumption 1. *(i) The auxiliary function $h(\cdot)$ is twice continuously differentiable, with $h'(a) > 0$ and $h''(a) < 0$ on $(0,1)$.*

(ii) For any possible posterior π', $a^* := \arg\max_a (1-\pi')u_0 h(1-a) + \pi' u_1 h(a) \in (0,1)$.

The assumption $h'(a) > 0$ implies that the decision-maker is better off the closer his bet is to the true state. The assumption $h''(a) < 0$ reflects the economic benefits of diversification: it allows for interior actions $a \in (0,1)$ to be relevant. If $h''(a) \geq 0$, then the only possible optimal actions are $a = 0$ and $a = 1$, and so communication is trivial, since the sender has enough reports at her disposal to perfectly recommend one or the other. Furthermore, part (ii) of Assumption 1 requires that the optimal action is *always* in the interior. This assumption makes it easier to illustrate how information compression works. An example of a sufficient condition guaranteeing this is $h'(1) = 0$.

The sender shares the preferences and prior belief of the decision-maker; that is, she faithfully serves his interests. We can justify this in terms of either pure or strategic loyalty. For the latter, suppose the sender can strategically position her reporting perspective described by $u_s(a;\theta)$ and π_s, and assume her profit from providing information services is increasing in the expected utility increase that her report adds for the decision-maker. Then obviously it is optimal for her to align her perspective with the decision-maker's $u(a;\theta)$ and π and aim to maximize his utility. This sender specification allows us to focus on information compression, without complications related to persuasion. It also makes the sender's commitment power irrelevant.

II.B Timing, Information, and Strategy

The timing is as follows: First, the sender receives N binary *signals* $s_1, ..., s_N \in \{0,1\}$ from nature. She then delivers n ($n \leq N$) binary *reported elements* $r_1, ..., r_n \in \{0,1\}$ to the decision-maker. Finally, the decision-maker chooses an action a.

A signal s_i represents a piece of evidence—a potential news story. I refer to a full profile of signal realizations $\mathrm{s} = (s_1, ..., s_N) \in \{0,1\}^N$ as a *scenario*. I assume $s_1, ..., s_N$ are conditionally independent on θ and $\Pr(s_i = \theta|\theta) = p > 1/2$ for every i.

A reported element r_i represents a piece of news that is covered. For now, I do not require reported elements to be truthful. I refer to the full collection of reported elements $\mathbf{r} = (r_1, ..., r_n) \in \{0,1\}^n$ as the *content*; I assume their order conveys no information. The positive integer n, which I call the *physical constraint*, represents the content length and is exogenously given. Note that such length is a requirement or convention: The sender cannot report more or fewer than n elements. In practice, newspaper space, broadcast time slots, or norms of report length are capacities neither exceeded nor partially filled. They may be endogenously determined ex ante under various considerations, but will be respected once pinned down.

The decision-maker's problem in the subgame determined by $\mathbf{r}$ is to choose the bet $a^*(\mathbf{r})$ that solves the program

$$\max_{a\in[0,1]} E[u(a;\theta)|\mathbf{r}, \{\sigma_{\mathbf{sr}}\}_{\mathbf{s}\in\{0,1\}^N, \mathbf{r}\in\{0,1\}^n}], \tag{2}$$

where $\{\sigma_{\mathbf{sr}}\}_{\mathbf{s}\in\{0,1\}^N, \mathbf{r}\in\{0,1\}^n}$ is the sender's information structure and $\sigma_{\mathbf{sr}} = \Pr(\mathbf{r}|\mathbf{s})$. Anticipating the action $a^*(\mathbf{r})$, the sender chooses an information structure that solves

$$\begin{aligned}
&\max_{\{\sigma_{\mathbf{sr}}\}} U = E[u(a^*(\mathbf{r};\theta))|\{\sigma_{\mathbf{sr}}\}_{\mathbf{s}\in\{0,1\}^N, \mathbf{r}\in\{0,1\}^n}] \\
&\text{ s.t. } \sigma_{\mathbf{sr}} \geq 0, \ \forall \mathbf{s} \in \{0,1\}^N, \ \mathbf{r} \in \{0,1\}^n; \ \sum_{\mathbf{r}} \sigma_{\mathbf{sr}} = 1, \ \forall \mathbf{s} \in \{0,1\}^N; \\
&\qquad \sigma_{\mathbf{sr_1}} = \sigma_{\mathbf{sr_2}}, \text{ for } \mathbf{r_1}'\mathbf{1}_{(n\times 1)} = \mathbf{r_2}'\mathbf{1}_{(n\times 1)}.
\end{aligned} \tag{3}$$

The last condition reflects the fact that the order of the reported elements does not matter.

Dimension reduction. This problem looks high-dimensional but can be simplified. Rather than dealing with the full scenario $\mathbf{s} = (s_1, ..., s_N)$, we can consider the *fundamental* $K := \sum_{i=1}^N s_i$ as its sufficient statistic. Since the order of reported elements is assumed irrelevant, given the content $\mathbf{r} = (r_1, ..., r_n) \in \{0,1\}^n$, we can consider the *report* $k := \sum_{i=1}^n r_i$. Whereas the full scenario space and content space are $\{0,1\}^N$ and $\{0,1\}^n$, the spaces of fundamentals and reports are simply $\{0,1,...,N\}$ and $\{0,1,...,n\}$. I summarize all the assumptions made about the signal distribution in the following Assumption 2(ii) about the fundamental distribution (and (ii) implies (i)).

Assumption 2. *(i) The Bayes factor* $\Lambda(K) := \Pr(K|\theta = 1)/\Pr(K|\theta = 0)$ *is strictly increasing in* K.

(ii) We have $K|\theta \sim Bi(N, p)$ *if* $\theta = 1$, *and* $K|\theta \sim Bi(N, 1-p)$ *if* $\theta = 0$.

Importantly, the fundamental and report can equivalently be represented by any affine transformations of K and k, including K/N and k/n, the proportions of ones in the scenario and the content. Such affine transformations preserve the equidistant nature of the fundamental and report sequences and hence do not affect the analysis of biases.

With this simplification, the sender's information structure can be represented as $\{\sigma_{Kk}\}_{K=0,...,N;\ k=0,...,n}$, where $\sigma_{Kk} := \Pr(k|K)$, and the decision-maker's action can be represented as $a^*(k)$. The sender finds an information structure $\{\sigma^*_{Kk}\}$ that solves

$$\begin{aligned}
&\max_{\{\sigma_{Kk}\}} U = E[u(a^*(k);\theta)|\{\sigma_{Kk}\}_{K=0,...,N;\ k=0,...,n}] \\
&\text{s.t. } \sigma_{Kk} \geq 0,\ \forall K \in \{0,...,N\}, k \in \{0,...,n\};\ \sum_k \sigma_{Kk} = 1,\ \forall K \in \{0,...,N\}.
\end{aligned} \tag{4}$$

This reformulation of the problem is central to our discussion of strategic information compression. Intuitively, the sender would like to separate all $N+1$ possible fundamentals for the decision-maker; however, she has only $n+1$ reports at her disposal, so she must pool some of the fundamentals. I call $n+1$ her *communication capacity* under the physical constraint.

From the perspective of information theory, we can view reports as codewords and the information structure as a codebook. The sender's aim is then to encode the fundamental optimally. The classical information-theoretic approach (which drives the use of mutual information in the literature on attention; see, e.g., Sims 2003), is to focus solely on the lengths of codewords, taking the codewords themselves as meaningless symbols. In contrast, this paper's approach is tailored to practical economic contexts in which reports (or content data) are not merely symbols; they have literal meanings that need to be modeled and are subject to real-world capacity constraints. Note that although reports can actually be arbitrary symbols for maximizing utility in (4), in Section II.D I will give them substance.

II.C Minimizing Compression Loss

To narrow down the search for solutions to the reformulated problem (4), I now characterize the equilibria of the game. The next three propositions, proved in the appendix, give necessary conditions that the sender must meet to minimize information loss when assembling her report.

First, the sender must use pure strategies. Intuitively, since she is loyal, she has no incentive to introduce unnecessary noise by using mixed strategies.

Proposition 1 (pure strategy)**.** *Under Assumptions 1 and 2(i), $\sigma^*_{Kk} \in \{0, 1\}$ for any K, k.*

Therefore, the sender's optimal strategy is to partition the set of fundamentals and map all the fundamentals in each partition set to the same report. Two questions follow: How many reports does she use in equilibrium? Which fundamentals does she pool together?

The next proposition states that she must use all available reports.

Proposition 2 (surjection)**.** *Under Assumptions 1 and 2(i), for each k there exists K such that $\sigma^*_{Kk} > 0$.*

The intuition here is that each additional report allows the sender to distinguish between more fundamentals by further refining the pooling structure; thus, reports should not be wasted. In other words, given a pooling structure (a map of partition sets to reports), if the sender has an extra report at her disposal, then she can split one of the partition sets into two parts, map one part to the original report, and map the second part to the new report; this will strictly increase expected utility. Therefore, an optimal strategy must use all $n + 1$ reports.

The last proposition describes how fundamentals are pooled.

Proposition 3 (cutoff structure)**.** *Under Assumptions 1 and 2(i), let* $\{B_0, ..., B_n\}$ *denote the partition of the fundamental space corresponding to an optimal information structure* $\{\sigma^*_{Kk}\}$*. Then there exist cutoffs* $\{K^*_1, ..., K^*_n\}$ *such that* $B_0 = \{0, ..., K^*_1\}$*,* $B_1 = \{K^*_1 + 1, ..., K^*_2\}$*, ... ,* $B_k = \{K^*_k + 1, ..., K^*_{k+1}\}$*, ... ,* $B_n = \{K^*_n + 1, ..., N\}$*.*

Intuitively, to minimize compression loss, the sender should pool "similar" fundamentals. The appropriate measure of similarity between two fundamentals turns out to be the closeness of their Bayes factors; by Assumption 2(i), this implies that an optimal solution involves an *ordered* partition, with adjacent fundamentals pooled. Thus, the partition is characterized by n cutoffs separating the fundamental space into $n+1$ partition sets. The following example visually illustrates this intuition.

Example 1. Let $u(a;\theta) = u_\theta \cos\left(\frac{\varpi}{2}|a-\theta|\right)$, where ϖ represents the mathematical constant pi. (This corresponds to setting $h(a) = \sin\left(\frac{\varpi}{2}a\right)$.) Under Assumption 2(ii), we have $\Lambda(K) = (p/(1-p))^{2K-N}$ and the sender's objective is

$$U = \sum_{0 \le i \le n} \| \sum_{\mathbf{v}_K \in B_i} \mathbf{v}_K \|,$$

where $\mathbf{v}_K = \left((1-\pi)u_0 C^K_N p^{N-K}(1-p)^K, \pi u_1 C^K_N p^K (1-p)^{N-K}\right) \in \mathbf{R}^2$, $\|\cdot\|$ is the Euclidean norm, and B_i is an element of the partition $\{B_0, ..., B_n\}$.

Geometrically, each fundamental K can be represented by a vector $\mathbf{v}_K$ (shown as a blue arrow in Fig. I). The sender calculates her utility as follows: First, she partitions the collection of all fundamental vectors into $n+1$ partition sets. Then, for each partition set, she calculates the vector sum of all of the fundamentals in that set. We call the resulting vector the *representative vector*; its Euclidean length equals the contribution of the fundamentals in that partition set to the sender's expected utility. Finally, she calculates the sum of the lengths of all of the representative vectors; this equals the expected utility.

[Insert Figure I here.]

To maximize the expected utility, the sender should choose a partition that minimizes the loss of length caused by summing the vectors in each partition set; thus, she should pool vectors with similar angles. The angle associated with $\mathbf{v_K}$ is $\tilde{a}(K) = \arctan \frac{\pi}{1-\pi} \frac{u_1}{u_0} \Lambda(K)$, which is strictly monotone in $\Lambda(K)$ and hence in K. Therefore, the sender should pool adjacent K. Figure I depicts the optimal partition in an example with $N = 5$.

Note that to maximize expected utility in the reformulated problem (4), it suffices to find an optimal partition; it does not matter which report is attached to each partition set, as long as each set gets a distinct report. This means multiple equilibria always exist, as the sender can permute the reports arbitrarily without affecting utility. The assignment of reports is simply an act of labeling. In the next section, we refine our perspective by associating reports with meanings.

II.D Connecting Symbols with Substance

In the real world, content data are not arbitrary symbols; readers of a newspaper, for example, will not accept its action recommendations unless they appear to be fact-based. Consequently, information intermediaries still insist on presenting evidence in practice when there are other forms of space-saving communication available, such as merely sending a short unsubstantiated summary or directly saying the action recommendation. They must frame their reports as a presentation of reality, able to withstand common-sense scrutiny from their audience.

To give symbols substance, I now introduce two criteria for the content in my model that reflect common audience expectations of trustworthiness and reliability. The first criterion is *(verifiable) honesty*, or the ability to survive fact-checks. This criterion says that reported elements must match actual signal realizations.

Definition 1 (the honesty criterion)**.** *An information structure is* honest *if, for each K, any k such that $\sigma_{Kk} > 0$ satisfies $K - (N - n) \leq k \leq K$.*

Importantly, this criterion does not require balanced coverage; it only says that no constituent piece of the content is fabricated. Suppose $N = 100$ and $n = 50$, and the realized scenario consists of 50 ones and 50 zeros. Then content consisting of 50 ones and no zeros would still be honest, though plainly biased. Hence the honesty criterion is not difficult to meet.

The second criterion is *(logical) self-consistency*, which says the sender cannot contradict herself. That means the report should be increasing as a function of the fundamental.

Definition 2 (the self-consistency criterion)**.** *An information structure is* self-consistent *if for any K_1, K_2 such that $K_1 < K_2$, the conditions $\sigma_{K_1k_1} > 0$ and $\sigma_{K_2k_2} > 0$ imply $k_1 \leq k_2$.*

Under self-consistency, if one fundamental favors $\theta = 1$ over $\theta = 0$ more than another fundamental, then the report on the former fundamental should exhibit more favor for $\theta = 1$ as well. To interpret this, note that there are two meanings associated with a report: a literal meaning, which comes from the fact that the content looks like a collection of signals, and a true meaning regarding the fundamental, which is implied by the information structure. Self-consistency requires the two meanings to move in the same direction. For instance, suppose $\theta = 1$ and $\theta = 0$ represent good and bad states of the economy. Self-consistency means that whenever the sender receives better news about the economy, her report looks more optimistic. Information intermediaries that are not self-consistent may appear untrustworthy.

Theorem 1, the first main result, says we can refine the equilibria for a desirable label structure.

Theorem 1. *Under Assumptions 1 and 2(i), there exists a solution to Eq. (4) that satisfies both honesty and self-consistency. Furthermore, any solution that satisfies self-consistency also satisfies honesty.*

The proof is straightforward: For any equilibrium, let the optimal partition sets $B_0, ..., B_n$ (given in increasing order, as in Proposition 3) map to the reports $0, ..., n$, in that order. This information structure is the only self-consistent one under the optimal partition, and it is obviously honest.

I call this self-consistent optimal information structure the *content-generating information structure*. It must exist[1], but I have not yet pinned it down by specifying where the optimal cut-offs are located. The analytical characterization is left for Section III, where I take the discrete baseline model to an asymptotic limit. For now, we can solve the discrete problem by exhaustively computing the utilities for all of the ordered partitions.

II.E Two Types of Bias, Illustrated

In the baseline model, *bias* refers to the difference between k/n and K/N. The discreteness of the model may contribute to this difference, but this contribution is unimportant and will disappear in the asymptotic model of Section III. Example 2 illustrates how two basic types of bias may arise under the content-generating information structure.

Example 2. A newspaper reports to an investor on tomorrow's market state θ, which is either boom (1) or bust (0). The investor has access to two assets, one paying off u_1 upon boom and zero upon bust, and the other zero upon boom and u_0 upon bust. The investor chooses a portfolio proxied by $a \in [0, 1]$, which represents his position in the former asset when the short-selling constraint is normalized to 0 and the budget to 1. The market belief for booms is π. The newspaper editor has $N = 5$ news stories about the economy, but can publish only $n = 3$. The content-generating information structure depends on the contextual economic variables, including π, u_1, u_0, p, and the shape of the investor's utility curve. Here I focus on π, u_1, and u_0.

Figure II gives an example in which πu_1 greatly exceeds $(1 - \pi)u_0$; that is, the investor prefers or expects the boom state. In this case, the bias of audience appeal emerges. The content-generating information structure, shown in the left panel of Fig. II, is given by the partition sets

1. For its existence, see Appendix A. It is also generally speaking unique, in the sense that asymptotically it is unique by Theorem 2.

$\{0\}$, $\{1\}$, $\{2\}$, and $\{3, 4, 5\}$, which are mapped to the reports 0, 1, 2, and 3, in that order. The right panel of Fig. II shows the *report curve*, a plot of k/n against K/N (given here as a scatter plot). Interpreting K/N and k/n as the levels of optimism in the scenario and the content, respectively, we see from the upward tilt of the report curve that the newspaper disproportionately omits negative stories, appearing to cater to the investor's belief or preference. For instance, when the scenario is 20% optimistic, with one positive story and four negative stories, the editor omits two of the latter, making the content 33% optimistic. When the scenario is 60% optimistic, with three positive and two negative stories, the editor omits both of the latter and publishes 100% optimistic content.

[Insert Figure II here.]

[Insert Figure III here.]

Figure III gives an example in which πu_1 and $(1-\pi)u_0$ are comparable. Here the bias of sensationalism emerges. The content-generating information structure has partition sets $\{0, 1\}$, $\{2\}$, $\{3\}$ and $\{4, 5\}$ mapped to the reports 0, 1, 2, and 3, respectively; that is, the newspaper is either exaggeratedly optimistic or exaggeratedly pessimistic, depending on the direction of the fundamental. For instance, when the scenario is 80% (20%) optimistic, the content is 100% (0%) optimistic. When the scenario is 60% (40%) optimistic, the content is 67% (33%) optimistic.

As we see in these examples, the first step in analyzing the bias is to interpret the partition. The sender chooses the optimal partition by assessing how much the potential knowledge of each scenario contributes to the decision-maker's expected utility; this tells her how aggressively to pool across the fundamental space. The assessment process is rigorously described by Eqs. (7) and (11).

In the example of Fig. II, where πu_1 is high, the investor strongly tends to bet close to 1. Therefore he does not highly value information confirming $\theta = 1$, because such information will not greatly alter his bet. Rather, he values precise information about scenarios far from $\theta = 1$, which will make him rethink his bet. Hence the newspaper editor pools optimistic fundamentals more aggressively than pessimistic ones.

In the example of Fig. III, where there are no strong audience-appeal effects, the investor is not very interested in precise information about tail scenarios. There are two reasons for this. First, the probability of a tail scenario is slim; thus, tail scenarios contribute little to expected utility. Second, whenever there is a tail scenario, the investor will make a similarly extreme bet regardless of how precisely the newspaper's policy distinguishes tail scenarios. Therefore the editor pools tail scenarios more aggressively than moderate ones.

The second step in analyzing the bias is to examine the mapping of fundamentals to reports. Two observations are helpful here. First, the highest and lowest fundamentals are respectively mapped to the highest and lowest reports, without bias. These unbiased reports on end scenarios provide anchors for the analysis of the middle scenarios. Second, the assignment of reports is more sensitive to changes in the fundamental for more newsworthy fundamentals (e.g., the pessimistic ones in our first example, or the moderate ones in our second example) than for less newsworthy ones. In other words, the same incremental increase in K/N will cause a larger increase in k/n if K/N is important than if it is unimportant. This variation in sensitivity stems from the variations in pooling intensity across the fundamental space.

The explanation of bias is now straightforward. Given an interval of highly newsworthy fundamentals K/N, the associated range of reports k/n is large (i.e., the report curve is steeper—more sensitive—there). These newsworthy fundamentals are contradictory to the decision-maker's preference or belief (as in Fig. II), so the report range may be so large that it includes less contradictory, or even confirmatory, reports. Across a fixed range of report values between 0 and 1, less newsworthy confirmatory fundamentals are associated with even more confirmatory reports. This gives rise to audience appeal. The newsworthy fundamentals are also moderate (as in Fig. III), so the report range may be large enough to encompass more extreme reports. Across a fixed report range between 0 and 1, less newsworthy near-extreme fundamentals are associated with more extreme reports. This gives rise to sensationalism.

Three remarks are in order. (1) The report curve, which depicts the relationship between fundamentals and reports, is the quantitative representation of the equilibrium. (2) Interestingly, the report curve also gives the cumulative distribution of cutoffs in the optimal partition. This is because the report k/n associated with a fundamental K/N can be interpreted as the percentage of cutoffs occurring below K/N. Intuitively, wherever the report curve has a high slope, the corresponding fundamentals are more precisely distinguished, which means there are more cutoffs appearing in that region. This observation is useful in the asymptotic model of Section III. (3) In addition to the usual form of audience appeal, an alarmist or reverse-appeal bias is also possible. It occurs when the marginal utility of extreme actions is extremely high, a situation discussed further in Section III.

