Content Bias and Information Compression

Jinge Liu*

April 20, 2024

Abstract

I model communication content in its economic context for micro-founding content bias and sentiment. Content-creating intermediaries must often report selectively to meet content length requirements. In the model, a sender, knowing many signals, must report a certain number of them to a receiver and help him make decisions. I show the content more accurately describes scenarios contradictory to the prior and preferred state and distant from extremes. This generates apparent content biases, including appealing to the audience and sensationalism, that are understood by the decision-maker. Such biases improve welfare. Asymptotically, the model is tractable and smooth, linking content to the reported fundamental information and the economic context. I discuss contextual effects on content. The model is applied to examine media slants and sentiment analysis, and extended to study product ratings.

*City University of Hong Kong. jingeliu@cityu.edu.hk. This paper is adapted from a chapter of my Ph.D. dissertation at Fuqua School of Business, Duke University. I thank my co-chairs S. "Vish" Viswanathan and Simon Gervais for their invaluable advice and support. I thank my committee members Felipe Varas and Arjada Bardhi for their invaluable support and suggestions. I also thank seminar and conference participants at the Asian Meetings of Econometric Society at Shenzhen and Tokyo, City University of Hong Kong, Duke Finance Student Theory Workshop, Hong Kong Junior Finance Workshop, and Stony Brook Game Theory Conference. Comments are welcome.

1 Introduction

Reports often reflect underlying information with bias, influenced by the context in which they are produced. For instance, even trusted news outlets may skew their content based on specific circumstances. This raises an intriguing question: How does content generated by such information intermediaries reflect reality? To interpret content in business and politics and use content data in economic research, we must understand the relationship among content, the underlying information, and the economic context.

In this paper, I propose a theory that micro-founds content by modeling its formation from the perspective of information selection in reporting. It is common for information intermediaries to selectively report and omit information in their content. Such senders often hold abundant information but, in many circumstances, face the problem that they cannot present all information due to some physical constraint, such as content length rules or conventions they must respect. In the news industry, for instance, the number of news stories must fit the newspaper space, broadcast time, or website front page 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."

To model this situation, I consider a sender that holds abundant news that can improve the choices of a decision-maker. News is binary with, say, positive or negative realizations. The sender assembles a certain number of news pieces as the content and presents it, hoping to maximize the decision-maker's utility.

Findings of the model include:

- The model can produce two well-documented report biases: (i) appealing to the audience and (ii) sensationalism. The degree of the biases depends on the economic context. Interestingly, the sender creates bias by selectively reporting and omitting information without telling a lie. Such apparent content biases are understood by the receiver and are a tacit arrangement for the information intermediary to effectively communicate the most useful information. They do not hurt welfare but lead to welfare maximization.
- The model is asymptotically tractable and smooth when the numbers of news signals and included signals in the content become many. Asymptotically, the fundamental information K, representing the realization of all binary news, is

conditionally Gaussian. The content measure is the proportion of positive news in the content. Its slope in K is 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 are the effects of higher-order curvatures of respectively the distribution and the preference, and $\tilde{a}$ is the hypothetical action supposing the receiver knows the perfect information. Contextual parameters affect these terms. This content measure matches how some empirical content research, particularly sentiment research, often quantifies content.

Bias refers to the imbalance of the two realizations included in the content compared to the full underlying information. Although the sender works for the decision-maker, the sender will not simply describe reality in an apparently close way. That is because the sender's job is to recommend actions for different scenarios. And to be trusted, the sender must recommend by presenting evidence. Knowing many signals, the sender can separate many scenarios for herself but not for the decision-maker under a fixed content length. The sender must pool some scenarios to generate the same content. Therefore, the sender must ask: Which scenarios should be pooled? What is the content that labels some pooled scenarios?

The intuition for bias comes from their answers. First, when choosing the best pooling, the sender will minimize information compression loss, a procedure including evaluating how important each scenario is for the decision-maker to separate. The more valuable scenarios are those deemed more probable or prompting higher utility improvements upon knowing. They will be pooled less aggressively. Second, I show the sender can pool optimally while still maintaining truth-telling and logical self-consistency in the content used for labeling these pools. The implication of such a pool-and-label approach is as follows. The sender uses adjacent content labels for adjacent pools of scenarios. While content labels are equidistant from each other, and so are the scenarios, the different levels of pooling aggressiveness create non-linearity in the relationship between a scenario and its content label. This non-linearity manifests as the apparent content bias.

Specifically, two typical report biases, (i) appealing to the audience and (ii) sensationalism, can be produced. In audience appealing, the sender includes more news that favors the version of reality the decision-maker believes or prefers to be true. In sensationalism, the report looks more extreme than the underlying information. This model posits that these biases may be the consequences of information selection. The sender does not highly value elaborating on scenarios close to the extremes or on the high side of prior belief or payoff relevance, thus creating a particular type of the aforementioned non-linearity. Interestingly, such content bias is only apparent. Different from the literature, this paper regards bias as the best report policy under the physical communication constraint and as welfare maximizing.