II.F Welfare

Contrary to the conventional wisdom that biases hurt welfare, my model shows that biases may actually reflect efforts to maximize welfare. The belief that biases hurt welfare is often based on the presumption that they result from agency problems. This paper points out, however, that they may arise even in the absence of agency problems, simply to increase communication efficiency under a capacity constraint. In my model, biases promote the welfare of both the decision-maker and society, since, in both cases, welfare is measured by the sender's ex-ante maximized utility.

My results also imply that bias-free communication policies are suboptimal both for the decision-maker and for social efficiency. Such policies notably include two that are widely accepted as ethical approaches: (1) The sender produces a report that resembles the fundamental as closely as possible; (2) the sender fully randomizes her reporting without any deliberate selection. In the latter case, her ex-ante expected utility is the same as that of choosing n reported elements out of n signals, since the sender can simply report the first n of the N signals.

The parameters N and n both affect welfare. The maximized utility is strictly increasing in N, because richer fundamental information enables the sender to recommend better actions within the same communication constraints. The maximized utility is also strictly increasing in n, as a corollary of Proposition 2: Given $n_1 < n_2$, the optimal information structure under n_1 is suboptimal under n_2, because it fails to make use of an available report. Intuitively, welfare loss comes solely from compression, and a bigger n implies a smaller loss.

III THE ASYMPTOTIC MODEL

In this section I extend the baseline model to an asymptotic model by letting N and n go to infinity. This extension has several benefits: First, a large N reflects the abundance of information in the real world, while a large n reflects the complexity of real-world content data. Second, the asymptotic model is analytically tractable, with equilibria characterized by smooth functions, which simplifies the interpretation and empirical analysis.

III.A Model Setup and Solution

The agents are the same as in the baseline model, with utility depending on $h(\cdot)$. I impose the following assumption, which is analogous to Assumption 1 in the baseline model.

Assumption 3. *The function $h(\cdot)$ is six times continuously differentiable. We have $h'(a) > 0$ and $h''(a) < 0$ on $(0, 1)$.*

Note that Assumption 3 requires more smoothness than Assumption 1, since the asymptotic analysis will involve higher-order derivatives. Also, there is no requirement analogous to Assumption 1(ii). The latter was included in the baseline model simply for convenience, to rule out the possibility that multiple fundamentals trigger the same action of 1 (or 0). In such cases, the sender can pool these fundamentals without any information loss, essentially reducing N to a smaller value N'. If $N' > n$, then Theorem 1 still applies (since the pool has the most extreme Bayes factor). If $N' \leq n$, no information compression is necessary. This possibility is cumbersome to include in the baseline model, but straightforward in the asymptotic model, so I do not need to rule it out here.

Now, let the binary signals s_i take values in $\{-\sigma/\sqrt{N}, \sigma/\sqrt{N}\}$ (where $\sigma > 0$ is a parameter) instead of $\{0, 1\}$. For a fixed N, this is simply an affine transformation of the signal space; the resulting information environment is equivalent to the original one. The fundamental $K = \sum_{i=1}^{N} s_i$ continues to serve as a sufficient statistic for the underlying scenario. Note that K is no longer necessarily a nonnegative integer; it is is a real number and may be negative.

Let $\mu > 0$ be another parameter, and for each i, assume

$$\Pr\left(s_i = \frac{\sigma}{\sqrt{N}}\middle|\,\theta\right) = \frac{1}{2}\left(1 + \frac{\mu_\theta}{\sigma\sqrt{N}}\right),$$

where $\mu_1 = \mu$ and $\mu_0 = -\mu$. This definition relates the probability p from the baseline model to N. Note that the ratio μ/σ measures how informative each signal is about the state θ. For fixed N and n, this setup is equivalent to the baseline model, so Propositions 1, 2 and 3 and Theorem 1 apply, and the content-generating information structure is given by n cutoffs in the fundamental space.

Out of the possible paths by which N and n may go to infinity, I choose the following two-step process. First, I fix n and let N go to infinity (i.e., the content remains small while the scenario becomes increasingly complex). For every finite N, the content-generating information structure is given by n cutoffs in the fundamental space; thus, in the limit as N goes to infinity, the information structure is given by n cutoffs on the real line (denoted by $\mathbf{R}$). Second, I let n go to infinity (i.e., the content becomes more sophisticated). Then, the (limiting) content-generating information structure is characterized by an infinite collection of cutoffs, normalized to a unit measure, that are continuously distributed on $\mathbf{R}$, forming a cutoff density.

For the first step, with n fixed and $N \to \infty$, the analysis is as follows. By standard arguments for infill asymptotics, the fundamental under N, denoted by $K^{(N)}$, satisfies

$$K^{(N)}|\theta \Rightarrow N(\mu_\theta, \sigma^2),$$

where $\Rightarrow$ stands for convergence in law. The limiting fundamental K follows a Gaussian mixture distribution and is supported on $\mathbf{R}$. The asymptotic analogue to Assumption 2(ii) is thus Assumption 4(ii) below. (As before, (ii) implies (i).)

Assumption 4. *(i) The conditional distributions of the fundamental, $F_{K|\theta=1}$ and $F_{K|\theta=0}$, have the following properties:*

(a) They are absolutely continuous and six times continuously differentiable.

(b) The conditional density satisfies $f_{K|\theta=1}(x) > 0$ if and only if $x \in (\underline{K}^{(1)}, \bar{K}^{(1)})$, and $f_{K|\theta=0}(x) > 0$ if and only if $x \in (\underline{K}^{(0)}, \bar{K}^{(0)})$, with $-\infty \leq \underline{K}^{(0)} \leq \underline{K}^{(1)} < \bar{K}^{(0)} \leq \bar{K}^{(1)} \leq +\infty$.

(c) The likelihood ratio $f_{K|\theta=1}(x)/f_{K|\theta=0}(x)$ is strictly increasing on $(\underline{K}^{(1)}, \bar{K}^{(0)})$.

(d) The range of $(1-\pi)u_0h'(1-a)/\pi u_1 h'(a)$ for $a \in (0,1)$ is a subset of the range of $f_{K|\theta=1}(x)/f_{K|\theta=0}(x)$ for $x \in (\underline{K}^{(1)}, \bar{K}^{(0)})$.

(ii) We have $K|\theta \sim N(\mu_\theta, \sigma^2)$.

As in the baseline model, content formation and bias are analyzed through the *report curve*, which is the same as the cumulative distribution of cutoffs like before. The report curve is defined as follows. Let $\{B_0, B_1, B_2, ..., B_n\}$ be the ordered partition used in the content-generating information structure, and let $\kappa^*(n) = \{K_1^*, ..., K_n^*\}$ be the corresponding set of cutoffs, so that the partition is $\{(-\infty, K_1^*), [K_1^*, K_2^*), [K_2^*, K_3^*), ..., (K_n^*, +\infty)\}$. Define the report curve as

$$\beta_n(K) = \frac{1}{n} \sum_{K' \in \kappa^*(n)} \mathbf{1}_{K' \leq K}. \tag{5}$$

This curve fully characterizes the equilibrium, with $\kappa^*(n)$ maximizing the utility

$$\sum_{i=0}^{n} \left\{ \pi u_1 \Pr\left(K \in B_i | \theta = 1\right) h(a_i^*) - (1-\pi) u_0 \Pr\left(K \in B_i | \theta = 0\right) h(1 - a_i^*) \right\}. \tag{6}$$

For each i, a_i^* in Eq. (6) is the recommended action for B_i and hence is subject to

$$\frac{\pi u_1 \Pr\left(K \in B_i | \theta = 1\right)}{(1-\pi) u_0 \Pr\left(K \in B_i | \theta = 0\right)} = \frac{h'(1-a_i^*)}{h'(a_i^*)}.$$

The sender's problem is to choose cutoffs for the Gaussian-mixture limiting fundamental. Here she avoids the complexities arising from the discreteness of the fundamental space in the baseline model. Note that the equilibrium in the limit as $N \to \infty$ can be viewed as satisfying Propositions 1, 2 and 3, as well as Theorem 1. As in the baseline model, the content-generating information structure is honest and self-consistent.

For the second step, with $n \to \infty$, the sender seeks to find the curve $\beta_\infty(K) := \lim_{n\to\infty} \beta_n(K)$, which is both the report curve and the cumulative distribution of cutoffs in the limit. The limiting equilibrium can again be viewed as satisfying Propositions 1, 2 and 3, as well as Theorem 1, and again it is honest and self-consistent.

Obviously, $\beta_\infty(K)$ meets the criteria for being a cumulative distribution function: It is non-decreasing, right-continuous, and defined on $\mathbf{R}$, with 0 and 1 as its limits at $-\infty$ and ∞. Hence it induces a canonical probability space $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{P}')$, where, for any $S \in \mathcal{B}(\mathbf{R})$, $\mathbf{P}'(S)$ is the Lebesgue measure of $\beta_\infty(S)$. This probabilistic perspective on the report curve has two implications.

First, in the asymptotic model, the sender's communication capacity is captured by a unit measure. In the baseline model, the communication capacity is given by the number of partition sets $n+1$, or, equivalently, the n cutoffs dividing them. As $n \to \infty$, this collection of cutoffs determines a unit measure, which serves as the capacity measure in the limit.

Second, instead of solving for the optimal $\beta_\infty(K)$, the sender may solve for its derivative $\beta'_\infty(K)$, which equivalently characterizes the equilibrium. When $\beta_\infty(K)$ is absolutely continuous, $\beta'_\infty(K)$ can be viewed as the density of cutoffs. I call $\beta'_\infty(K)$ the *newsworthiness curve*, because it specifies the importance of each fundamental: If $\beta'_\infty(K)$ is large at a given fundamental K, that means the sender inserts more cutoffs around that K, which reflects her assessment that K

deserves more elaborate coverage. This observation echoes and makes rigorous the discussion in Section II.E on the value of each scenario.

By analogy to the attention allocation curve which appears in the literature on limited attention, I also refer to $\beta'_\infty(K)$ as the *capacity allocation curve*: It describes how the sender allocates the scarce resource of cutoffs across the fundamental space. In fact, although I view the communication capacity as a constraint binding the sender, it can also be interpreted as a constraint binding the decision-maker's information-receiving capacity. In that case, $\beta'_\infty(K)$ can also be named the attention allocation curve. This novel perspective on capacity or attention is based on an economic motivation and complements the information-theoretic perspective in the literature.

What is the equilibrium $\beta'_\infty(K)$? To answer this question, I introduce the *perfect-information optimal action* $\tilde{a}(K)$, which is the decision-maker's hypothetical best action assuming he knows K. Let

$$R(t) = \frac{\pi u_1 f_{K|\theta=1}(t)}{(1-\pi)u_0 f_{K|\theta=0}(t)}.$$

In the case $h'(1)/h'(0) < R(K) < h'(0)/h'(1)$, $\tilde{a}(K) \in (0,1)$ is the solution to $R(K) = h'(1-\tilde{a})/h'(\tilde{a})$. Otherwise, if $R(K) \le h'(1)/h'(0)$, then $\tilde{a}(K) = 0$; if $R(K) \ge h'(0)/h'(1)$, then $\tilde{a}(K) = 1$. Under Assumption 4(i)(d), the range of $\tilde{a}(K)$ includes the interval $(0,1)$. If, in a given economic context, some fundamentals induce the extreme action of 1 (or 0), then these fundamentals can be pooled without loss and mapped to the extreme report of 1 (or 0). Thus it is only necessary to pin down $\beta'_\infty(K)$ for fundamentals K such that $\tilde{a}(K) \in (0,1)$. I denote the interval of such fundamentals by $(\underline{K}, \bar{K})$.

The equilibrium is characterized in Theorem 2, this paper's second main result.

Theorem 2 (asymptotic capacity allocation)**.** *(i) Suppose Assumptions 3 and 4(i) hold. Then on* $(\underline{K}, \bar{K})$*,*

$$\beta'_\infty(K) \propto \lambda_h(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} \tilde{a}'(K)^{\frac{1}{2}} \tag{7}$$

(provided the right-hand side is integrable), where

$$\lambda_h(t) = -\left(h'(\tilde{a}(t))h''(1-\tilde{a}(t)) + h'(1-\tilde{a}(t))h''(\tilde{a}(t))\right),$$

$$\lambda_F(t) = F''_{K|\theta=1}(t)F'_{K|\theta=0}(t) - F'_{K|\theta=1}(t)F''_{K|\theta=0}(t).$$

(ii) Suppose Assumptions 3 and 4(ii) hold. Then

$$\lambda_F(t) \propto \exp\left(-\frac{t^2}{\sigma^2}\right).$$

Theorem 2 decomposes the newsworthiness curve into three components, which are powers of $\lambda_h(K)$ (the curvature of the utility function), $\lambda_F(K)$ (the curvature of the fundamental distribution), and $\tilde{a}'(K)$ (the sensitivity of the perfect-information optimal action). In logarithmic form, Theorem 2 says that the log cutoff density is linear in the logs of $\lambda_h(K)$, $\lambda_F(K)$, and $\tilde{a}'(K)$, with weights of $1/6$, $1/6$, and $1/2$, respectively. The report curve is an antiderivative of $\beta'_\infty(K)$, scaled to be a cumulative distribution function. The proof of Theorem 2 is in Appendix B.

Another way to present Eq. (7) is to define $H_1(K) := h(\tilde{a}(K))$, $H_0(K) := h(1-\tilde{a}(K))$, and $\lambda_H(K) := H'_1(K)H''_0(K) + H''_1(K)H'_0(K)$. Then Eq. (7) says

$$\beta'_\infty(K) \propto \lambda_H(K)^{\frac{1}{6}}\lambda_F(K)^{\frac{1}{6}}.$$

Theorem 2 makes it easy to calculate $\beta'_\infty(K)$ for many common utility functions. Table I lists several examples. In some cases the report curve takes a particularly nice form; for instance, with an exponential utility function, the report curve is the cumulative distribution function for a truncated $N(0, 3\sigma^2)$ distribution with asymmetric tail cutoffs. For a cosine-difference utility function, the factor $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is proportional to a logistic density; for a quadratic utility it is proportional to a hyperbolic secant density, and for log and power utilities it is proportional to certain powers of a hyperbolic secant density. Figure IV shows the capacity allocation and report curves for an example with a cosine-difference utility.

[Insert Table I here.]

[Insert Figure IV here.]

Importantly, Theorem 2 is not about information revelation, even though it implies that the limiting equilibrium reporting policy entails no compression loss. In fact, now that the fundamental and the report take values from two continuums, there exist many one-to-one maps from the fundamental to the report, and any of them gives a loss-free reporting policy. The real point of Theorem 2 is that it specifies an *economically meaningful* one-to-one map—namely, the map that arises as the limiting solution to the baseline problem in Eq. (4), or to the asymptotic problem in Eq. (6). The analytical expression in Theorem 2 gives a tractable asymptotic approximation of the equilibrium information structure solving either Eq. (4) or Eq. (6).

As a final observation, let β_∞^{-1} be the inverse of β_∞ on $(\underline{K}, \bar{K})$, and let Φ be the standard Gaussian cumulative distribution function. Under Assumption 4(ii), the decision-maker's posterior belief on seeing a report $\rho \in (0,1)$ is determined by $K = \beta_\infty^{-1}(\rho)$, i.e.,

$$\Pr(\theta = 1|\rho) = \frac{\pi \exp\left(\frac{2\mu}{\sigma^2}\beta_\infty^{-1}(\rho)\right)}{\pi \exp\left(\frac{2\mu}{\sigma^2}\beta_\infty^{-1}(\rho)\right) + 1 - \pi}.$$

For $\rho = 0$ and $\rho = 1$, the posteriors are respectively

$$\frac{\pi\Phi(\frac{\underline{K}-\mu}{\sigma})}{\pi\Phi(\frac{\underline{K}-\mu}{\sigma}) + (1-\pi)\Phi(\frac{\underline{K}+\mu}{\sigma})} \quad \text{and} \quad \frac{\pi(1-\Phi(\frac{\bar{K}-\mu}{\sigma}))}{\pi(1-\Phi(\frac{\bar{K}-\mu}{\sigma})) + (1-\pi)(1-\Phi(\frac{\bar{K}+\mu}{\sigma}))}.$$

III.B Factors Determining Capacity Allocation

The three components in Eq. (7), which capture the higher-order curvatures in the problem, fall into two groups. The first group consists of $\lambda_F(K)^{\frac{1}{6}}$, which depends solely on the conditional distributions of the fundamental; it describes how the likelihood of a scenario directly affects its newsworthiness. It comes from the optimality condition for a_i^* under Assumption 3 and Assumption 4:

$$\frac{F_{K|\theta=1}(K_{i+1}^*) - F_{K|\theta=1}(K_i^*)}{F_{K|\theta=0}(K_{i+1}^*) - F_{K|\theta=0}(K_i^*)} = \frac{(1-\pi)u_0 h'(1-a_i^*)}{\pi u_1 h'(a_i^*)} \equiv \frac{f_{K|\theta=1}(\tilde{K}(a_i^*))}{f_{K|\theta=0}(\tilde{K}(a_i^*))}. \tag{8}$$

Here, $\tilde{K}(a)$ is the inverse of $\tilde{a}(K)$ on $(\underline{K}, \bar{K})$ and stands for the perfect-information fundamental equivalence for a fundamental partition set inducing action a. The second equality in Eq. (8) follows from the definition of $\tilde{a}(K)$. This condition implies that the relative location of $\tilde{K}(a_i^*)$ within $[K_i^*, K_{i+1}^*]$ is[2]

$$\frac{\tilde{K}(a_i^*) - K_i^*}{K_{i+1}^* - K_i^*} \approx \frac{1}{2} + \frac{1}{24} \left(\frac{d \ln \lambda_F(K)}{dK} \bigg|_{K \in [K_i^*, K_{i+1}^*]} \right) (K_{i+1}^* - K_i^*). \tag{9}$$

The second group consists of $\lambda_H(K)^{\frac{1}{6}} = \lambda_h(K)^{\frac{1}{6}} \tilde{a}'(K)^{\frac{1}{2}}$. Intuitively, $\tilde{a}'(K)$ describes how sensitive the action is to the fundamental. The term $\lambda_h(K)$, which depends on K only through $\tilde{a}(K)$, describes how the sensitivity of the action translates to sensitivity of the utility. By the optimality condition for K_i^*,

$$\frac{h(a_i^*) - h(a_{i-1}^*)}{h(1 - a_i^*) - h(1 - a_{i-1}^*)} = \frac{(1-\pi) u_0 f_{K|\theta=0}(\tilde{a}(K_i^*))}{\pi u_1 f_{K|\theta=1}(\tilde{a}(K_i^*))} \equiv \frac{h'(\tilde{a}(K_i^*))}{h'(1 - \tilde{a}(K_i^*))},$$

the position of $\tilde{a}(K_i^*)$ within $[a_{i-1}^*, a_i^*]$, or equivalently that of K_i^* within $[\tilde{K}(a_{i-1}^*), \tilde{K}(a_i^*)]$, is

$$\frac{K_i^* - \tilde{K}(a_{i-1}^*)}{\tilde{K}(a_i^*) - \tilde{K}(a_{i-1}^*)} \approx \frac{1}{2} + \frac{1}{24} \left(\frac{d \ln \lambda_H(K)}{dK} \bigg|_{K \in [\tilde{K}(a_{i-1}^*), \tilde{K}(a_i^*)]} \right) (\tilde{K}(a_i^*) - \tilde{K}(a_{i-1}^*)). \tag{10}$$

Clearly, to study how a unit measure's worth of cutoffs are distributed across the real line, it suffices to study their relative allocation between any two neighborhoods. Locally, for a cutoff K_i^* to lie near a high concentration of other cutoffs, it should be closer to $\tilde{K}(a_i^*)$ than the next cutoff K_{i+1}^* is, while $\tilde{K}(a_i^*)$ in turn may need to be closer to K_{i+1}^* than the more distant $\tilde{K}(a_{i+1}^*)$ is. Hence, both $\frac{1}{24} \frac{d \ln \lambda_F(K)}{dK}$ and $\frac{1}{24} (\frac{d \ln \lambda_H(K)}{dK})$ matter in characterizing local relative cutoff concentration between nearby neighborhoods. Then, for two neighborhoods K_1 and K_2 located apart from each other, their relative cutoff allocation can be calculated by properly aggregating these characterizations of local concentrations for all the in-between neighborhoods, an exercise that leads to Theorem 2.

2. Eq. (9) and Eq. (10) are respectively variants of Eq. (B.3) and Eq. (B.4).

What is the intuitive economic interpretation of $\lambda_F(K)$ and $\lambda_H(K)$? Notice that

$$\lambda_F(K) = f_{K|\theta=0}(K) f_{K|\theta=1}(K) \left(\ln \frac{f_{K|\theta=1}(K)}{f_{K|\theta=0}(K)} \right)', \; \lambda_H(K) = H_0'(K) H_1'(K) \left(\ln \frac{H_1'(K)}{H_0'(K)} \right)'.$$

That is, $\lambda_F(K)$ is increasing in the two conditional likelihoods and the elasticity of the likelihood ratio, while $\lambda_H(K)$ is increasing in the two conditional value sensitivities and the elasticity of the value sensitivity ratio. Each of the latter can be further decomposed into a utility sensitivity term and an action sensitivity term, $\tilde{a}'(K)$, via the chain rule. These terms fully describe newsworthiness, with Eq. (7) becoming

$$\beta_\infty'(K) \propto f_{K|\theta=0}(K)^{\frac{1}{6}} f_{K|\theta=1}(K)^{\frac{1}{6}} \left(\ln \frac{f_{K|\theta=1}(K)}{f_{K|\theta=0}(K)} \right)'^{\frac{1}{6}} H_0'(K)^{\frac{1}{6}} H_1'(K)^{\frac{1}{6}} \left(\ln \frac{H_1'(K)}{H_0'(K)} \right)'^{\frac{1}{6}}. \quad (11)$$

III.C Characterizing Bias in the Report Curve

Audience appeal. To identify the audience-appeal bias in the reporting policy, I compare the location of $\beta_\infty'(K)$ with the "midpoint" of the fundamental space. It is appropriate to take this midpoint to be zero, since it is zero for any finite N. Relative to zero, the more $\beta_\infty'(K)$ leans against the direction given by the sign of $\frac{\pi u_1}{(1-\pi)u_0} - 1$, the more the reporting policy should be viewed as biased toward that direction. Let $K_{1/2}$ denote the quantity $-\frac{\sigma^2}{2\mu} \ln \frac{\pi u_1}{(1-\pi)u_0}$, which solves $\tilde{a}(K) = 1/2$ under Assumption 4(ii). Its sign is opposite to that of $\frac{\pi u_1}{(1-\pi)u_0} - 1$.

Definition 3. *The reporting policy $\beta_\infty(K)$ is* strongly appealing *if $\beta_\infty'(K) \geq \beta_\infty'(-K)$, and* strongly alarmist *if $\beta_\infty'(K) \leq \beta_\infty'(-K)$, for any K such that $KK_{1/2} > 0$.*

This is a strong definition: If a strongly appealing $\beta_\infty(K)$ is the distribution of some random variable, then its mean (if well-defined), its median, and the average of its upper and lower αth percentiles for any α all have the same sign as $K_{1/2}$.

What is the source of the audience-appeal bias? Under Assumption 4(ii), $\lambda_F(K)^{\frac{1}{6}}$ is proportional to a Gaussian $N(0, 3\sigma^2)$ density and does not skew. Hence bias comes from $\lambda_h(K)^{\frac{1}{6}} \tilde{a}'(K)^{\frac{1}{2}}$.

Its shape is described in Proposition 4.[3]

Proposition 4. *Under Assumptions 3 and 4(ii), the following hold:*

(i) $\tilde{a}(K)$ *is symmetric about* $(K_{1/2}, \frac{1}{2})$*, and* $\tilde{a}'(K)$ *is symmetric about* $K = K_{1/2}$*;*

(ii) $\lambda_h(K)$ *is symmetric about* $K = K_{1/2}$*.*

Proposition 4 implies that $\beta'_\infty(K)$ is proportional to the product of two symmetric curves: $\lambda_F(K)^{\frac{1}{6}}$, which is symmetric about zero, and $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$, which is symmetric about $K_{1/2}$. This insight leads to Proposition 5, which gives sufficient conditions for the strongly appealing and strongly alarmist properties.

Proposition 5. *Assume the integrability of the right-hand side of Eq. (7). Under Assumptions 3 and 4(ii), the following hold:*

(i) $\beta_\infty(K)$ *is strongly appealing if* $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ *is hump-shaped, i.e., increasing on* $(\underline{K}, K_{1/2})$ *and decreasing on* $(K_{1/2}, \bar{K})$*;*

(i)* $\beta_\infty(K)$ *is strongly alarmist if* $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ *is U-shaped, i.e., decreasing on* $(\underline{K}, K_{1/2})$ *and increasing on* $(K_{1/2}, \bar{K})$*;*

(ii) $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ *is hump-shaped (U-shaped) if and only if, for* $a < \frac{1}{2}$*,*

$$\frac{d}{da}\left(\frac{h'(a)^3 h'(1-a)^3}{(-h'(1-a)h''(\tilde{a}) - h''(1-a)h'(a))^2}\right) \geq (\leq)\ 0; \tag{12}$$

(iii) $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ *is hump-shaped if both* $\frac{h''(a)}{h'(a)}$ *and* $\frac{h'''(a)}{h'(a)}$ *are decreasing in* a*.*

The conditions in Proposition 5 involve only h and are simple to verify. Many common utility functions, including cosine-difference, quadratic, log, power (with $\gamma \leq 2$), and exponential utilities, satisfy Proposition 5(i), and all of these except for the log and power utilities satisfy Proposition 5(iii).

Intuitively, Proposition 5 states that the report curve is strongly appealing whenever customizing extreme action recommendations is not too important to the overall utility from the perspective

3. The proofs of Propositions 4, 5 and 6 are in Appendix C.

of $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$. Notice that $\tilde{a}'(K)$ is always hump-shaped, meaning the action recommendation is not sensitive to fundamentals near the extremes. Hence, for $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ to be hump-shaped, $\lambda_h(K)$ must not explode too rapidly near extreme scenarios. This means $\lambda_h(K)$ may itself be hump-shaped, so that customizing extreme action recommendations has little importance; or it may explode but at a rate controllable by $\tilde{a}'(K)$, so that the product $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is still hump-shaped. In the latter situation, customizing extreme action recommendations is highly valuable due to high marginal utility, but nevertheless contributes little to newsworthiness since the action recommendation has little sensitivity to the scenario.

Example 3. Consider the constant relative risk-aversion (CRRA) utility functions $h(a) = \frac{a^{1-\gamma}}{1-\gamma}$ ($\gamma > 0$ and $\gamma \neq 1$) and $h(a) = \ln(a)$ ($\gamma = 1$). The relative risk-aversion γ captures the curvatures of these utilities.

Figure V shows $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ and $\beta'_\infty(K)$ for various γ. The tails of $\lambda_h(K)^{\frac{1}{6}}$ tend to zero for $\gamma \leq 1$ but explode for $\gamma > 1$. When $\gamma \leq 2$, the behavior of $\tilde{a}'(K)^{\frac{1}{2}}$ prevails over that of $\lambda_h(K)^{\frac{1}{6}}$, and the product $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ satisfies Proposition 5(i), implying the strongly appealing property. When $\gamma > 2$, however, $\lambda_h(K)^{\frac{1}{6}}$ prevails and $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is U-shaped, fitting Proposition 5(i*). Now, $\beta'_\infty(K)$ is given by $\lambda_F(K)^{\frac{1}{6}}$ times $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$. Because $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is symmetric about $K_{1/2}$ and is bigger for fundamentals farther from $K_{1/2}$, it scales up $\lambda_F(K)$ more for K with an opposite sign to $K_{1/2}$, making the capacity allocation curve somewhat larger at confirmatory K values. This leads to an interesting alarmist bias for high γ values.

[Insert Figure V here.]

Ultimately, the bias type depends on Eq. (12), which can be rewritten as

$$\frac{d}{da}\ln h'(a) + \frac{d}{da}\ln h'(1-a) + \left(-2\frac{d}{da}\ln\left(\frac{d}{da}\ln\frac{h'(1-a)}{h'(a)}\right)\right) \geq (\leq)\ 0.$$

The first two terms on the left-hand side are the elasticities for the conditional marginal utilities, which express a type of "level sensitivity" that measures the absolute stakes of trembling at a.

Their sum is negative for utilities with non-increasing absolute risk-aversion, which include all of the examples in Table I. The sign reflects a fear of choosing a near zero, rather than near $\frac{1}{2}$, because of concerns about loss coming from betting extreme. The third term is minus twice the elasticity of the elasticity of the conditional marginal utility ratio, a type of "ratio sensitivity" that measures the relative stakes of trembling at a in alternative states. For the utilities in Table I it is positive, capturing the importance of learning information in order to customize actions near $\frac{1}{2}$. This ratio sensitivity dominates the level sensitivity for many utilities, leading to the strongly appealing property. On the other hand, for CRRA utilities with $\gamma > 2$, the opposite occurs, leading to the strongly alarmist property.

Sensationalism. Under Assumption 4(ii) and with integrability in Theorem 2, sensationalism is inevitable. Sensationalism can be described as the sender's dumping two vast regions of tail scenarios into two small bins of extreme reports; more precisely, it occurs when the slope of the report curve becomes very small as K tends to $\pm\infty$. Essentially, the source of sensationalism is the structural assumption that N greatly exceeds n as these quantities tend to infinity. This assumption also determines the proper rescaling of fundamentals and reports used for the asymptotic exercise. The model predicts that sensationalism is inevitable as long as the complexity of the scenario is far greater than an information intermediary can represent, however nuanced her reports may be—a feature of many real-world situations.

Of the three factors in Theorem 2, $\tilde{a}'(K)^{\frac{1}{2}}$ and $\lambda_F(K)^{\frac{1}{6}}$ promote sensationalism with their well-behaved tails while $\lambda_h(K)^{\frac{1}{6}}$ may work for or against it, depending on the tail behavior of the utility.

III.D Contextual Effects

Report distribution. While other variables may be latent, the report is observable. Under Assumption 4(ii), the conditional distribution of the report ρ is

$$\begin{cases} \Pr(\rho = 0|\theta) = \Phi\left(\frac{\underline{K}-\mu_\theta}{\sigma}\right) & \text{if } \rho = 0, \\ \rho|\theta \sim \Phi\left(\frac{\beta_\infty^{-1}(\rho)-\mu_\theta}{\sigma}\right) & \text{if } \rho \in (0,1), \\ \Pr(\rho = 1|\theta) = 1 - \Phi\left(\frac{\bar{K}-\mu_\theta}{\sigma}\right) & \text{if } \rho = 1, \end{cases} \tag{13}$$

and its unconditional distribution is a mixture with mixing probability π.

Below I discuss the effects of several contextual parameters on the report curve and distribution under Assumption 4(ii), in which case $\lambda_H(K)$ depends on K only via $\tilde{a}(K)$, or equivalently via $\frac{2\mu}{\sigma^2}K$.

Definition 4 compares the degrees of audience appeal or alarmism in report curves.

Definition 4. *The report curve* $\beta_\infty^{(1)}(K)$ leans more positive *than another report curve* $\beta_\infty^{(2)}(K)$ *if* $\beta_\infty^{(2)}(K)$ *has first-order stochastic dominance (FOSD) over* $\beta_\infty^{(1)}(K)$.

Relative payoff relevance u_1/u_0. Obviously, u_1 and u_0 are not separately identifiable in report data. The term u_1/u_0, which is separately identifiable from π as is discussed later, affects the report distribution through the report curve. Consider $u_1^{(1)}/u_0^{(1)} > u_1^{(2)}/u_0^{(2)}$ with corresponding report curves $\beta_\infty^{(1)}(K)$ and $\beta_\infty^{(2)}(K)$. Then for cosine-difference, quadratic, log, and power ($\gamma < 2$) utilities, the "likelihood ratio" $\beta_\infty^{(1)\prime}(K)/\beta_\infty^{(2)\prime}(K)$ is strictly decreasing, so that $\beta_\infty^{(2)}(K)$ has FOSD over $\beta_\infty^{(1)}(K)$ and thus $\beta_\infty^{(1)}(K)$ leans more positive. For an exponential utility, it is trivial that $\beta_\infty^{(1)}(K)$ leans more positive. In such situations, the reporting policy will lean more towards the state that has higher payoff relevance. By contrast, for a power utility with $\gamma > 2$, $\beta_\infty^{(1)\prime}(K)/\beta_\infty^{(2)\prime}(K)$ is strictly increasing and so $\beta_\infty^{(1)}(K)$ leans less positive.

When $\beta_\infty^{(1)}(K)$ leans more positive, the conditional and unconditional report distributions for $u_1^{(1)}/u_0^{(1)}$ have FOSD over the corresponding report distributions for $u_1^{(2)}/u_0^{(2)}$. This is because

$\beta_\infty^{(1)}(K) \geq \beta_\infty^{(2)}(K)$ by the FOSD order of report curves, implying $\beta_\infty^{(1)-1}(\rho) \leq \beta_\infty^{(2)-1}(\rho)$ and $\underline{K}^{(1)} \leq \underline{K}^{(2)}$ in Eq. (13).

Belief π. The belief has two effects on the report distribution. First, it affects the report curve in the same way as u_1/u_0. Second, it equals the mixing probability. The latter means π and u_1/u_0 are separately identifiable in report data.

Suppose $\pi^{(1)} > \pi^{(2)}$ and $\beta_\infty^{(1)}(K)$ leans more positive than $\beta_\infty^{(2)}(K)$. Then the conditional report distributions under $\pi^{(1)}$ have FOSD over those under $\pi^{(2)}$, by reasoning analogous to that used for u_1/u_0. The unconditional report distribution under $\pi^{(1)}$ also has FOSD over the one under $\pi^{(2)}$, because, as well as the fact that both conditional distributions are higher in the former, the mixture gives more weight to the conditional distribution for $\theta = 1$ which is higher.

Informativeness μ/σ. The parameters μ and σ are not separately identifiable in report data, although their ratio μ/σ, which captures how informative nature's signals are, may be. To see this, consider setting 1 with parameters $\mu^{(1)}$ and $\sigma^{(1)}$ and setting 2 with parameters $\mu^{(2)} = C\mu^{(1)}$ and $\sigma^{(2)} = C\sigma^{(1)}$, where C is a constant. Then a fundamental $K^{(1)} = K$ in setting 1 is equivalent to the fundamental $K^{(2)} = CK$ in setting 2, for any K. That is, $K^{(1)}$ and $K^{(2)}$ are the same quantiles in their respective conditional distributions and also have the same report. Therefore, the conditional and unconditional report distributions in both settings are the same. Essentially, applying an affine transformation to the fundamental space does not change the problem; what matters is the standardized fundamental K/μ. For the same reason, to identify the fundamental $K = \beta_\infty^{-1}(\rho)$ corresponding to some observed $\rho \in (0, 1)$, we must first pin down either μ or σ, which we can do without loss of generality.

Informativeness contributes to sensationalism. For instance, let h satisfy Proposition 5(i), and fix μ. Let σ tend to zero. Then the hump-shaped curves $\lambda_H(K)^{\frac{1}{6}}$ and $\lambda_F(K)^{\frac{1}{6}}$ concentrate their masses around $K = 0$, and the limit of the curve $\beta_\infty(K)$ becomes extremely sensationalist.

Proposition 6. *Assume the integrability of the right-hand side of Eq.* (7)*. Under Assumptions 3 and 4(i), we have* $\beta_\infty(K) \to 0$ *for* $K < 0$ *and* $\beta_\infty(K) \to 1$ *for* $K \geq 0$ *as* $\sigma \to 0$*, if* $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ *is hump-shaped.*

This happens because high informativeness makes the state very clear for side scenarios, so the sender allocates most capacity in the intermediate region of the fundamental space, creating high sensitivity of the report to the fundamental. Consequently, the conditional report distributions tend to a mass of one at 0 and 1, respectively, and the unconditional distribution, a mixture of the two, is highly dispersed.

Utility. Proposition 5 and Example 3 outline the effects of the choice of utility function: The utility affects the shape of $\lambda_H(K)$ and thus changes the bias implications of the other contextual parameters.

IV IMPLICATIONS: INTUITION AND METHODOLOGY

This paper makes two major contributions. First, it offers a novel perspective on content bias—a phenomenon relevant to business, politics, and everyday life—by relating it to a simple, practically motivated information selection mechanism. Second, it provides a micro-foundation for content analysis. In the literature, the study of content data is generally limited to reduced-form analysis or data mining, which makes it difficult to see beyond the literal meaning of the content, to account for the environment in which it was produced, and to rigorously interpret the methodology and results of the analysis. I address this gap by proposing a tractable model of content formation.

As a tool for structural analysis of content, my model has four important features. (1) It is not a black box; rather, it is based on a common content formation mechanism. (2) It distinguishes between the literal meaning and the underlying true meaning of the content, while linking both to the context. This allows the researcher, as a third-party observer, to interpret the content from the perspective of its intended audience, within the original economic environment. (3) The model clearly shows contextual effects. (4) The conditionally Gaussian fundamental and the sentiment- or frequency-based content measure match specifications frequently seen in the literature.

Scope of applicability. This model applies to any form of content with the following characteristics: (1) The content facilitates decision-making by presenting information about two competing hypotheses. (2) The content is subject to fairly rigid length limits, necessitating information selection. (3) The analysis is not focused on audience heterogeneity or agency issues.

Media reports are an obvious example of such content. Other examples include briefings for busy decision-makers, consultancy reports, filings publishing information, and essays that selectively present evidence for argumentation.

IV.A Media Bias

Media outlets often produce narratives that skew reality. Demand-side theories of bias (see, e.g., Suen 2004, Mullainathan and Shleifer 2005, or Gentzkow and Shapiro 2006) attribute this to the incentive to satisfy an audience, and they all require some form of discrepancy between the beliefs or preferences of the agents.

This paper proposes a novel demand-side framework in which bias arises from information selection, independent of any such discrepancies, and rationalizes slant and media narratives. A media outlet's coverage may exhibit bias simply because its editors select the pieces that will convey the most useful information, given the audience's preferences and beliefs and the size constraints of the platform. In particular, the biases of audience appeal and sensationalism may arise simply as manifestations of optimal communication efficiency as tacitly agreed on by the outlet and its audience, not as the result of any intention to mislead or pander to the audience.

For instance, the state θ may be the subject of conflicting assertions of fact by left- and right-wing politicians, and individuals may then consult a newspaper for information about the true state in order to choose a policy to support. Suppose an individual has more confidence in the rightist view, or will benefit more if it is true. To help such an individual maximize his utility, the newspaper's selection of stories should be biased toward the rightist view. The same, of course, applies for leftist readers. In addition, to maximize communication efficiency, the newspaper should exaggerate the direction (left or right) of the evidence available to it.

Should we worry about such biases in media coverage? The model says no. Of course, it assumes full rationality, which may not hold in the real world, and it disregards other possible channels for bias, as well as sociological and cultural factors. Nonetheless, it provides a perspective on why the existence of bias may be reasonable.

IV.B Empirical Implications

Data and model preparation. In practice, content data usually need to be quantified, or *tokenized*, for analysis (see the survey of Gentzkow, Kelly, and Taddy 2019). Tokenization is the division of the full content into basic tokens—such as events, pieces of evidence, or phrases—that correspond to the content elements in a model. For example, in my model, each token should correspond to a reported element r_i supporting one realization of the state over the other.

One feature of such tokenization is the irrelevance of the order of tokens. My model assumes that the relative locations of reported elements (e.g., news stories) carry no information, as long as they appear in the same piece of content (e.g., the same newspaper). This is a rather strong assumption if, for example, our content is a text and the tokens are individual phrases. However, similar assumptions are common in the empirical literature—for example, in the bag-of-words approach to textual analysis.

Example 4 illustrates how results such as Theorem 2 and Eq. (13) can help us establish a structural model given a concrete problem.

Example 4. Consider an investor with CRRA utility $u(w) = \frac{1}{1-\gamma} w^{1-\gamma}$ for a portfolio worth w, who allocates one dollar between two categories of assets: the A category, which returns $R_{A,1}$ in a boom ($\theta = 1$) and $R_{A,0}$ in a bust ($\theta = 0$), and the B category, which returns $R_{B,1}$ in a boom and $R_{B,0}$ in a bust. So that neither asset is dominant, assume $R_{A,1} > R_{B,1}$ and $R_{A,0} < R_{B,0}$. (These returns are parameters that can be calibrated using real-world data, e.g., data on mean returns for certain universes of assets in a boom or a bust.) The investor shares the market belief π in the likelihood of a boom. A financial newspaper targeting him reports on a conditionally Gaussian shock K. To match this setup to the model, let

$$h(a) = \frac{1}{1-\gamma}(a+C)^{1-\gamma},\ u_1 = (R_{A,1} - R_{B,1})^{1-\gamma},\text{ and } u_0 = (R_{B,0} - R_{A,0})^{1-\gamma},$$

where

$$C = \frac{1}{2}\left(\frac{R_{B,1}}{R_{A,1} - R_{B,1}} + \frac{R_{B,0}}{R_{B,0} - R_{A,0}} - 1\right) \geq 0.$$

Then $\underline{K} = -\frac{\gamma\sigma^2}{2\mu}\ln\frac{1+C}{C} \geq -\infty$ and $\bar{K} = \frac{\gamma\sigma^2}{2\mu}\ln\frac{1+C}{C} \leq \infty$. By Theorem 2,

$$\beta'_\infty(K) \propto \exp\left(-\frac{K^2}{6\sigma^2}\right)\left((\pi u_1)^{\frac{1}{\gamma}}\exp\left(\frac{\mu}{\gamma\sigma^2}K\right) + ((1-\pi)u_0)^{\frac{1}{\gamma}}\exp\left(-\frac{\mu}{\gamma\sigma^2}K\right)\right)^{\frac{\gamma-2}{3}},$$

which is the distribution for the power case in Table I, truncated to $(\underline{K}, \bar{K})$.