To structurally analyze content, going asymptotic brings several advantages. First, the asymptotic setup is relevant. The full underlying information consisting of all signals can be summarized by a conditional Gaussian fundamental signal, which is common in economic models with learning. The content measure becomes the proportion of positive signals in the content, also a common empirical content quantification that is especially popular in sentiment analysis. Second, the asymptotic tractability clarifies content's interpretation and facilitates its empirical analysis, opening up the black box of content measures. I show the report policy has a nice form for common utility functions. Asymptotically, appealing to the audience is common and sensationalism is inevitable.

The economic context matters. It refers to the value of contextual economic parameters, including prior belief, payoff relevance, news informativeness, and preference shapes. It affects how the sender views scenarios and hence the report policy and the report distribution. I show how the belief and the payoff relevance affect the degree of appealing and how increasing informativeness leads to more sensationalism. An important model feature is the separation between content's literal meaning and implied true meaning. The former comes from the fact that the report looks like a collection of facts, while the latter is inferred from the report policy. The rational decision-maker inside the context discerns the bias and knows the implied meaning. Researchers, as third parties outside the context analyzing content data and inferring economic information, should consider the context structurally to use data correctly.

The model adapts to many settings. I discuss the source of news narratives and slants, providing a novel perspective on their rationalization. Methodologically, I discuss how the model micro-founds sentiment analysis, a first in the literature.

Furthermore, because this model features the sender compressing complex information to simple reports with monotone labeling and creating non-linearity in the report policy, its structure has broader implications beyond textual content. I discuss how the same model can be applied to micro-found the analysis of a product's star ratings or a student's 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, this paper provides a novel theoretical framework for communication under limited capacity.

This paper is related to the work on media bias as a demand-side theory, as is named in Mullainathan and Shleifer (2005), that attributes media bias to attempts to attract an audience. Seminal work includes Mullainathan and Shleifer (2005) on media competing for a heterogeneous audience and Gentzkow and Shapiro (2006) on the sender's reputation, both requiring some belief or preference heterogeneity. This paper departs by not assuming heterogeneity while still retaining the biases. This paper also connects to work in communication games intersected with limited attention and bias, such as Che and Mierendorff (2019) and Perego and Yuksel (2020).

In methodology, this paper is related to the literature on communication with limited capacity, often branded as limited attention following Simon (1959) if attributing the capacity to the receiver. It intersects the works on Bayesian persuasion (Kamenica and Gentzkow (2011)) with limited attention, including Gentzkow and Kamenica (2014) and Bloedel and Segal (2020) and discusses optimal information compression and attention allocation, while departs by incorporating a practically motivated capacity measure in place of an information-theoretic one (see Chapter 5 of Cover and Thomas (2006), and, e.g., Sims (2003)).

This paper models labels. It is crucial to empirical relevance as the content data we observe are all labels. This feature is novel in communication games with commitment (Kamenica and Gentzkow (2011), Bergemann and Morris (2019)) that only focuses on posteriors and abstracts away from how labels look. It also connects to the literature on the partial disclosure of information or hard evidence (e.g., Milgrom and Roberts (1986), Dye (1985), and Green and Laffont (1986)). In a way, this paper shares similar spirits to signal jamming models such as Stein (1989) in the sense that the sender misreports even though it seems futile in front of rational receivers.

This paper fills the vacuum of economic foundations in the empirical literature of content analysis (see the survey of Gentzkow et al. (2019)). Because of how content is modeled, this paper particularly speaks to founding such analysis that uses textual frequency measures to investigate tendency or sentiment between two competing hypotheses, such as boom and bust, politically left and right, or stable and unstable. Examples include Antweiler and Frank (2004), Tetlock (2007), Tetlock et al. (2008), and Loughran and McDonald (2011) that use frequencies of linguistic tokens and Gentzkow and Shapiro (2010) and Baker et al. (2016) that use frequencies of articles or covered events. This paper provides a method to extract information from such textual measures and parameterize a model for content data, proposing a solution to a standing challenge of studying content data in context.

The rest of the paper proceeds as follows. Section 2 introduces the baseline model and illustrates the biases. Section 3 takes the baseline model to its asymptotic limit, derives the solution and examines it. Section 4 discusses the model's implications. Section 5 extends the model to ratings analysis. Section 6 concludes.

2 The Baseline Model

2.1 Agents

A sender (*she*) reports to a decision-maker (*he*) information about the binary true 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 of the two states, and $h(\cdot)$ is an auxiliary function defined on $[0, 1]$, capturing the closeness between the true state and the bet. I make the following assumption about $h(\cdot)$.

Assumption 1. *The auxiliary function $h(\cdot)$ has the following properties:*
(i) h is twice continuously differentiable, and $h'(a) > 0$ 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 first-order assumption implies that the decision-maker is better off if his bet is closer to the true state. The second-order assumption captures some economic benefits of diversification. It allows interior actions $a \in (0,1)$ to be relevant. Otherwise, if $h''(a) \leq 0$, then only $a = 0$ or 1 are relevant for optimization, and communication is thorough and trivial, with only two actions to recommend but varied reports at the sender's disposal. I further assume $h(\cdot)$ ensures the optimal action is always in the interior to illustrate information compression conveniently. An example of a sufficient condition guaranteeing this is $h'(1) = 0$.