Parameter identification from report data. As previously noted, the parameters that may be identifiable from report data include π, u_1/u_0, μ/σ, and certain parameters governing the shape of the utility function given its form. For a fixed μ or σ, we can extract the fundamental underlying a given report. If also given the data of proxies for other parameters or for the fundamental, we may have access to not just the marginal report distribution but the joint distribution, in which case Theorem 2 and Eq. (13) will enable us to parameterize a model accounting for context changes.

Sentiment analysis and model (mis-)specification. My model is particularly relevant for sentiment analysis. Conventional sentiment research proceeds in two steps. The first is to define a measure based on the frequency of certain elements in the content data (for instance, the proportion of negative words in a text, or of positive stories in a collection). This measure may be called something like the sentiment, tone, attitude, or pessimism; it is seen as a proxy for variables such as beliefs, preferences, or fundamentals. The second step is to perform regressions or data mining using that measure. In this process, however, it is often unclear exactly how information is embedded in the measure, which makes interpretation difficult. My paper speaks to this issue, since the reports k/n closely resemble a frequency measure.

The model in this paper predicts that a sentiment measure (or other frequency measure) derived from content data is actually a nonlinear combination of fundamentals, preferences, and beliefs. This means that researchers performing sentiment analysis should carry out an intermediate step: Rather than directly analyzing the measured sentiment data, they should extract the quantities of interest from these data. For instance, a researcher who wants to study a text-based fundamental K should estimate a model which predicts K before the second step, or integrate the textual model with the second step and do a full analysis. If she skips this intermediate step, she risks misspecification, e.g., by confusing the report k (or ρ) with the fundamental (or shock) K, or with learning outcomes, such as the posterior or the action.

One potential source of misspecification is the nonlinearity of the report curve. Suppose a researcher mistakes the sentiment (i.e., the report) for the shock and uses it to explain a shock proxy y (consisting of the shock plus noise) in a regression. Sensationalism predicts that the impact of extreme shocks on y is overstated compared to that of moderate shocks, because

$$\frac{\partial E[y|K]}{\partial \rho} = \frac{\partial E[y|K]}{\partial K}\frac{\partial K}{\partial \rho},$$

where $\partial K/\partial \rho$ is large for extreme shocks and small for moderate shocks. The audience-appeal (alarmist) bias predicts that the impact of confirmatory (contradictory) shocks becomes obscure. With audience appeal, for instance, a report that is over 50% confirmatory may represent both confirmatory shocks and some moderately contradictory shocks. If an indicator variable of over 50% confirmatory sentiment is used to explain a shock proxy in a regression, the significance and magnitude may both be mitigated. If such an indicator is for contradictory sentiment, however, the significance and magnitude may both be exaggerated.

Misspecification may also occur if contextual variables are not properly accounted for when the context changes. Consider for instance a dataset in which some data are generated under a high u_1/u_0 and others under a low u_1/u_0. Then, even if we assume that the data-generating process for the fundamental is the same, the reports will have different conditional means on the two contexts. Therefore, if a researcher mistakes the sentiment for the shock and tries to use it to explain a variable y that is related to the payoff relevance u_1/u_0 but orthogonal to the shock, she may find a spurious significance in regressing y on the sentiment and reach a false conclusion.

V BEYOND CONTENT: AN ANALYSIS OF RATINGS

My model also applies to non-content forms of data, such as product star ratings.

Consider a customer (*she*) who rates her experience of a product on a five-star scale ($k = 0, ..., 4$) for the benefit of a later shopper (*he*). The product's type θ is either good (1) or bad (0). The customer experience K is a random variable whose conditional distributions on θ satisfy the monotone likelihood ratio property. The later shopper observes the rating and chooses an action a, which stands for the probability or amount of purchase. Both agents' preferences and beliefs are aligned.

While a star rating is not a summary of reported elements as in my original model, this problem is similar in structure to the earlier problem and can be solved using a slight adjustment of the model. It is natural to assume that the customer experience K is highly complex, so the assignment of a star rating k to it involves significant information compression; that is, as before, the fundamental space is much larger than the report space.

For simplicity, as in the baseline model, I start by assuming the fundamental space is a finite set of integers. Importantly, Propositions 1, 2 and 3 extend to this problem, because their proofs do not depend on the specific distributions of the signals s_i or even K, but only on the strict monotone likelihood ratio property of $K|\theta$. Therefore the consumer's information structure (or rating strategy) is a pure strategy, surjective, and characterized by cutoffs. To find the optimal ("rating-generating") information structure, we can impose self-consistency (which is natural since, in practice, customers give higher ratings for better experiences). Next I let the fundamental space become dense and satisfy (i) or (ii) of Assumption 4. The report curve is then obtained by numerically solving Eq. (6), or approximated using Theorem 2.

My results help explain why real-world ratings often look skewed. For example, if a product has many five- and four-star ratings but few low ratings, it may be that customers go in expecting it to be good, or benefit more from purchasing a good product than from avoiding the purchase of a bad product, so that they overrate even moderately positive experiences and thus have more lower star ratings available to distinguish between bad experiences. A highly dispersed distribution of ratings may indicate that customer experience is highly informative about quality.

My model can also be extended to study, for example, students' exam scores as a reflection of their skill, with the scores serving as input for a decision-maker's choices.

Essence of the model. This extension reveals the model's essential mathematical structure. The fundamental values and the report values form equidistant sequences. An information structure is an increasing mapping from the fundamental to the report that takes the lowest (highest) fundamental to the lowest (highest) report. That mapping must involve some monotone pooling, and the sender's problem is to pool in a way that maximizes the expected utility. Such pooling creates non-linearity in the mapping, which manifests as various interesting phenomena in practice.

VI CONCLUDING REMARKS

This paper identifies information selection due to physical capacity constraints as a cause of content bias, including audience appeal and sensationalism, and gives asymptotic characterizations of these phenomena. Bias stems from the sender's strategy of compressing fundamental information based on its newsworthiness. In particular, bias in my model does not mislead the receiver and improves welfare.

The content generation channel I model is independent of any discrepancies between the agents' preferences or payoffs. If such discrepancies are of concern in a real-world problem, other persuasion-related channels may also be in place. In that case, the optimal information structure may involve mixed strategies, and the criteria of honesty and self-consistency may conflict or constrain optimization. However, this paper's findings remain relevant if information selection is also a concern.

Importantly, my model distinguishes between the content's literal meaning and its underlying true meaning, while connecting the two via a tractable and smooth function that incorporates parameters capturing the economic context. It can be applied in many settings involving information selection, including the study of media slant, sentiment analysis (and other forms of content analysis involving frequency measures), and ratings analysis. It may also be useful in strengthening empirical analyses of content data, by helping researchers account for contextual factors and avoid specification errors.

Author Affiliation: City University of Hong Kong.

Appendix: Content Bias and Information Compression

Jinge Liu

A Propositions 1, 2, and 3

Let $p_K := \Pr(K|\theta = 1)$, $q_K := \Pr(K|\theta = 0)$. The posterior beliefs are $\pi'_k := \Pr(\theta = 1|k) = \pi \Pr(k|\theta = 1)/\Pr(k)$. Hence the ex-ante utility is

$$\begin{aligned} U &= \sum_{k=0}^{n} \Pr(k) \left\{ \pi'_k u_1 h(a^*(k)) + (1-\pi'_k) u_0 h(1-a^*(k)) \right\} \\ &= \sum_{k=0}^{n} \left\{ \pi \Pr(k|\theta=1) u_1 h(a^*(k)) + (1-\pi) \Pr(k|\theta=0) u_0 h(1-a^*(k)) \right\} := \sum_{k=0}^{n} U_k, \end{aligned}$$

where $a^*(k)$ is $\arg\max_a \pi'_k u_1 h(a) + (1-\pi'_k) u_0 h(1-a)$ if $\sum_K \sigma_{Kk} > 0$, and may be any value otherwise. The domain of U is $[0,1]^{(N+1)\times(n+1)}$ for $\{\sigma_{Kk}\}_{K=0,\ldots,N;k=0,\ldots,n}$. The utility is continuous on a compact set, so there always exists a solution to maximizing this utility.

Proofs of Propositions 1 and 2. We will show that Proposition $1'$, below, implies Proposition 2, which in turn implies Proposition 1.

Proposition $1'$. *Under Assumption 1, there exists a pure-strategy equilibrium.*

Proof of Proposition $1'$. For σ_{Kk},

$$\begin{aligned} \frac{\partial U}{\partial \sigma_{Kk}} = \frac{\partial U_k}{\partial \sigma_{Kk}} &= \frac{\partial U_k(\{\sigma_{Kk}\}_{K=0,\ldots,N}, a^*(k))}{\partial \sigma_{Kk}} + \frac{\partial U_k(\{\sigma_{Kk}\}_{K=0,\ldots,N}, a^*(k))}{\partial a^*(k)} \frac{\partial a^*(k)}{\partial \sigma_{Kk}} \\ &= \frac{\partial U_k(\{\sigma_{Kk}\}_{K=0,\ldots,N}, a^*(k))}{\partial \sigma_{Kk}} \qquad \text{(by optimality of } a^*(k)\text{)} \\ &= \pi u_1 p_K h(a^*(k)) + (1-\pi) u_0 q_K h(1-a^*(k)). \end{aligned} \tag{A.1}$$

The optimality of $a^*(k)$ for k such that $\sum_K \sigma_{Kk} > 0$ is characterized by the first-order condition

$0 = \pi'_k u_1 h'(a) - (1-\pi'_k)u_0 h'(1-a)$. By the implicit function theorem,

$$\frac{\partial a^*(k)}{\partial \sigma_{Kk}} = -\frac{\pi u_1 p_K h'(a^*(k)) - (1-\pi)u_0 q_K h'(1-a^*(k))}{\pi u_1(\sum_{K'=1}^{N} p_{K'}\sigma_{K'k})h''(a^*(k)) + (1-\pi)u_0(\sum_{K'=1}^{N} q_{K'}\sigma_{K'k})h''(1-a^*(k))}.$$

Hence,

$$\begin{aligned}\frac{\partial^2 U}{\partial \sigma_{Kk}^2} &= (\pi u_1 p_K h'(a^*(k)) - (1-\pi)u_0 q_K h'(1-a^*(k)))\frac{\partial a^*(k)}{\partial \sigma_{Kk}} \\ &= -\frac{(\pi u_1 p_K h'(a^*(k)) - (1-\pi)u_0 q_K h'(1-a^*(k)))^2}{\pi u_1(\sum_{K'=1}^{N} p_{K'}\sigma_{K'k})h''(a^*(k)) + (1-\pi)u_0(\sum_{K'=1}^{N} q_{K'}\sigma_{K'k})h''(1-a^*(k))},\end{aligned} \tag{A.2}$$

which is ≥ 0 since $h''(\cdot) < 0$. Also,

$$\frac{\partial^2 U}{\partial \sigma_{Kk_1}\partial \sigma_{Kk_2}} = 0, \tag{A.3}$$

implying that $\partial U/\partial \sigma_{Kk}$ does not depend on $\sigma_{Kk'}$ for $k' \neq k$.

Suppose $\sigma^*_{Kk_1} \in (0,1)$ is in an optimal information structure. Then there must exist $\sigma^*_{Kk_2} \in (0,1)$. Both $\sum_K \sigma^*_{Kk_1} > 0$ and $\sum_K \sigma^*_{Kk_2} > 0$ hold, so Eqs. (A.2) and (A.3) hold. Writing U for $U(\sigma_{Kk_1}, \sigma_{Kk_2})$, we consider the following cases.

Case (i): $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} \neq \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$ (without loss of generality, we may assume $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} > \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$). Then for a small ε, $\sigma^{**}_{Kk_1} = \sigma^*_{Kk_1} + \varepsilon$ and $\sigma^{**}_{Kk_2} = \sigma^*_{Kk_2} - \varepsilon$ will improve U. This is because $\frac{\partial U}{\partial \sigma_{Kk_1}}$ is continuous in σ_{Kk_1}, as is $\frac{\partial U}{\partial \sigma_{Kk_2}}$ in σ_{Kk_2}; thus there exists $\varepsilon > 0$ such that $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}+\varepsilon'} > \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}-\varepsilon'}$ for all $\varepsilon' \in (0, \varepsilon]$. The utility increase from changing to $\sigma^{**}_{Kk_1}$ and $\sigma^{**}_{Kk_2}$ is $U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2}) = U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) + U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2})$, which equals $\int_0^\varepsilon \frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}+\varepsilon'} d\varepsilon' - \int_0^\varepsilon \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}-\varepsilon''} d\varepsilon'' > 0$ by Eq. (A.3). This contradicts optimality.

Case (ii): $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} = \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$, and there is no $\delta > 0$ such that both $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} = 0$ on $(\sigma^*_{Kk_1} - \delta, \sigma^*_{Kk_1} + \delta)$ and $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} = 0$ on $(\sigma^*_{Kk_2} - \delta, \sigma^*_{Kk_2} + \delta)$; assume without loss that $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} > 0$ on $(\sigma^*_{Kk_1}, \sigma^*_{Kk_1} + \delta_0]$. Then $\sigma^{**}_{Kk_1} = \sigma^*_{Kk_1} + \delta_0$ and $\sigma^{**}_{Kk_2} = \sigma^*_{Kk_2} - \delta_0$ will improve U. Here is why: The utility increase is $U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2}) = U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) + U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2})$, which equals $U(\sigma^*_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2}) + U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) -$

$U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2})$ by Eq. (A.3). Define a continuous function f by $f(x) := U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2} + x)$ if $x \in [-\delta_0, 0]$ and $f(x) := U(\sigma^*_{Kk_1} + x, \sigma^*_{Kk_2})$ if $x \in (0, \delta_0]$; then f is convex on $[-\delta_0, \delta_0]$ and strictly convex on $(0, \delta_0]$. The utility increase is $f(\delta_0) + f(-\delta_0) - 2f(0)$, which is positive by Jensen's inequality, contradicting optimality.

Case (iii): $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} = \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$, and there exists $\delta > 0$ such that both $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} = 0$ on $(\sigma^*_{Kk_1} - \delta, \sigma^*_{Kk_1} + \delta)$ and $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} = 0$ on $(\sigma^*_{Kk_2} - \delta, \sigma^*_{Kk_2} + \delta)$. Then one of the following two cases will occur.

Case (iii-a): $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} > 0$ or $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} > 0$ somewhere on $(0, \sigma^*_{Kk_1} + \sigma^*_{Kk_2})$; assume without loss that for some constant $\sigma^0 \in (\sigma^*_{Kk_1}, \sigma^*_{Kk_1} + \sigma^*_{Kk_2})$, $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} > 0$ when $\sigma_{Kk_1} \in (\sigma^0, \sigma^*_{Kk_1} + \sigma^*_{Kk_2})$, $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} = 0$ when $\sigma_{Kk_1} \in (\sigma^*_{Kk_1}, \sigma^0]$, and $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} = 0$ when $\sigma_{Kk_2} \in (\sigma^*_{Kk_2} - (\sigma^0 - \sigma^*_{Kk_1}), \sigma^*_{Kk_2})$. Letting $\sigma^0_{Kk_1} = \sigma^0$ and $\sigma^0_{Kk_2} = \sigma^*_{Kk_2} - (\sigma^0 - \sigma^*_{Kk_1})$, we get a new information structure falling under case (ii), with the same utility. Applying the reasoning of case (ii) by taking $\sigma^0_{Kk_1}$ and $\sigma^0_{Kk_2}$ as the new $\sigma^*_{Kk_1}$ and $\sigma^*_{Kk_2}$, we get a contradiction to optimality.

Case (iii-b): $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} = 0$ and $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} = 0$ for $\sigma_{Kk_1}, \sigma_{Kk_2} \in (0, \sigma^*_{Kk_1} + \sigma^*_{Kk_2})$. Let $\sigma^{**}_{Kk_1} = \sigma^*_{Kk_1} + \sigma^*_{Kk_2}$ and $\sigma^{**}_{Kk_2} = 0$.

Case (iii-b-1): $\sigma^*_{Kk_1} + \sigma^*_{Kk_2} = 1$. In this case, the new strategy has the same utility with no mixing for K. If there is no mixing for any other fundamental K', then we have found a pure strategy that delivers the same utility as the optimal mixed strategy. If some other K' has mixing, then let K' be the new K and iterate the discussion of cases (i)–(iii).

Case (iii-b-2): $\sigma_{Kk_1^*} + \sigma_{Kk_2^*} < 1$. Then there exists k_3 such that $\sigma^*_{Kk_3} \in (0, 1)$. Let k_3 be the new k_2 and iterate the discussion of cases (i)–(iii).

Hence the only case in which the existence of mixing is plausible is (iii-b-1). An equilibrium mixed strategy exists only when it delivers the same utility as a pure strategy. Therefore, a pure-strategy equilibrium always exists. □

Lemma 1. *Under Assumptions 1 and 2(i), the equilibrium involves either a pure strategy, or a mixed strategy that yields the same utility as a pure strategy in which* $\sum_K \sigma_{Kk} = 0$ *for some* k.

Proof of Lemma 1. From the proof of Proposition $1'$, mixing can occur only in case (iii-b-1), when $\sigma^*_{Kk_1} + \sigma^*_{Kk_2} = 1$. In that case, $\partial U / \partial \sigma_{Kk_1}$ is constant, so $a^*(k_1)$ is constant with respect to σ_{Kk_1} by Eq. (A.1). Hence, by the first-order condition, σ_{Kk_1} does not affect $\pi'_{k_1}/(1-\pi'_{k_1})$. Since the values of p_K/q_K differ across K, we have $\sigma_{K'k_1} = 0$ for all $K' \neq K$. The same applies to k_2; thus K is the only fundamental value that maps to k_1 or k_2. Therefore the utility-equivalent pure strategy must have either $\sum_K \sigma^{**}_{Kk_1} = 0$ or $\sum_K \sigma^{**}_{Kk_2} = 0$; i.e., some k ends up unused. $\square$

Proof of Proposition 2. By Proposition $1'$ and Lemma 1, it suffices to consider pure-strategy equilibria. The proof is by contradiction. Suppose, in a pure-strategy equilibrium, $\sum_K \sigma_{Kk_1} = 0$. Then $N+1$ fundamentals are mapped to at most $(n+1)-1 = n \leq N$ reports, so some report is associated with $m \geq 2$ fundamentals. Let k_2 be such a report, representing the set of fundamentals $\{K^{(1)}, ..., K^{(m)}\}$. The contribution of k_2 to the utility is U_{k_2}, and the optimal action for k_2 is $a^*(k_2)$. Consider an alternative pure strategy in which $K^{(1)}$ is mapped to k_1, $\{K^{(2)}, ..., K^{(m)}\}$ is mapped to k_2, and the rest of the strategy is unchanged. Let U'_{k_1} and U'_{k_2} denote the contributions of k_1 and k_2 to the utility, and $a'^*(k_1)$ and $a'^*(k_2)$ the optimal actions. Then

$$
\begin{aligned}
U_{k_2} &= \Pr\Big(K \in \{K^{(1)}, ..., K^{(m)}\}\Big) E[u(a^*(k_2);\theta) | K \in \{K^{(1)}, ..., K^{(m)}\}] \\
&= \Pr\Big(K \in \{K^{(1)}, ..., K^{(m)}\}\Big) E[E[u(a^*(k_2);\theta) | \tilde{1}_{K=K^{(1)}}] | K \in \{K^{(1)}, ..., K^{(m)}\}] \\
&< \Pr\Big(K \in \{K^{(1)}, ..., K^{(m)}\}\Big) E[E[\max_a u(a;\theta) | \tilde{1}_{K=K^{(1)}}] | K \in \{K^{(1)}, ..., K^{(m)}\}] \\
&= \Pr\Big(K = K^{(1)}\Big) E[u(a'^*(k_1);\theta) | K = K^{(1)}; K \in \{K^{(1)}, ..., K^{(m)}\}] + \\
&\quad ... + \Pr\Big(K \in \{K^{(2)}, ..., K^{(m)}\}\Big) E[u(a'^*(k_2);\theta) | K \neq K^{(1)}, K \in \{K^{(1)}, ..., K^{(m)}\}] \\
&= U'_{k_1} + U'_{k_2},
\end{aligned}
$$

where $\tilde{1}_{K=K^{(1)}}$ is a random variable that equals 1 when $K = K^{(1)}$ and 0 otherwise. The inequality is strict because $a^*(k_2)$, $a'^*(k_1)$ and $a'^*(k_2)$ cannot coincide, since p_K/q_K is strictly monotone. This contradicts optimality. $\square$

Proof of Proposition 1. Proposition 2 rules out the mixed-strategy case in Lemma 1. □

Proof of Proposition 3. Let $x_K = \pi u_1 p_K$ and $y_K = (1-\pi)u_0 q_K$. Each K is fully characterized by (x_K, y_K). Let $\eta_K := \frac{y_K}{x_K} = \frac{\pi}{1-\pi}\frac{u_1}{u_0}\Lambda(K)$.

Lemma 2. *Let Assumption 1 hold, let the fundamental space be $\{K_0, ..., K_N\}$ with all η_{K_i} different, denote the optimal partition by $\{B_k\}_{k=0}^{n}$, and suppose B_n consists of $m \geq 2$ fundamentals (by Proposition 2, $m \leq N-n+1$). Define the fundamental v by $x_v = \sum_{K \in B_n} x_K$ and $y_v = \sum_{K \in B_n} y_K$. Then, for the alternative problem with fundamental space $\{K_0, ..., K_N\} \cup \{v\} \backslash B_n$ and $n+1$ reports, the optimal partition is $\{B_k\}_{k=0}^{n-1} \cup \{\{v\}\}$.*

Proof of Lemma 2. Under the partition $\{B_k\}_{k=0}^{n}$, the utility is

$$U = \sum_{k=0}^{n} \left\{ \left(\sum_{K \in B_k} x_K \right) h(a^*(k)) + \left(\sum_{K \in B_k} y_K \right) h(1 - a^*(k)) \right\},$$

where $a^*(k)$ is determined by the conditional probabilities of fundamentals through $\sum_{K \in B_k} x_K / \sum_{K \in B_k} y_K$. Therefore, any utility delivered by a feasible partition of $\{K_0, ..., K_N\} \cup \{v\} \backslash B_n$ can be delivered by a feasible partition of $\{K_0, ..., K_N\}$, specifically by replacing v, in the partition set of $\{K_0, ..., K_N\} \cup \{v\} \backslash B_n$ containing v, by all of the elements of B_n. That is, the utilities obtained by partitioning $\{K_0, ..., K_N\} \cup \{v\} \backslash B_n$ form a subset of those obtained by partitioning $\{K_0, ..., K_N\}$. Therefore, if $\{B_k\}_{k=0}^{n}$ is optimal, then $\{B_k\}_{k=0}^{n-1} \cup \{\{v\}\}$, as a partition of $\{K_0, ..., K_N\} \cup \{v\} \backslash B_n$ that delivers the same utility, must be optimal. □

Proof of Proposition 3. The proof is by induction. Let $P(N, n)$ be the problem of choosing n reported elements from N signals. We want to show that the solution of $P(N, n)$ involves an ordered partition of the fundamental space. We first induct from the base cases $P(2, 1)$ and $P(3, 1)$ to establish the assertion for $P(N, 1)$, then deduce it for $P(N, n)$.

Step 1. We prove that the solution to $P(N, 1)$ involves an ordered partition.

Step 1.1. Here we prove that the solution to $P(2, 1)$ involves an ordered partition.

The proof is by contradiction. Consider the following strategy: (1) The fundamentals $0, 1, 2$ map to k_0 with probabilities $1-s, 0, 1-t$ respectively; the optimal action is a. (2) The fundamentals $0, 1, 2$ map to k_1 with probabilities $s, 1, t$ respectively; the optimal action is b. Thus,

$$
\begin{aligned}
U = ((1-s)x_0 + (1-t)x_2)h(a) + ((1-s)y_0 + (1-t)y_2)h(1-a) + ... \\
+ (sx_0 + x_1 + tx_2)h(b) + (sy_0 + y_1 + ty_2)h(1-b).
\end{aligned}
$$

We need to show that $s = t = 0$ is a suboptimal strategy. By the envelope theorem,

$$
\frac{\partial U}{\partial s} = x_0(h(b) - h(a)) + y_0(h(1-b) - h(1-a)), \tag{A.4}
$$

$$
\frac{\partial U}{\partial t} = x_2(h(b) - h(a)) + y_2(h(1-b) - h(1-a)). \tag{A.5}
$$

We consider two cases.

Case (i): $a \neq b$ at $s = t = 0$. Then as long as one of the two partial derivatives is positive at $s = 0$ or $t = 0$—say, $\frac{\partial U}{\partial s}|_{s=0} > 0$—the pure strategy at $s = t = 0$ is strictly worse than a strategy with a small positive s. But the latter is a mixed strategy and thus suboptimal by Proposition 1. Therefore we only need to show that Eq. (A.4) or Eq. (A.5) is positive. We do so by contradiction. Suppose both Eq. (A.4) and Eq. (A.5) are nonpositive at $s = t = 0$. Then $\eta_1 = \frac{h'(b)}{h'(1-b)}$ by the optimality of b. On the other hand, if $a > b$, then

$$
\eta_2 \leq \frac{h(a) - h(b)}{h(1-b) - h(1-a)} = \frac{(h(a) - h(b))/(a-b)}{(h(1-b) - h(1-a))/((1-b) - (1-a))}
$$

by Eq. (A.5), so $\eta_1 < \eta_2 \leq \frac{(h(a)-h(b))/(a-b)}{(h(1-b)-h(1-a))/((1-b)-(1-a))}$. Since $h(\cdot)$ is strictly increasing and concave, we have $\frac{h(a)-h(b)}{a-b} < h'(b)$ and $\frac{h(1-b)-h(1-a)}{(1-b)-(1-a)} > h'(1-b)$; hence $\eta_1 < \frac{h'(b)}{h'(1-b)}$, a contradiction. And if $a < b$, then

$$
\eta_0 \geq \frac{h(a) - h(b)}{h(1-b) - h(1-a)} = \frac{(h(a) - h(b))/(a-b)}{(h(1-b) - h(1-a))/((1-b) - (1-a))}
$$

by Eq. (A.4), and hence, by analogous arguments, $\eta_1 > \frac{h'(b)}{h'(1-b)}$, a contradiction.

Case (ii): $a = b$ at $s = t = 0$. Then $\frac{\partial U}{\partial s}|_{s=0} = \frac{\partial U}{\partial t}|_{t=0} = 0$. The utility is $U|_{s=t=0} = (x_0 + x_1 + x_2)h(a) + (y_0 + y_1 + y_2)h(1-a)$. Consider the strategy at $s = t = 1$. By the optimality of a and b, $\frac{y_0+y_2}{x_0+x_2} = \frac{h'(a)}{h'(1-a)} = \frac{h'(b)}{h'(1-b)} = \frac{y_1}{x_1}$, so $\frac{h'(a)}{h'(1-a)} = \frac{h'(b)}{h'(1-b)} = \frac{y_0+y_1+y_2}{x_0+x_1+x_2}$. Therefore, for $s = t = 1$, the optimal action is also a. The utility satisfies $U|_{s=t=1} = U|_{s=t=0}$. However, the strategy at $s = t = 1$ does not use both reports and thus is suboptimal by Proposition 2. Hence the strategy at $s = t = 0$ is also suboptimal. This completes Step 1.1.

Step 1.2. Prove that the solution to $P(3,1)$ involves an ordered partition.

The proof is by contradiction. The possible non-ordered partitions $\{B_k\}_{k=0}^1$ are $\{\{1\},\{0,2,3\}\}$, $\{\{2\},\{0,1,3\}\}$, $\{\{0,3\},\{1,2\}\}$, and $\{\{0,2\},\{1,3\}\}$. Below we examine each.

Case (i): $\{\{1\},\{0,2,3\}\}$ or $\{\{2\},\{0,1,3\}\}$ is optimal. Suppose without loss of generality that $\{\{1\},\{0,2,3\}\}$ is optimal. Consider another problem with fundamentals $\{0,1,v\}$, where v is defined by $x_v = x_2 + x_3$ and $y_v = y_2 + y_3$. By Lemma 2, the optimal partition must be $\{\{1\},\{0,v\}\}$. However, the problem is $P(2,1)$, and since $\eta_0 < \eta_1 < \eta_v$, by Step 1.1, $\{\{1\},\{0,v\}\}$ is suboptimal because it is not ordered, a contradiction.

Case (ii): $\{\{0,3\},\{1,2\}\}$ is optimal. Consider another problem with fundamentals $\{0,v,3\}$ where v is defined by $x_v = x_1 + x_2$ and $y_v = y_1 + y_2$. By Lemma 2, the optimal partition must be $\{\{v\},\{0,3\}\}$. However, the problem is $P(2,1)$, and since $\eta_0 < \eta_v < \eta_3$, by Step 1.1, $\{\{v\},\{0,3\}\}$ is suboptimal because it is not ordered, a contradiction.

Case (iii): $\{\{0,2\},\{1,3\}\}$ is optimal. We consider three cases.

Case (iii-a): $a > b$. Consider the following strategy: (1) The fundamentals $0,1,2,3$ map to k_0 with probabilities $1, s, 1-t, 0$ respectively; the optimal action is a. (2) The fundamentals $0,1,2,3$ map to k_1 with probabilities $0, 1-s, t, 1$ respectively; the optimal action is b. Then

$$U = (x_0 + sx_1 + (1-t)x_2)h(a) + (y_0 + sy_1 + (1-t)y_2)h(1-a) + ...$$
$$+ ((1-s)x_1 + tx_2 + x_3)h(b) + ((1-s)y_1 + ty_2 + y_3)h(1-b).$$

We need to show that $s = t = 0$ is a suboptimal strategy. By the envelope theorem,

$$\frac{\partial U}{\partial s} = x_1(h(a) - h(b)) + y_1(h(1-a) - h(1-b)),$$
$$\frac{\partial U}{\partial t} = -x_2(h(a) - h(b)) - y_2(h(1-a) - h(1-b)).$$

As in Step 1.1, it now suffices to show that one of the two partial derivatives is positive at $s = 0$ or $t = 0$. Suppose both are nonpositive; then $\eta_2 \leq \frac{h(a)-h(b)}{h(1-b)-h(1-a)} \leq \eta_1$, contradicting the assumption that $\eta_2 > \eta_1$.

Case (iii-b): $a < b$. Consider the following strategy: (1) The fundamentals $0, 1, 2, 3$ map to k_0 with probabilities $1-t, 0, 1, s$ respectively; the optimal action is a. (2) The fundamentals $0, 1, 2, 3$ map to k_1 with probabilities $t, 1, 0, 1-s$ respectively; the optimal action is b. Then

$$U = ((1-t)x_0 + x_2 + sx_3)h(a) + ((1-t)y_0 + y_2 + sy_3)h(1-a) + ...$$
$$+ (tx_0 + x_1 + (1-s)x_3)h(b) + (ty_0 + y_1 + (1-s)y_3)h(1-b).$$

We need to show that $s = t = 0$ is a suboptimal strategy. By the envelope theorem,

$$\frac{\partial U}{\partial s} = x_3(h(a) - h(b)) + y_3(h(1-a) - h(1-b)),$$
$$\frac{\partial U}{\partial t} = -x_0(h(a) - h(b)) - y_0(h(1-a) - h(1-b)).$$

As in Step 1.1, it suffices to show that one of the two partial derivatives is positive at $s = 0$ or $t = 0$. Suppose both are nonpositive; then $\eta_3 \leq \frac{h(a)-h(b)}{h(1-b)-h(1-a)} \leq \eta_0$, contradicting the assumption that $\eta_3 > \eta_0$.

Case (iii-c): $a = b$. Consider the same strategy as (iii-a). Then $\frac{\partial U}{\partial s}|_{s=0} = \frac{\partial U}{\partial t}|_{t=0} = 0$. The utility is $U|_{s=t=0} = (x_0 + x_1 + x_2 + x_3)h(a) + (y_0 + y_1 + y_2 + y_3)h(1-a)$. By the optimality of a and b, $\frac{y_0+y_2}{x_0+x_2} = \frac{h'(a)}{h'(1-a)} = \frac{h'(b)}{h'(1-b)} = \frac{y_1+y_3}{x_1+x_3}$, and hence $\frac{h'(a)}{h'(1-a)} = \frac{h'(b)}{h'(1-b)} = \frac{y_0+y_1+y_2+y_3}{x_0+x_1+x_2+x_3}$. Therefore, the alternative strategy with the partition $\{\{0, 1, 2, 3\}, \varnothing\}$ yields the same optimal actions and utility as $s = t = 0$. However, it does not use both reports, so by Proposition 2 it is suboptimal. Hence the strategy at $s = t = 0$ is also suboptimal. This completes Step 1.2.

Step 1.3. Given that the solution to $P(N-1,1)$ involves an ordered partition, prove that the solution of $P(N,1)$ involves an ordered partition (for $N \geq 4$).

The proof is by contradiction. Suppose the optimal partition $\{B_0, B_1\}$ for $P(N,1)$ is not ordered. We consider two cases.

Case (i): There exist neighboring fundamentals $i, i+1 \in B_0$ (or B_1; here we pick B_0 without loss of generality). In this case, let v be a fundamental and $(x_v, y_v) = (x_i + x_{i+1}, y_i + y_{i+1})$. Then $\Lambda(i) < \Lambda(v) < \Lambda(i+1)$. Consider the $P(N-1,1)$ problem of partitioning $\{0, 1, ..., i-1, v, i+2, ..., N\}$. Since $\{B_0, B_1\}$ is optimal, by Lemma 2 the solution has to be $\{B_0', B_1'\}$ where $B_0' = B_0 \cup \{v\} \backslash \{i, i+1\}$ and $B_1' = B_1$. However, this is not an ordered partition, a contradiction.

Case (ii): There are no patterns as in case (i), i.e. $B_0 = \{0, 2, 4, ...\}$ and $B_1 = \{1, 3, 5, ...\}$. Let v be a fundamental and $(x_v, y_v) = (x_0 + x_2, y_0 + y_2)$. Then $\Lambda(v) < \Lambda(2)$. Consider the $P(N-1,1)$ problem of partitioning $\{v, 1, 3, 4, ..., N\}$. Since $\{B_0, B_1\}$ is optimal, by Lemma 2 the solution has to be $\{B_0', B_1'\}$ with $B_0' = B_0 \cup \{v\} \backslash \{0, 2\}$ and $B_1' = B_1$. However, $\Lambda(v) < \Lambda(3) < \Lambda(4)$, so this is not an ordered partition, a contradiction.

By Steps 1.1, 1.2, and 1.3, $P(N,1)$ has ordered-partition solutions.

Step 2. Prove that the solution to $P(N,n)$ involves an ordered partition.

The problem is to partition $\{0, 1, ..., N\}$ into $n+1$ sets. Denote the solution by $\{B_0, ..., B_n\}$. For any $0 \leq i, j \leq n$ ($i \neq j$), set $B_i = \{K_1^i, K_2^i, ..., K_{m_i}^i\}$ and $B_j = \{K_1^j, K_2^j, ..., K_{m_j}^j\}$ and consider the problem $P(m_i + m_j, 1)$ with fundamental space $\{K_1^i, K_2^i, ..., K_{m_i}^i, K_1^j, K_2^j, ..., K_{m_j}^j\}$. Since $\{B_0, ..., B_n\}$ is optimal, the solution has to be $\{B_i, B_j\}$ to avoid contradiction. By Step 1, the latter is an ordered partition. Therefore, for any two partition sets in $\{B_0, ..., B_n\}$, there is a cutoff such that the two partition sets lie on different sides of it. This implies that $\{B_0, ..., B_n\}$ is an ordered partition. $\square$

B Theorem 2

Lemma 3 (properties of $\tilde{a}(K)$)**.** *Under Assumptions 3 and 4(i),*

(i) $\tilde{a}(K)$ *is well-defined, strictly increasing, and five times continuously differentiable on* $(\underline{K}, \bar{K})$*,*

(ii) for any n*,* $a_0^* < \tilde{a}(K_1^*) < a_1^* < \tilde{a}(K_2^*) < \ldots < \tilde{a}(K_n^*) < a_n^*$*, and*

(iii) $\kappa^*(n) \subset (\underline{K}, \bar{K}) \subset (\underline{K}^{(1)}, \bar{K}^{(0)})$ *for all* n*.*

Proof of Lemma 3. (i) The function $\tilde{a}(K)$ is implicitly determined by

$$\frac{f_{K|\theta=1}(K)}{f_{K|\theta=0}(K)} = \frac{(1-\pi)u_0 h'(1-a)}{\pi u_1 h'(a)}.$$

Because $\frac{\partial}{\partial a} h'(1-a)/h'(a) \neq 0$ on $(0,1)$, by the implicit function theorem, $\tilde{a}(K)$ is well-defined on $(\underline{K}, \bar{K})$ and inherits the smoothness of LHS minus RHS, so it is five times continuously differentiable. (By an analogous argument, $\tilde{K}(a)$, the inverse of $\tilde{a}(K)$ on $(0,1)$, is also well-defined and five times continuously differentiable.) By Assumption 4(i)(c), the left-hand side is strictly increasing in K. By Assumption 3, the right-hand side is strictly increasing in a. Hence $\tilde{a}(K)$ is strictly increasing.

(ii) Given a partition interval (K_1, K_2), its optimal action a^* satisfies

$$\frac{\Pr(K \in (K_1, K_2)|\theta = 1)}{\Pr(K \in (K_1, K_2)|\theta = 0)} = \frac{(1-\pi)u_0 h'(1-a)}{\pi u_1 h'(a)}.$$

The right-hand side is strictly increasing in a. The left-hand side lies in $\left(\frac{f_{K|\theta=1}(K_1)}{f_{K|\theta=0}(K_1)}, \frac{f_{K|\theta=1}(K_2)}{f_{K|\theta=0}(K_2)}\right)$ by Assumption 4(i)(c). Hence, $\tilde{a}(K_1) < a^* < \tilde{a}(K_2)$. This result applies to all partition sets.

(iii) The optimal action a_0^* for $(-\infty, K_1^*)$ satisfies $a_0^* > 0$, so $\tilde{a}(K_1^*) > 0$, i.e., $K_1^* > \underline{K}$. Similarly, for $(K_n^*, +\infty)$, $a_n^* < 1$, so $\tilde{a}(K_n^*) < 1$, i.e., $K_n^* < \bar{K}$. □

Lemma 4 (cutoffs are dense in the limit)**.** *Under Assumptions 3 and 4(i),*

(i)
$$\lim_{n\to\infty} \max_{K_i^*, K_{i+1}^* \in \kappa^*(n)} |K_{i+1}^* - K_i^*| = 0, \text{ and}$$

(ii)
$$\lim_{n\to\infty} K_1^* = \underline{K}, \quad \lim_{n\to\infty} K_n^* = \bar{K}.$$

Proof of Lemma 4. Let $v(a, K_1, K_2) := \Pr(K \in (K_1, K_2)) E[u(a;\theta)|K \in (K_1, K_2)]$, $v^*(K_1, K_2) :=$

$\max_a v(a, K_1, K_2)$, and denote the optimal action on (K_1, K_2) by $a(K_1, K_2)$.

The proof is by contradiction. For Lemma 4(i), suppose there exists $\delta > 0$ such that for a subsequence $\{n_j\}$,

$$\max_{K_i^*, K_{i+1}^* \in \kappa^*(n_j)} |K_{i+1}^* - K_i^*| > \delta.$$

By a slight abuse of notation, let (K_i^*, K_{i+1}^*) denote the longest interval in the partition given by $\kappa^*(n_j)$. Using the cutoff set $\kappa^*(n_j) \cup \{(K_{i+1}^* + K_i^*)/2\}$ improves $\kappa^*(n_j)$ by at least $w := \min_{K \in (\underline{K}, \bar{K} - \delta)} v^*(K, K + \delta/2) + v^*(K + \delta/2, K + \delta) - v^*(K, K + \delta)$. By Assumptions 3 and 4(i)(c), $a(K, K + \delta/2)$ and $a(K + \delta/2, K + \delta)$ cannot coincide for any K. Hence $w > 0$.

For Lemma 4(ii), I present the proof for $\underline{K}$. Suppose there exists $K' > \underline{K}$ such that for a subsequence $\{n_j\}$, $K_1^* > K'$. Define K^{**} as $(\underline{K} + K')/2$ if $\underline{K} > -\infty$ and $K' - 1$ if $\underline{K} = -\infty$. Using the cutoff set $\kappa^*(n_j) \cup \{K^{**}\}$ improves $\kappa^*(n_j)$ by at least $w' := \inf_{n_j} v^*(\underline{K}, K^{**}) + v^*(K^{**}, K_1^*) - v^*(\underline{K}, K_1^*) \geq (v^*(\underline{K}, K^{**}) - v(a(\underline{K}, K_1^*), \underline{K}, K^{**})) + \inf_{n_j} \{v^*(K^{**}, K_1^*) - v(a(\underline{K}, K_1^*), K^{**}, K_1^*)\} > 0$. Here the inequality $>$ holds because by Assumptions 3 and 4(i)(c), $a(\underline{K}, K^{**})$ and $a(\underline{K}, K_1^*)$ cannot coincide.

Let $\hat{\kappa}_1(n_1), \hat{\kappa}_2(n_2) \subset (\underline{K}, \bar{K})$ be any two cutoff sets satisfying $\hat{\kappa}_1(n_1) \subset \hat{\kappa}_2(n_2)$, with expected utility levels $\hat{U}_1$ and $\hat{U}_2$. Obviously, $\hat{U}_1 \leq \hat{U}_2$, $\hat{U}_2 \leq M := h(1) \max\{u_0, u_1\} < +\infty$, and $\hat{U}_1 \geq m := \max_a \pi u_1 h(a) + (1-\pi) u_0 h(1-a) > -\infty$. Hence $\hat{U}_2 - \hat{U}_1 \leq M - m$. Under n_j, let $\hat{\kappa}_2(n_2)$ be $\kappa^*(n_j)$ and let $\hat{\kappa}_1(n_1)$ be $\{K_i^* \in \kappa^*(n_j) | i \text{ is even}\}$. Then, setting $K_0^* := \underline{K}$ and $K_{n+1}^* := \bar{K}$, we get $\hat{U}_2 - \hat{U}_1 = \sum_{t=1,\ldots,\lfloor (n_j+1)/2 \rfloor} \nu_t$, where $\nu_t = v^*(K_{2t-2}^*, K_{2t-1}^*) + v^*(K_{2t-1}^*, K_{2t}^*) - v^*(K_{2t-2}^*, K_{2t}^*)$ for $t < \lfloor (n_j+1)/2 \rfloor$ and $\nu_t = v^*(K_{2t-2}^*, K_{2t-1}^*) + \ldots + v^*(K_{n_j}^*, \bar{K}) - v^*(K_{2t-2}^*, \bar{K})$ for $t = \lfloor (n_j+1)/2 \rfloor$. Each ν_t is nonnegative, and since $M - m$ is the upper bound, $\min_t \nu_t \leq 2(M-m)/n_j$. As $n_j \to \infty$, $2(M-m)/n_j \to 0$, so $\min_t \nu_t \to 0$. Therefore, given w or w', there exists $\tilde{n}_j$ such that $\min_t \nu_t$ is less than w or w'. Let this minimum be achieved at $\tilde{t}$. Thus, the cutoff set $\kappa^*(\tilde{n}_j) \cup \{(K_i^* + K_{i+1}^*)/2\} \backslash \{K_{2\tilde{t}-1}^*\}$ improves on $\kappa^*(\tilde{n}_j)$ by at least $w - \min_t \nu_t > 0$, a contradiction that proves Lemma 4(i). The cutoff set $\kappa^*(\tilde{n}_j) \cup \{K^{**}\} \backslash \{K^*_{2\tilde{t}-1}\}$ improves on $\kappa^*(\tilde{n}_j)$ by at least $w' - \min_t \nu_t > 0$, a contradiction that proves Lemma 4(ii). $\square$

Proof of Theorem 2. By Lemma 4, it suffices to show that the functions

$$\hat{\beta}_n(a) = \begin{cases} \beta_n(K_i^*) + \frac{\beta_n(K_{i+1}^*)-\beta_n(K_i^*)}{K_{i+1}^*-K_i^*}(K - K_i^*) \text{ if } K \in [K_i^*, K_{i+1}^*), \\ \frac{\beta_n(K_1^*)}{K_1^*-\underline{\underline{K}}}(K - \underline{\underline{K}}) \text{ if } a \in (\underline{\underline{K}}, K_1^*), \quad 0 \text{ if } K \leq \underline{\underline{K}}, \quad 1 \text{ if } a \geq K_n^* \end{cases}$$

converge, where $\underline{\underline{K}} := \max\{\underline{K}, K_1^* - 1\}$. $\hat{\beta}_n(K)$ are absolutely continuous cumulative distribution functions. By the theorem in Scheffé (1947), it suffices to show their densities,

$$b_n(a) = \begin{cases} \frac{\beta_n(K_{i+1}^*)-\beta_n(K_i^*)}{K_{i+1}^*-K_i^*} \text{ if } K \in [K_i^*, K_{i+1}^*), \\ \frac{\beta_n(K_1^*)}{K_1^*-\underline{\underline{K}}} \text{ if } K \in (\underline{\underline{K}}, K_1^*), \quad 0 \text{ otherwise,} \end{cases}$$

converge pointwise to some limiting density almost everywhere. We proceed in two steps.

Step 1. Find the target density. By Lemma 3, we can define $\tilde{K} : (0,1) \to (\underline{K}, \bar{K})$ as the inverse function of $\tilde{a}(K)$ on $(\underline{K}, \bar{K})$. To simplify notation, let $t_1(K) = h(\tilde{a}(K))$, $t_0 = h(1 - \tilde{a}(K))$, $I_i = K_{i+1}^* - K_i^*$, and $J_i = \tilde{K}(a_i^*) - \tilde{K}(a_{i-1}^*)$. Then the first-order conditions for a_i^* and K_i^* are respectively

$$\frac{F_1(K_{i+1}^*) - F_1(K_i^*)}{F_0(K_{i+1}^*) - F_0(K_i^*)} = \frac{F_1'(\tilde{K}(a_i^*))}{F_0'(\tilde{K}(a_i^*))}, \tag{B.1}$$

$$\frac{t_1(\tilde{K}(a_i^*)) - t_1(\tilde{K}(a_{i-1}^*))}{t_0(\tilde{K}(a_i^*)) - t_0(\tilde{K}(a_{i-1}^*))} = \frac{t_1'(K_i^*)}{t_0'(K_i^*)}. \tag{B.2}$$

In (B.1), Taylor-expanding $F_1(K_{i+1}^*)$, $F_1(K_i^*)$, $F_0(K_{i+1}^*)$, and $F_0(K_i^*)$ to fourth order at $\bar{K}_i := (K_{i+1}^* + K_i^*)/2$ and $F_0(K_{i+1}^*)$ and $F_0(K_i^*)$ to first order at $\bar{K}_i$, we get

$$\begin{aligned} \tilde{K}(a_i^*) - \bar{K}_i =& \frac{1}{24} \frac{F_0'''(\bar{K}_i)F_1'(\bar{K}_i) - F_1'''(\bar{K}_i)F_0'(\bar{K}_i)}{F_1'(\bar{K}_i)F_0''(\bar{K}_i) - F_0'(\bar{K}_i)F_1''(\bar{K}_i)} J_i^2 + \frac{R_1 + R_2 + R_3 + R_4 + R_5 + R_6}{F_1'(\bar{K}_i)F_0''(\bar{K}_i) - F_0'(\bar{K}_i)F_1''(\bar{K}_i)} \\ :=& \Gamma(\bar{K}_i)I_i^2 + Rem_i^a, \qquad \text{where} \end{aligned} \tag{B.3}$$

$$
\begin{aligned}
R_1 &= C_1(F_0'(\bar{K}_i)F_1'''(K_d) - F_1'(\bar{K}_i)F_0'''(K_d')) && \times (\tilde{K}(a_i^*) - \bar{K}_i)^2, \\
R_2 &= C_2(F_0'''(\bar{K}_i)F_1''(\bar{K}_i) - F_1'''(\bar{K}_i)F_0''(\bar{K}_i) && \times I_i^2(\tilde{K}(a_i^*) - \bar{K}_i), \\
R_3 &= C_3(F_0'''(\bar{K}_i)F_1'''(K_d) - F_1'''(\bar{K}_i)F_0'''(K_d')) && \times I_i^2(\tilde{K}(a_i^*) - \bar{K}_i)^2, \\
R_4 &= C_4((F_0^{(5)}(K_c'') - F_0^{(5)}(K_c'''))F_1'(\bar{K}_i) - (F_1^{(5)}(K_c) - F_1^{(5)}(K_c'))F_0'(\bar{K}_i)) && \times I_i^4, \\
R_5 &= C_5((F_0^{(5)}(K_c'') - F_0^{(5)}(K_c'''))F_1''(\bar{K}_i) - (F_1^{(5)}(K_c) - F_1^{(5)}(K_c'))F_0''(\bar{K}_i)) && \times I_i^4(\tilde{K}(a_i^*) - \bar{K}_i), \\
R_6 &= C_6((F_0^{(5)}(K_c'') - F_0^{(5)}(K_c'''))F_1'''(K_d) - (F_1^{(5)}(K_c) - F_1^{(5)}(K_c'))F_0'''(K_d')) && \times I_i^4(\tilde{K}(a_i^*) - \bar{K}_i)^2.
\end{aligned}
$$

In (B.2), Taylor-expanding $t_1(\tilde{K}(a_i^*))$, $t_1(\tilde{K}(a_{i-1}^*))$, $t_0(\tilde{K}(a_i^*))$, and $t_0(\tilde{K}(a_{i-1}^*))$ to fourth order at $\bar{\bar{K}}_i := (\tilde{K}(a_i^*) + \tilde{K}(a_{i-1}^*))/2$ and $t_1'(K_i^*)$ and $t_0'(K_i^*)$ to first order at $\bar{\bar{K}}_i$, we get

$$
\begin{aligned}
K_i^* - \bar{\bar{K}}_i =& \frac{1}{24}\frac{t_0'''(\bar{\bar{K}}_i)t_1'(\bar{\bar{K}}_i) - t_1'''(\bar{\bar{K}}_i)t_0'(\bar{\bar{K}}_i)}{t_1'(\bar{\bar{K}}_i)t_0''(\bar{\bar{K}}_i) - t_0'(\bar{\bar{K}}_i)t_1''(\bar{\bar{K}}_i)} I_i^2 + \frac{S_1 + S_2 + S_3 + S_4 + S_5 + S_6}{t_1'(\bar{\bar{K}}_i)t_0''(\bar{\bar{K}}_i) - t_0'(\bar{\bar{K}}_i)t_1''(\bar{\bar{K}}_i)} \\
:=& T(\bar{\bar{K}}_i)J_i^2 + Rem_i^K, \qquad \text{where}
\end{aligned}
\tag{B.4}
$$

$$
\begin{aligned}
S_1 &= C_1(t_0'(\bar{\bar{K}}_i)t_1'''(K_{dd}) - t_1'(\bar{\bar{K}}_i)t_0'''(K_{dd}')) && \times (K_i^* - \bar{\bar{K}}_i)^2, \\
S_2 &= C_2(t_0'''(\bar{\bar{K}}_i)t_1''(\bar{\bar{K}}_i) - t_1'''(\bar{\bar{K}}_i)t_0''(\bar{\bar{K}}_i)) && \times J_i^2(K_i^* - \bar{\bar{K}}_i), \\
S_3 &= C_3(t_0'''(\bar{\bar{K}}_i)t_1'''(K_{dd}) - t_1'''(\bar{\bar{K}}_i)t_0'''(K_{dd}')) && \times J_i^2(K_i^* - \bar{\bar{K}}_i)^2, \\
S_4 &= C_4((t_0^{(5)}(K_{cc}'') - t_0^{(5)}(K_{cc}'''))t_1'(\bar{\bar{K}}_i) - (t_1^{(5)}(K_{cc}) - t_1^{(5)}(K_{cc}'))t_0'(\bar{\bar{K}}_i)) && \times J_i^4, \\
S_5 &= C_5((t_0^{(5)}(K_{cc}'') - t_0^{(5)}(K_{cc}'''))t_1''(\bar{\bar{K}}_i) - (t_1^{(5)}(K_{cc}) - t_1^{(5)}(K_{cc}'))t_0''(\bar{\bar{K}}_i)) && \times J_i^4(K_i^* - \bar{\bar{K}}_i), \\
S_6 &= C_6((t_0^{(5)}(K_{cc}'') - t_0^{(5)}(K_{cc}'''))t_1'''(K_{dd}) - (t_1^{(5)}(K_{cc}) - t_1^{(5)}(K_{cc}'))t_0'''(K_{dd}')) && \times J_i^4(K_i^* - \bar{\bar{K}}_i)^2.
\end{aligned}
$$

Here, K_d and K_d' (between $\tilde{K}(a_i^*)$ and $\bar{K}_i$) are in the remainders of $F_1'(\tilde{K}(a_i^*))$ and $F_0'(\tilde{K}(a_i^*))$; K_c and K_c'' (between K_{i+1}^* and $\bar{K}_i$) are in the remainders of $F_1(K_{i+1}^*)$ and $F_0(K_{i+1}^*)$; and K_c' and K_c''' (between $\bar{K}_i$ and $\tilde{K}(a_i^*)$) are in the remainders of $F_1(K_i^*)$ and $F_0(K_i^*)$. We omit the subscripts i in this notation. Differentiation is possible by the smoothness in Assumption 4. The denominator $F_1'(\bar{K}_i)F_0''(\bar{K}_i) - F_0'(\bar{K}_i)F_1''(\bar{K}_i)$ is positive when $\bar{K}_i \in (\underline{K}, \bar{K})$, since $(F_1'(K)/F_0'(K))' > 0$ by Assumption 4(i).

Also, K_{dd} and K'_{dd} (between K_i^* and $\bar{\bar{K}}_i$) are in the remainders of $t'_1(K_i^*)$ and $t'_0(K_i^*)$; K_{cc} and K''_{cc} (between $\tilde{K}(a_i^*)$ and $\bar{\bar{K}}_i$) are in the remainders of $t_1(\tilde{K}(a_i^*))$ and $t_0(\tilde{K}(a_i^*))$; and K'_{cc} and K'''_{cc} (between $\bar{\bar{K}}_i$ and $\tilde{K}(a_{i-1}^*)$) are in the remainders of $t_1(\tilde{K}(a_{i-1}^*))$ and $t_0(\tilde{K}(a_{i-1}^*))$. Again we omit the subscripts i. Differentiation is possible by the smoothness in Assumption 3 and Lemma 3(i), and $t'_1(\bar{\bar{K}}_i)t''_0(\bar{\bar{K}}_i) - t'_0(\bar{\bar{K}}_i)t''_1(\bar{\bar{K}}_i) > 0$ when $\bar{\bar{K}}_i \in (\underline{K}, \bar{K})$ by Assumption 3. Thus,

$$I_i - I_{i-1} = -2\Gamma(\bar{K}_i)I_i^2 - 2\Gamma(\bar{K}_{i-1})I_{i-1}^2 - 4T(\bar{\bar{K}}_i)J_i^2 - 2Rem_i^a - 2Rem_{i-1}^a - 4Rem_i^K, \quad \text{(B.5)}$$

$$J_i - I_{i-1} = -2\Gamma(\bar{K}_{i-1})I_{i-1}^2 - 2T(\bar{\bar{K}}_i)J_i^2 - 2Rem_{i-1}^a - 2Rem_i^K. \quad \text{(B.6)}$$

To find the limit of $b_n(x)$, we first investigate $b_n(x)/b_n(y)$ on $(\underline{K}, \bar{K})$. Choose any closed interval $[K_L, K_H] \subset (\underline{K}, \bar{K})$ and examine $b_n(x)/b_n(y)$, for any $x, y \in [K_L, K_H]$ $(x < y)$. Define $i_z := \max\{i : K_i^* \leq z\}$. By Lemma 4, $\lim_{n\to\infty} K_1^* = \underline{K}$, so for a large n,

$$b_n(x) = \frac{\beta_n(K_{i_x+1}^*) - \beta_n(K_{i_x}^*)}{K_{i_x+1}^* - K_{i_x}^*} = \frac{1/n}{I_{i_x}}.$$

Thus, for any large n,

$$\frac{b_n(x)}{b_n(y)} = \frac{(1/n)/I_{i_x}}{(1/n)/I_{i_y}} = \frac{I_{i_y}}{I_{i_y-1}} \times ... \times \frac{I_{i_x+1}}{I_{i_x}} = \exp\left(\ln\left(\frac{I_{i_y}}{I_{i_y-1}}\right) + ... + \ln\left(\frac{I_{i_x+1}}{I_{i_x}}\right)\right). \quad \text{(B.7)}$$

The first-order Taylor expansion of $\ln\left(\frac{I_i}{I_{i-1}}\right)$ at 1 gives $\ln\left(\frac{I_i}{I_{i-1}}\right) = \frac{I_i}{I_{i-1}} - 1 - \frac{1}{2t_i}(\frac{I_i}{I_{i-1}} - 1)^2$, where t_i is between $\frac{I_i}{I_{i-1}}$ and 1, so Eq. (B.7) becomes

$$\frac{b_n(x)}{b_n(y)} = \exp\left(\sum_{i=i_x+1}^{i_y} (\frac{I_i}{I_{i-1}} - 1)\right) / \exp\left(\sum_{i=i_x+1}^{i_y} \frac{1}{2t_i}(\frac{I_i}{I_{i-1}} - 1)^2\right). \quad \text{(B.8)}$$

By Eqs. (B.5) and (B.6), we have

$$\frac{I_i}{I_{i-1}} = 1 - 2\Gamma(\bar{K}_i)I_i\frac{I_i}{I_{i-1}} - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 4T(\bar{\bar{K}}_i)J_i\frac{J_i}{I_{i-1}} - \frac{2Rem_i^a + 2Rem_{i-1}^a + 4Rem_i^K}{I_{i-1}}, \quad \text{(B.9)}$$

$$\frac{J_i}{I_{i-1}} = 1 - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 2T(\bar{\bar{K}}_i)J_i\frac{J_i}{I_{i-1}} - \frac{2Rem_{i-1}^a + 2Rem_i^K}{I_{i-1}}. \quad \text{(B.10)}$$

Now let us discuss Eq. (B.8). First, we show three useful results. Results 1 and 3 are with proofs, and Result 2 is a direct corollary of Result 1.

Result 1. We have $\max_i |I_i/I_{i-1} - 1| \to 0$ and $\max_i |J_i/I_{i-1} - 1| \to 0$ as $n \to \infty$.

Proof of Result 1. For brevity, let $X_i = I_i/I_{i-1}$ and $Y_i = J_i/I_{i-1}$. Equations (B.9) and (B.10) can be rearranged into the following system:

$$X_i = 1 - 2\Gamma(\bar{K}_i)I_iX_i - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 4T(\bar{\bar{K}}_i)J_iY_i - 2\frac{Rem_i^a/I_{i-1}}{X_i}X_i - 2Rem_{i-1}^a/I_{i-1} - \frac{4Rem_i^K/I_{i-1}}{Y_i}Y_i,$$

$$Y_i = 1 - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 2T(\bar{\bar{K}}_i)J_iY_i - 2Rem_{i-1}^a/I_{i-1} - \frac{2Rem_i^K/I_{i-1}}{Y_i}Y_i.$$

Solving this system, we obtain

$$X_i = \frac{1 - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 2Rem_{i-1}^a/I_{i-1} - (4T(\bar{\bar{K}}_i)J_i + \frac{4Rem_i^K/I_{i-1}}{Y_i})Y_i}{1 + 2\Gamma(\bar{K}_i)I_i + 2\frac{Rem_i^a/I_{i-1}}{X_i}}, \tag{B.11}$$

$$Y_i = \frac{1 - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 2Rem_{i-1}^a/I_{i-1}}{1 + 2T(\bar{\bar{K}}_i)J_i + \frac{2Rem_i^K/I_{i-1}}{Y_i}}. \tag{B.12}$$

Here, Γ, T and all functions appearing in Rem^a and Rem^K that consist of higher-order derivatives of F and t are continuous on $[K_L, K_H]$ by the continuous differentiability assumption, and are therefore bounded. Let $M > 0$ be a uniform upper bound on their absolute values.

Notice that in Rem_{i-1}^a we have $|\tilde{K}(a_{i-1}^*) - \bar{K}_{i-1}| \leq I_{i-1}$, since $\tilde{K}(a_{i-1}^*), \bar{K}_{i-1} \in [K_{i-1}^*, K_i^*]$ by Lemma 3(ii). Hence, by the triangle inequality, $|Rem_{i-1}^a| \leq M(I_{i-1}^2 + I_{i-1}^3 + I_{i-1}^4 + I_{i-1}^4 + I_{i-1}^5 + I_{i-1}^6)$, so $|Rem_{i-1}^a/I_{i-1}| \leq M(I_{i-1} + I_{i-1}^2 + 2I_{i-1}^3 + I_{i-1}^4 + I_{i-1}^5)$. Since $\max_i I_{i-1} \to 0$ in Lemma 4, $\max_i |Rem_{i-1}^a/I_{i-1}| \to 0$.

Also, in Rem_i^K we have $|K_i^* - \bar{\bar{K}}_i| \leq J_i$, since $\tilde{K}_i^*, \bar{\bar{K}}_i \in [\tilde{K}(a_{i-1}^*), \tilde{K}(a_i^*)]$. Hence $|Rem_i^K| \leq M(J_i^2 + J_i^3 + J_i^4 + J_i^4 + J_i^5 + J_i^6)$, and so $|(Rem_i^K/I_{i-1})/Y_i| = |Rem_i^K/J_i| \leq M(J_i + J_i^2 + 2J_i^3 + J_i^4 + J_i^5)$. Since $\max_i J_i \to 0$ in Lemma 4, $\max_i |(Rem_i^K/I_{i-1})/Y_i| \to 0$.

We conclude that in Eq. (B.12), $|\Gamma|, |T| \leq M$, and all indexed terms uniformly converge to 0.

Thus, $Y_i \to 1$ uniformly on $[K_L, K_H]$.

Furthermore, in Rem_i^a we have $|\tilde{K}(a_i^*) - \bar{K}_i| \leq I_i$, since $\tilde{K}(a_i^*), \bar{K}_i \in [K_i^*, K_{i+1}^*]$. Hence $|Rem_i^a| \leq M(I_i^2 + I_i^3 + I_i^4 + I_i^4 + I_i^5 + I_i^6)$, and so $|(Rem_i^a/I_{i-1})/X_i| = |Rem_i^a/I_i| \leq M(I_i + I_i^2 + 2I_i^3 + I_i^4 + I_i^5)$. Since $\max_i I_i \to 0$ in Lemma 4, $\max_i |(Rem_i^a/I_{i-1})/X_i| \to 0$.

Therefore, in Eq. (B.11), $|\Gamma|, |T| \leq M$, Y_i uniformly converges to 1, and all other indexed terms uniformly converge to 0. Thus, $X_i \to 1$ uniformly on $[K_L, K_H]$. □

Result 2. There exists $\eta \in (0, 1)$ such that $I_i/I_{i-1}, J_i/I_{i-1} \in [1-\eta, 1+\eta]$ for any i and large n.

Definition. Let $(a_m)_{m=1}^M$, $(b_m)_{m=1}^M$, and $(c_m)_{m=1}^M$ be vectors of nonnegative integers and let $z = \min(a_m + b_m + c_m)_{m=1}^M$. Let $o_i(z)$ denote $\sum_{m=1}^M I_{i-1}^{a_m} I_i^{b_m} J_i^{c_m}$.

Result 3. We have $\sum_{i=i_x+1}^{i_y} o_i(z) \to 0$ for $x, y \in [K_L, K_H]$ if $z \geq 2$.

Proof of Result 3. Let $a, b, c \geq 0$ be integers. It suffices to show that $\sum_{i_x+1}^{i_y} I_{i-1}^a I_i^b J_i^c \to 0$ for $x, y \in [K_L, K_H]$ if $a + b + c \geq 2$. Without loss of generality, let $a > 1$; then $0 \leq \sum_{i_x+1}^{i_y} I_{i-1}^a I_i^b J_i^c \leq (\sum_{i_x+1}^{i_y} I_{i-1})(\max_i I_{i-1})^{a-1}(\max_i I_i)^b(\max_i J_i)^c = (y-x)(\max_i I_{i-1})^{a-1}$
$(\max_i I_i)^b(\max_i J_i)^c \to 0$. This proof also works for $b > 1$ or $c > 1$. □

To understand the term $I_i/I_{i-1} - 1$ in Eq. (B.8), substitute Eq. (B.10) into the right-hand side of Eq. (B.9), and subtract 1 from both sides to get

$$\frac{I_i}{I_{i-1}} - 1 = -2\Gamma(\bar{K}_i)I_i - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 4T(\bar{\bar{K}}_i)J_i + Residual_i, \tag{B.13}$$

where $Residual_i =$

$$\begin{aligned}
&-(2Rem_i^a + 2Rem_{i-1}^a + 4Rem_i^K)/I_{i-1} \\
&-2\Gamma(\bar{K}_i)I_i\left(-2\Gamma(\bar{K}_i)I_i\frac{I_i}{I_{i-1}} - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 4T(\bar{\bar{K}}_i)J_i\frac{J_i}{I_{i-1}} - \frac{2Rem_i^a + 2Rem_{i-1}^a + 4Rem_i^K}{I_{i-1}}\right) \\
&-4T(\bar{\bar{K}}_i)J_i\left(-2\Gamma(\bar{K}_{i-1})I_{i-1} - 2T(\bar{\bar{K}}_i)J_i\frac{J_i}{I_{i-1}} - \frac{2Rem_{i-1}^a + 2Rem_i^K}{I_{i-1}}\right).
\end{aligned}$$

In $Residual_i$, we first examine Rem_i^a. Since $|\tilde{K}(a_i^*) - \bar{K}_i| \leq I_i$ (by the proof of Result 1), each of R_2, R_3, R_4, R_5, R_6 divided by its denominator should have absolute value at most a constant times I_i^3, I_i^4, I_i^4, I_i^5, I_i^6, respectively. Since $|R_1|$ is at most a constant times I_i^2, by Eq. (B.4), $R_1 = C_1(F_0'(\bar{K}_i)F_1'''(K_d) - F_1'(\bar{K}_i)F_0'''(K_d'))(\Gamma(\bar{K}_i)I_i^2 + Rem_i^a)^2$. The terms here have absolute values at most a constant times I_i^4, $I_i^2 Rem_i^a$ (which is at most a constant times $I_i^2(I_i^2+I_i^3+I_i^4+I_i^4+I_i^5+I_i^6)$), and $(Rem_i^a)^2$ (which is at most a constant times $I_i^4+...+I_i^{12}$). Thus, $|R_1|$ is at most a constant times I_i^4 plus higher-order terms. Hence, $|Rem_i^a| \leq \frac{M_1}{1+\eta}(I_i^3 + ... + I_i^{12})$, for a constant M_1. By Result 2, for large n, $|\frac{Rem_i^a}{I_{i-1}}| \leq \frac{M_1}{1+\eta}\frac{I_i}{I_{i-1}}(I_i^2+...+I_i^{11}) \leq M_1 o_i(2)$. Similarly, $|\frac{Rem_{i-1}^a}{I_{i-1}}| \leq M_2(I_{i-1}^2+...+I_{i-1}^{11}) \leq M_2 o_i(2)$.

By analogous arguments, $|Rem_i^K| \leq \frac{M_3}{1+\eta}(J_i^3 + ... + J_i^{12})$, where M_3 is a constant. For large n, $|\frac{Rem_i^K}{I_{i-1}}| \leq \frac{M_3}{1+\eta}\frac{J_i}{I_{i-1}}(J_i^2 + ... + J_i^{11}) \leq M_3 o_i(2)$.

Hence, for large n such that $0 < 1 - \eta \leq I_i/I_{i-1}, J_i/I_{i-1} \leq 1 + \eta$ and a constant M_4,

$$|Residual_i| \leq M_4 o_i(2) \tag{B.14}$$

on $[K_L, K_H]$. By Result 3,

$$\sum_{i=i_x+1}^{i_y} Residual_i \to 0, \tag{B.15}$$

since $0 \leq |\sum_{i=i_x+1}^{i_y} Residual_i| \leq \sum_{i=i_x+1}^{i_y} |Residual_i| \leq \sum_{i=i_x+1}^{i_y} M_4 o_i(2) \to 0$. Squaring Eq. (B.13) and using Eq. (B.14), for large n and some constant M_5 we get $|\frac{I_i}{I_{i-1}} - 1|^2 \leq M_5 o_i(2)$. Hence for the denominator of Eq. (B.8) we have

$$\exp\left(\sum_{i=i_x+1}^{i_y} \frac{1}{2t_i}(\frac{I_i}{I_{i-1}} - 1)^2\right) \to 1, \tag{B.16}$$

because, with t_i between 1 and I_i/I_{i-1} and hence between $1 - \eta$ and $1 + \eta$ by Result 2,

$$0 \leq |\sum_{i=i_x+1}^{i_y} \frac{1}{2t_i}(\frac{I_i}{I_{i-1}} - 1)^2| \leq \frac{1}{2(1-\eta)} \sum_{i=i_x+1}^{i_y} (\frac{I_i}{I_{i-1}} - 1)^2 \leq \frac{1}{2(1-\eta)} \sum_{i=i_x+1}^{i_y} M_5 o_i(2) \to 0.$$

Now, substituting Eq. (B.13) in Eq. (B.8), we can rewrite Eq. (B.8) as

$$\begin{aligned}\frac{b_n(x)}{b_n(y)} &= \exp\left(\sum_{i=i_x+1}^{i_y} (-2\Gamma(\bar{K}_i)I_i - 2\Gamma(\bar{K}_{i-1})I_{i-1} - 4T(\bar{\bar{K}}_i)J_i)\right) \times Rest(n; [x, y]) \\ &= \exp\left(RS(-4\Gamma; [x, y]) + RS(-4T; [x, y])\right) \times Rest(n; [x, y]),\end{aligned} \tag{B.17}$$

where $Rest(n;[x,y]) \to 1$ by Eq. (B.15) and Eq. (B.16), and $RS(f,[x,y])$ denotes a Riemann sum of f on $[x,y]$. Because $\int_x^y f'(a)/f(a)da = \ln(|f(a)|)|_x^y$, when $x, y \in [K_L, K_H]$,

$$\frac{b_n(x)}{b_n(y)} \to \exp\left(-4\int_x^y \Gamma(K)dK - 4\int_x^y T(K)dK\right)$$

$$= \left|\frac{F_1'(x)F_0''(x) - F_0'(x)F_1''(x)}{F_1'(y)F_0''(y) - F_0'(y)F_1''(y)}\right|^{\frac{1}{6}} \left|\frac{t_1'(x)t_0''(x) - t_0'(x)t_1''(x)}{t_1'(y)t_0''(y) - t_0'(y)t_1''(y)}\right|^{\frac{1}{6}}$$

$$= \left|\frac{F_1'(x)F_0''(x) - F_0'(x)F_1''(x)}{F_1'(y)F_0''(y) - F_0'(y)F_1''(y)}\right|^{\frac{1}{6}} \left|\frac{h'(\tilde{a}(x))h''(1-\tilde{a}(x)) + h'(1-\tilde{a}(x))h''(\tilde{a}(x))}{h'(\tilde{a}(y))h''(1-\tilde{a}(y)) + h'(1-\tilde{a}(y))h''(\tilde{a}(y))}\right|^{\frac{1}{6}} \left|\frac{\tilde{a}'(x)}{\tilde{a}'(y)}\right|^{\frac{1}{2}}.$$

Under Assumption 3, $|h'(\tilde{a}(x))h''(1-\tilde{a}(x)) + h'(1-\tilde{a}(x))h''(\tilde{a}(x))| = -(h'(\tilde{a}(x))h''(1-\tilde{a}(x)) + h'(1-\tilde{a}(x))h''(\tilde{a}(x)))$. By the monotone likelihood ratio property in Assumption 4(i), $|F_1'(x)F_0''(x) - F_0'(x)F_1''(x)| = F_1''(x)F_0'(x) - F_0''(x)F_1'(x)$. By the integrability assumption in Theorem 2, the denominator as a function of y is integrable on $(\underline{K}, \bar{K})$. Let $m(y) > 0$ denote the denominator scaled by its integral on $(\underline{K}, \bar{K})$, so that $\int_{\underline{K}}^{\bar{K}} m(y)dy = 1$. Then, pointwise for $x, y \in [K_L, K_H]$,

$$\lim_{n\to\infty} \frac{b_n(x)}{b_n(y)} = \frac{m(x)}{m(y)}. \tag{B.18}$$

For any $x, y \in (\underline{K}, \bar{K})$, Eq. (B.18) holds because there always exist K_L, K_H such that $x, y \in [K_L, K_H]$.

Step 2. Prove convergence.

To show that for any $y \in (\underline{K}, \bar{K})$,

$$\lim_{n\to\infty} d_n(y) = m(y), \tag{B.19}$$

notice that by Eq. (B.18),

$$\frac{1}{m(y)} = \int_{\underline{K}}^{\bar{K}} \frac{m(x)}{m(y)}dx = \int_{\underline{K}}^{\bar{K}} \lim_{n\to\infty} \frac{d_n(x)}{d_n(y)}dx$$
$$\leq \liminf_{n\to\infty} \int_{\underline{K}}^{\bar{K}} \frac{d_n(x)}{d_n(y)}dx = \liminf_{n\to\infty} \frac{1}{d_n(y)} = \frac{1}{\limsup_{n\to\infty} d_n(y)},$$

where the inequality is by Fatou's lemma and the subsequent equality comes from the fact that $\int_{\underline{K}}^{\bar{K}} d_n(x)dx =$

1 for any n. Hence,

$$m(y) \geq \limsup_{n\to\infty} d_n(y) > 0. \tag{B.20}$$

Therefore $\limsup_{n\to\infty} d_n(y)$ and $d_n(y)$ are dominated by $m(y)$, which is integrable on $(\underline{K}, \bar{K})$. By the Fatou–Lebesgue theorem,

$$1 = \limsup_{n\to\infty} 1 = \limsup_{n\to\infty} \int_{\underline{K}}^{\bar{K}} d_n(y)dy \leq \int_{\underline{K}}^{\bar{K}} \limsup_{n\to\infty} d_n(y)dy \leq \int_{\underline{K}}^{\bar{K}} m(y)dy = 1,$$

so $\int_{\underline{K}}^{\bar{K}} \limsup_{n\to\infty} d_n(y)dy = \int_{\underline{K}}^{\bar{K}} m(y)dy$. Combined with Eq. (B.20), this implies that, almost everywhere,

$$\limsup_{n\to\infty} d_n(y) = m(y). \tag{B.21}$$

Next, I show by contradiction that $\liminf_{n\to\infty} d_n(y) = m(y)$. Suppose otherwise; then by Eq. (B.21), there exists $y_0 \in (\underline{K}, \bar{K})$ such that $\liminf_{n\to\infty} d_n(y_0) = m(y_0) - \delta \in [0, m(y_0))$. Hence there is a convergent subsequence $\{d_{n_k}(y_0)\}$ such that $\lim_{k\to\infty} d_{n_k}(y_0) = m(y_0) - \delta$. By Eq. (B.18), for any $x \in (\underline{K}, \bar{K})$,

$$\lim_{k\to\infty} d_{n_k}(x) = \frac{m(y_0) - \delta}{m(y_0)} m(x).$$

Then, on the one hand,

$$\int_{\underline{K}}^{\bar{K}} \lim_{k\to\infty} d_{n_k}(x)dx = \frac{m(y_0) - \delta}{m(y_0)},$$

but on the other hand, because $d_{n_k}(x) \leq \limsup_{k\to\infty} d_{n_k}(x) \leq \limsup_{n\to\infty} d_n(x) \leq m(x)$, by the dominated convergence theorem we have

$$\int_{\underline{K}}^{\bar{K}} \lim_{k\to\infty} d_{n_k}(x)dx = \lim_{k\to\infty} \int_{\underline{K}}^{\bar{K}} d_{n_k}(x)dx = \lim_{k\to\infty} 1 = 1,$$

a contradiction. This proves Eq. (B.19). Hence we have $\beta_\infty(K) := \beta_n(K) \to \int_{\underline{K}}^{K} m(y)dy$ as discussed in the beginning of the proof. Therefore, $\beta'_\infty(K) = m(K) \propto \lambda_h(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} \tilde{a}'(K)^{\frac{1}{2}}$. $\square$

C Propositions 4, 5 and 6

Proof of Proposition 4. (i) By the definition of $\tilde{a}(K)$,

$$\frac{h'(1-a)}{h'(a)} = \frac{u_1\pi}{u_0(1-\pi)} \exp\left(\frac{2\mu}{\sigma^2}K\right). \tag{C.1}$$

Hence,

$$\tilde{K}(a) = \frac{\sigma^2}{2\mu}\left(\ln h'(1-a) - \ln h'(a) - \ln \frac{u_1\pi}{u_0(1-\pi)}\right) \tag{C.2}$$

for $a \in (0,1)$, and $\tilde{K}(a) = 2K_{1/2} - \tilde{K}(1-a)$, implying $\tilde{K}(a)$ is symmetric about $(\frac{1}{2}, K_{1/2})$. Thus, $\tilde{a}(K)$ is symmetric about $(K_{1/2}, \frac{1}{2})$ on $(\underline{K}, \bar{K})$. This symmetry obviously holds outside $(\underline{K}, \bar{K})$. The symmetry of $\tilde{a}(K)$ implies the symmetry of $\tilde{a}'(K)$ about $K = K_{1/2}$.

(ii) Evaluating $\lambda_h(K)$ at $K_{1/2}+\delta$ and $K_{1/2}-\delta$ with $\tilde{a}(K_{1/2}+\delta) + \tilde{a}(K_{1/2}-\delta) = 1$, we get $\lambda_h(K_{1/2}+\delta) = \lambda_h(K_{1/2}-\delta)$. □

Proof of Proposition 5. (i) Since the hump-shaped $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is symmetric about $K_{1/2}$, its value is higher if K is closer to $K_{1/2}$. Then, for K such that $KK_{1/2} > 0$, $\beta'_\infty(K) = \exp\left(-\frac{K^2}{6\sigma^2}\right)\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}} \geq \exp\left(-\frac{K^2}{6\sigma^2}\right)\lambda_h(-K)^{\frac{1}{6}}\tilde{a}'(-K)^{\frac{1}{2}} = \beta'_\infty(-K)$. The inequality holds because $K_{1/2}$ and K are closer together than $K_{1/2}$ and $-K$.

(i*) By arguments analogous to those used for (i), the value of the U-shaped $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is lower if K is closer to $K_{1/2}$. Then, for K such that $KK_{1/2} > 0$, we have $\beta'_\infty(K) = \exp\left(-\frac{K^2}{6\sigma^2}\right)\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ $\leq \exp\left(-\frac{K^2}{6\sigma^2}\right)\lambda_h(-K)^{\frac{1}{6}}\tilde{a}'(-K)^{\frac{1}{2}} = \beta'_\infty(-K)$.

(ii) Differentiating both sides of Eq. (C.1) with respect to K, we get

$$\tilde{a}' = \frac{2\mu}{\sigma^2}\frac{h'(\tilde{a})^2}{\lambda_h(K)}\frac{u_1\pi}{u_0(1-\pi)}\exp\left(\frac{2\mu}{\sigma^2}K\right).$$

Differentiating the multiplicative inverse of both sides, we get

$$\tilde{a}' = \frac{2\mu}{\sigma^2}\frac{h'(1-\tilde{a})^2}{\lambda_h(K)} \Bigg/ \left(\frac{u_1\pi}{u_0(1-\pi)}\exp\left(\frac{2\mu}{\sigma^2}K\right)\right).$$

The product of these two equations gives $\tilde{a}'^2$. Its square root is

$$\tilde{a}' = \frac{2\mu}{\sigma^2} \frac{h'(\tilde{a})h'(1-\tilde{a})}{\lambda_h(K)},$$

implying that

$$(\lambda_h^{\frac{1}{6}}(\tilde{a}')^{\frac{1}{2}})^6 = \lambda_h(\tilde{a}')^3 = (\frac{2\mu}{\sigma^2})^3 \frac{h'(\tilde{a})^3 h'(1-\tilde{a})^3}{\lambda_h(K)^2} = (\frac{2\mu}{\sigma^2})^3 \frac{h'(\tilde{a})^3 h'(1-\tilde{a})^3}{(-h'(1-\tilde{a})h''(\tilde{a}) - h''(1-\tilde{a})h'(\tilde{a}))^2}.$$

Hence, $\lambda_h^{\frac{1}{6}}(\tilde{a}')^{\frac{1}{2}}$ is increasing (decreasing) if and only if

$$\frac{d}{d\tilde{a}} \left(\frac{h'(\tilde{a})^3 h'(1-\tilde{a})^3}{(-h'(1-\tilde{a})h''(\tilde{a}) - h''(1-\tilde{a})h'(\tilde{a}))^2} \right) \times \frac{da}{dK} \geq (\leq)\ 0,$$

i.e.,

$$\frac{d}{da} \left(\frac{h'(a)^3 h'(1-a)^3}{(-h'(1-a)h''(\tilde{a}) - h''(1-a)h'(a))^2} \right) \geq (\leq)\ 0. \tag{C.3}$$

For $\lambda_h^{\frac{1}{6}}(\tilde{a}')^{\frac{1}{2}}$ to be hump-shaped in K, this needs to hold if and only if $a < \frac{1}{2}$.

(iii) A sufficient condition for Eq. (C.3) ($\geq$) to hold is that both $\frac{h''(a)}{h'(a)}$ and $\frac{h'''(a)}{h'(a)}$ be decreasing in a. If so, then for $a < \frac{1}{2}$ we have $a < 1-a$, and hence

$$3\left(\frac{h''(a)}{h'(a)} - \frac{h''(1-a)}{h'(1-a)}\right)\lambda_h + 2\left(\frac{h'''(a)}{h'(a)} - \frac{h'''(1-a)}{h'(1-a)}\right) \geq 0.$$

This implies that Eq. (C.3) ($\geq$) holds. □

Proof of Proposition 6. To show Proposition 6, we only need to show that

$$\lim_{\sigma \to 0} (\beta_\infty(r) - \beta_\infty(-r)) = 1$$

for any $r > 0$.

Consider the case when $K_{1/2} > 0$ without loss of generality. Since $K_{1/2} \to 0$ as $\sigma \to 0$, we have $K_{1/2} < \frac{r}{2}$ for small σ values. By Proposition 4, this means $\lambda_H(K)^{\frac{1}{6}} \geq \lambda_H(r)^{\frac{1}{6}}$ for $K \in (0, r)$ and $\lambda_H(K)^{\frac{1}{6}} \leq \lambda_H(r)^{\frac{1}{6}}$ for $K > r$. Also, it is obvious that $\lambda_H(K)^{\frac{1}{6}} \geq \lambda_H(-r)^{\frac{1}{6}}$ for $K \in (-r, 0)$ and

$\lambda_H(K)^{\frac{1}{6}} \leq \lambda_H(r)^{\frac{1}{6}}$ for $K < -r$. Therefore, we have

$$\frac{1-\beta_\infty(r)}{\beta_\infty(r)-\beta_\infty(0)} = \frac{\int_r^{+\infty} \lambda_H(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} dK}{\int_0^r \lambda_H(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} dK} \leq \frac{\lambda_H(r)^{\frac{1}{6}} \int_r^{+\infty} \lambda_F(K)^{\frac{1}{6}} dK}{\lambda_H(r)^{\frac{1}{6}} \int_0^r \lambda_F(K)^{\frac{1}{6}} dK} = \frac{\int_r^{+\infty} \lambda_F(K)^{\frac{1}{6}} dK}{\int_0^r \lambda_F(K)^{\frac{1}{6}} dK},$$

$$\frac{\beta_\infty(-r)-0}{\beta_\infty(0)-\beta_\infty(-r)} = \frac{\int_{-\infty}^{-r} \lambda_H(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} dK}{\int_{-r}^0 \lambda_H(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} dK} \leq \frac{\lambda_H(-r)^{\frac{1}{6}} \int_{-\infty}^{-r} \lambda_F(K)^{\frac{1}{6}} dK}{\lambda_H(-r)^{\frac{1}{6}} \int_{-r}^0 \lambda_F(K)^{\frac{1}{6}} dK} = \frac{\int_{-\infty}^{-r} \lambda_F(K)^{\frac{1}{6}} dK}{\int_{-r}^0 \lambda_F(K)^{\frac{1}{6}} dK}.$$

We know $\lambda_F(K)^{\frac{1}{6}}$ is a $N(0, 3\sigma^2)$ density, so

$$\frac{\int_r^{+\infty} \lambda_F(K)^{\frac{1}{6}} dK}{\int_0^r \lambda_F(K)^{\frac{1}{6}} dK} = \frac{\int_{-\infty}^{-r} \lambda_F(K)^{\frac{1}{6}} dK}{\int_{-r}^0 \lambda_F(K)^{\frac{1}{6}} dK},$$

and hence, as $\sigma \to 0$,

$$0 \leq \frac{1-\beta_\infty(r)+\beta(-r)-0}{\beta_\infty(r)-\beta_\infty(-r)} \leq \frac{\int_r^{+\infty} \lambda_F(K)^{\frac{1}{6}} dK}{\int_0^r \lambda_F(K)^{\frac{1}{6}} dK} + \frac{\int_{-\infty}^{-r} \lambda_F(K)^{\frac{1}{6}} dK}{\int_{-r}^0 \lambda_F(K)^{\frac{1}{6}} dK} \to 0.$$

This means $\beta_\infty(r) - \beta_\infty(-r) \to 1$ and hence proves the result. □

References

Antweiler, Werner, and Murray Z Frank. 2004. "Is all that talk just noise? The information content of internet stock message boards." *The Journal of Finance* 59 (3): 1259–1294.

Baker, Scott R, Nicholas Bloom, and Steven J Davis. 2016. "Measuring economic policy uncertainty." *The Quarterly Journal of Economics* 131 (4): 1593–1636.

Bergemann, Dirk, and Stephen Morris. 2019. "Information design: A unified perspective." *Journal of Economic Literature* 57 (1): 44–95.

Bloedel, Alex, and Ilya Segal. 2021. "Persuading a rationally inattentive agent." *Working Paper.*

Che, Yeon-Koo, and Konrad Mierendorff. 2019. "Optimal dynamic allocation of attention." *American Economic Review* 109 (8): 2993–3029.

Cover, Thomas M, and Joy A Thomas. 2006. *Elements of Information Theory.* John Wiley & Sons.

Dye, Ronald A. 1985. "Disclosure of nonproprietary information." *Journal of Accounting Research,* 123–145.

Gentzkow, Matthew, and Emir Kamenica. 2014. "Costly persuasion." *American Economic Review* 104 (5): 457–62.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as data." *Journal of Economic Literature* 57 (3): 535–74.

Gentzkow, Matthew, and Jesse M Shapiro. 2006. "Media bias and reputation." *Journal of Political Economy* 114 (2): 280–316.

———. 2010. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78 (1): 35–71.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review* 101 (6): 2590–2615.

Loughran, Tim, and Bill McDonald. 2011. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of Finance* 66 (1): 35–65.

Milgrom, Paul, and John Roberts. 1986. "Relying on the information of interested parties." *The RAND Journal of Economics,* 18–32.

Milgrom, Paul R. 1981. "Good news and bad news: Representation theorems and applications." *The Bell Journal of Economics,* 380–391.

Mullainathan, Sendhil, and Andrei Shleifer. 2005. "The market for news." *American Economic Review* 95 (4): 1031–1053.

Perego, Jacopo, and Sevgi Yuksel. 2022. "Media competition and social disagreement." *Econometrica* 90 (1): 223–265.

Scheffé, Henry. 1947. "A useful convergence theorem for probability distributions." *The Annals of Mathematical Statistics* 18 (3): 434–438.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50 (3): 665–690.

Suen, Wing. 2004. "The self-perpetuation of biased beliefs." *The Economic Journal* 114 (495): 377–396.

Tetlock, Paul C. 2007. "Giving content to investor sentiment: The role of media in the stock market." *The Journal of Finance* 62 (3): 1139–1168.

Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. "More than words: Quantifying language to measure firms' fundamentals." *The Journal of Finance* 63 (3): 1437–1467.

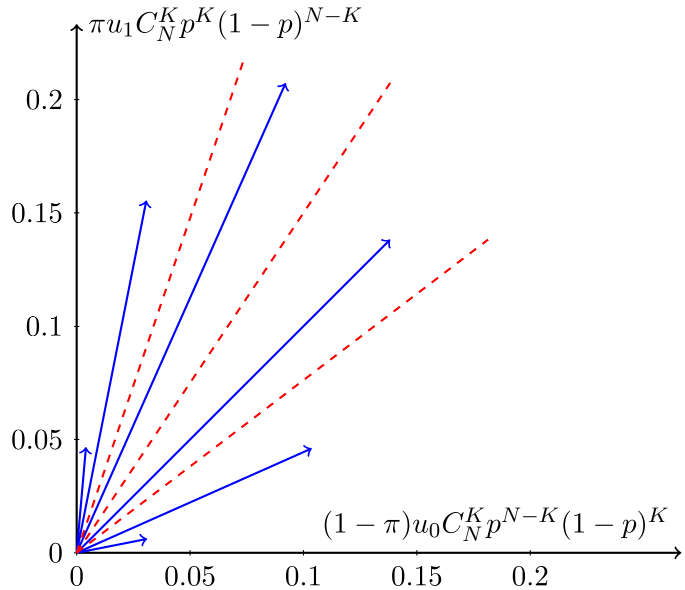
TABLE I:

Equilibrium for Some Common Utility Functions Under Assumption 4(ii)

	$h(a)$	$\beta'_\infty(K) \propto \exp\left(-\frac{K^2}{6\sigma^2}\right) \times ...$
Cosine difference	$\sin\left(\frac{\varpi}{2}a\right)$	$\left((\pi u_1)^2 \exp\left(\frac{2\mu}{\sigma^2}K\right) + ((1-\pi)u_0)^2 \exp\left(-\frac{2\mu}{\sigma^2}K\right)\right)^{-\frac{1}{2}}$
Quadratic	$A(1-a)^2 + B$ $(A < 0)$	$\left(\pi u_1 \exp\left(\frac{\mu}{\sigma^2}K\right) + (1-\pi)u_0 \exp\left(-\frac{\mu}{\sigma^2}K\right)\right)^{-1}$
Log	$\ln(a)$	$\left(\pi u_1 \exp\left(\frac{\mu}{\sigma^2}K\right) + (1-\pi)u_0 \exp\left(-\frac{\mu}{\sigma^2}K\right)\right)^{-\frac{1}{3}}$
Power	$\frac{1}{1-\gamma}a^{1-\gamma}$ $(\gamma > 0,\ \gamma \neq 1)$	$\left((\pi u_1)^{\frac{1}{\gamma}} \exp\left(\frac{\mu}{\gamma\sigma^2}K\right) + ((1-\pi)u_0)^{\frac{1}{\gamma}} \exp\left(-\frac{\mu}{\gamma\sigma^2}K\right)\right)^{\frac{\gamma-2}{3}}$
Exponential	$C_0 - C_1 \exp(-Aa)$ $(A,\ C_1 > 0)$	$\begin{cases} 1 & \text{for } -\frac{\sigma^2}{2\mu}\ln\left(\frac{\pi u_1}{(1-\pi)u_0}\right) - \frac{A\sigma^2}{2\mu} < K \text{ ...} \\ & \text{...} < -\frac{\sigma^2}{2\mu}\ln\left(\frac{\pi u_1}{(1-\pi)u_0}\right) + \frac{A\sigma^2}{2\mu}; \\ 0 & \text{otherwise} \end{cases}$

FIGURE I:

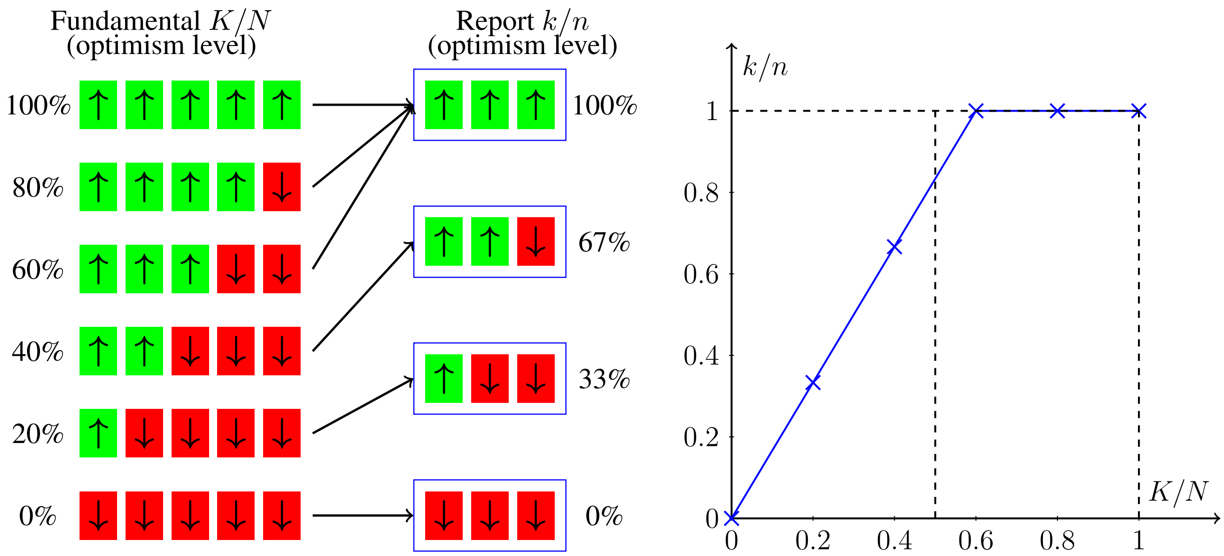
Illustration of Example 1



Notes: This illustration shows the fundamentals (given by arrows, corresponding to $K = 0, ..., 5$ anticlockwise) and the partition (given by dashed lines) in Example 1. Parameter values are as follows: $N = 5$, $n = 3$, $\pi = p = 0.6$, $u_1 = u_0 = 1$.

FIGURE II:

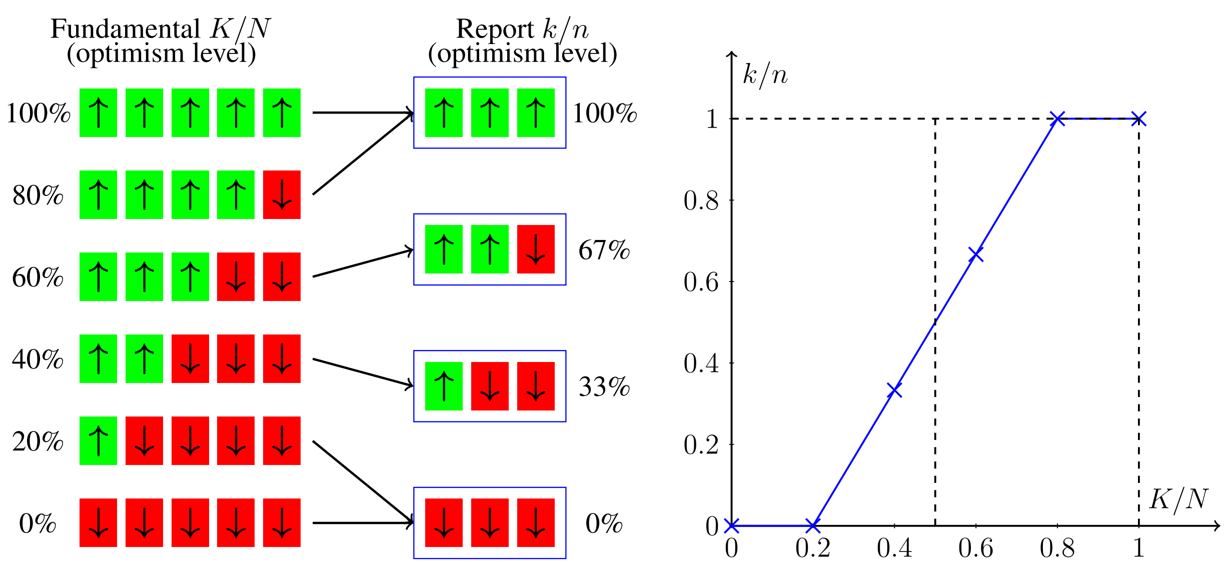
Audience Appeal in Example 2



Notes: This illustration shows the information structure (left) and the interpolated report curve (right), with $N = 5$, $n = 3$, $\pi u_1 = 0.9$, $(1 - \pi)u_0 = 0.1$, and $h(a) = \sin(\varpi a/2)$.

FIGURE III:

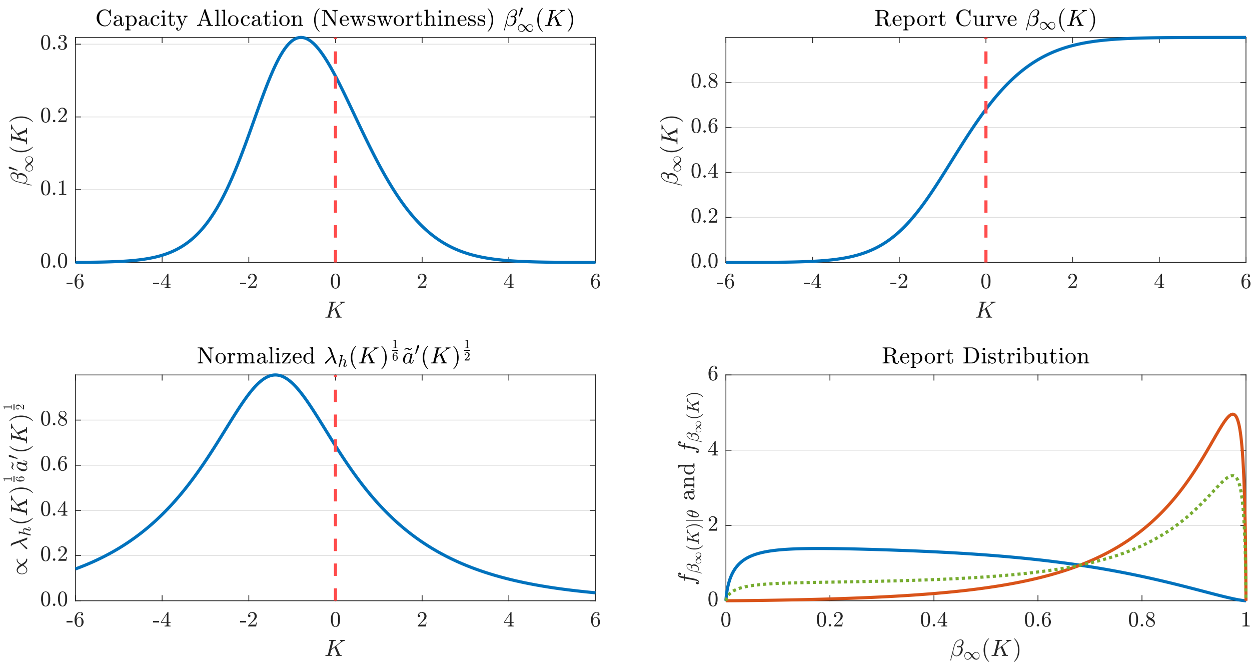
Sensationalism in Example 2



Notes: This illustration shows the information structure (left) and the interpolated report curve (right), with $N = 5$, $n = 3$, $\pi u_1 = 0.6$, $(1 - \pi)u_0 = 0.4$, and $h(a) = \sin(\varpi a/2)$.

FIGURE IV:

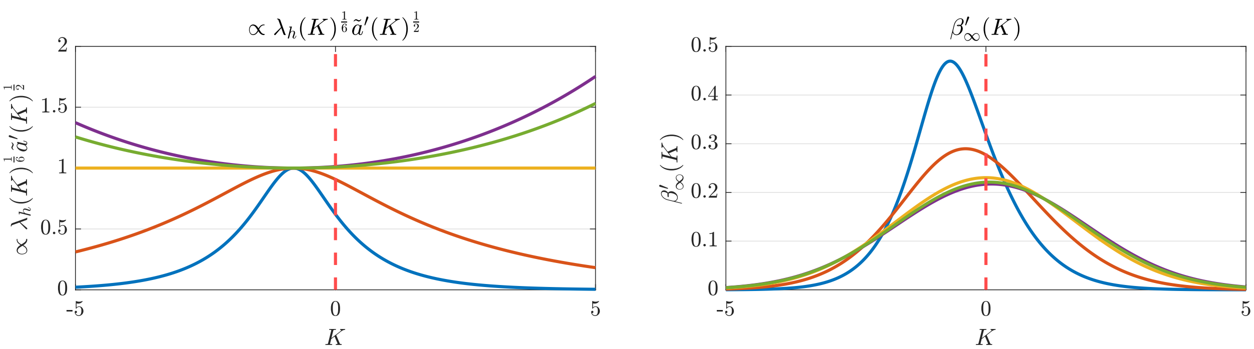
Illustration of Theorem 2



Notes: This is an illustration of Theorem 2 with cosine-difference utility. Parameter values are as follows: $\mu = 0.5$, $\sigma = 1$, $u_1/u_0 = 2$, $\pi = 2/3$. The term $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is normalized to equal 1 at $-\sigma^2 \ln(\pi u_1/(1-\pi)u_0)/2\mu$. In the plot of the report distributions, the blue (positively skewed), orange (negatively skewed), and dotted curves respectively show the conditional density on $\theta = 0$, the conditional density on $\theta = 1$, and the unconditional density.

FIGURE V:

Illustration of Example 3



Notes: This is an illustration of Example 3 with audience-appeal and alarmist biases. Parameter values are as follows: $\mu = 1$, $\sigma = 1$, $\pi u_1/(1-\pi)u_0 = 5$, $\pi = 2/3$. The term $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is normalized to equal 1 at $-\sigma^2 \ln(\pi u_1/(1-\pi)u_0)/2\mu$. The curves in blue, orange, yellow, green, and purple (from low to high for the left panel; from high to low in terms of values at $K = 0$ for the right panel) are for $\gamma = 0.5, 1, 2, 5, 10$, respectively.