The sender shares the same preference and prior belief. That is, the information intermediary faithfully serves the decision-maker to advance his interests. The rationale can be the sender's pure or strategic loyalty. To see strategic loyalty, consider the sender s can choose her own utility $u_s(a;\theta)$ and prior π_s as her strategic positioning and assume her profit from her information services increases in the value created for the decision-maker. Obviously, an optimal positioning is to align her preference and belief to the decision-maker's so that her communication can prompt the highest expected utility increase for him. Such sender specification allows us to focus on a clear-cut information compression effect without infusing the persuasion effect.

2.2 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. Then the sender 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* refers to an s_i and represents a news story or a piece of evidence in nature. It is material used to form content. An economic scenario is characterized by a vector of signal realizations $\mathbf{s} = (s_1, ..., s_N) \in \{0,1\}^N$, namely the scenario's full underlying information. Assumption 2 describes the signal distribution.

Assumption 2. *$s_1, ..., s_N$ are conditionally independent on θ;* $\Pr(s_i = \theta|\theta) = p > \frac{1}{2}$.

A *reported element* refers to an r_i and represents a covered news story or piece of evidence. I do not require it to be truthful at this moment. *Content* is a collection of reported elements $\mathbf{r} = (r_1, ..., r_n) \in \{0,1\}^n$. I assume their order does not convey information. The positive integer n is the *physical constraint* of communication, representing the content length and is exogenously given. In practice, content length requirements or conventions may be endogenously determined in advance with various considerations, but once pinned down, the communication content should fit that length, no more or less.

The decision-maker's problem in the subgame of $\mathbf{r}$ is to choose the best bet $a^*(\mathbf{r})$ that is essentially recommended by the sender to solve 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 $a^*(\mathbf{r})$, the sender chooses an information structure that solves

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

Dimension Reduction This problem looks high-dimensional but can be simplified. I define the *fundamental* of a scenario as $K := \sum_{i=1}^N s_i$. It is a sufficient statistic for the scenario's full underlying information that summarizes all the signals in Bayesian learning. Its distribution is a binomial mixture, with $K|\theta \sim Bi(N,p)$ if $\theta = 1$ and $Bi(N, 1-p)$ if $\theta = 0$. I define the *report* as $k := \sum_{i=1}^n r_i$. It summarizes content because reported elements are orderless. The fundamental space for K is $\{0, 1, ..., N\}$ and the report space for k is $\{0, 1, ..., n\}$, simplified from $\{0,1\}^N$ and $\{0,1\}^n$.

Importantly, all affine transformations of K and k also serve as equivalent variables for the fundamental and the report, including K/N and k/n, the proportions of ones in the full underlying information and the reported elements. Such affine transformations preserve the equidistant nature of grids in the fundamental and report spaces and hence do not affect the analysis of biases.

The information structure is thus $\{\sigma_{Kk}\}_{K=0,...,N;\ k=0,...,n}$ where $\sigma_{Kk} := \Pr(k|K)$.

The decision-maker chooses $a^*(k)$. The sender finds the optimal $\{\sigma^*_{Kk}\}$ that solves

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

This reformulated problem is central to our discussion of strategic information compression. Intuitively, the sender tries to separate $N+1$ fundamentals in $\{0,...,N\}$ for the decision-maker, but facing the physical constraint n, she only has $n+1$ reports in $\{0,...,n\}$ at her disposal. She must strategically communicate rich fundamentals with limited reports. Her *communication capacity* is thus defined as $n+1$ under the physical constraint n.

If relating this model to information theory, we may call reports codewords and the information structure a codebook. The aim is to code input fundamentals properly. That highlights where this paper departs from the attention literature that bounds mutual information (Sims (2003)) or entropy. The classical information theoretic perspective to coding, which drives the use of mutual information, is to take codewords as meaningless symbols and study expected codeword lengths. In contrast, this paper's approach to coding is rooted in economics with three features: First, the objective is utility. Second, the capacity is practically motivated. Third, it is based on the practical observation that content data, or reports in the model, are not mere meaningless codewords but have their literal substance that needs to be modeled. Although report values can be any $n+1$ distinctive symbols now, later I will give them substance.

2.3 Minimizing Compression Loss

To find the equilibrium, I characterize it to narrow down the search. The following three propositions are necessary conditions that a solution to the reformulated problem must satisfy. They describe how the sender, who compresses the more complex fundamental to the simpler report, minimizes loss. Their proofs are in Appendix A.

First, the sender must use pure strategies. The intuition is straightforward: if the sender is loyal, she has no incentive to infuse unnecessary noise in her reporting by using mixed strategies that confuse the decision-maker.

Proposition 1. *(Pure strategy) Under Assumption 1,* $\sigma^*_{Kk} \in \{0,1\}$ *for any* K, k.

Therefore, the sender's optimal strategy is to partition the set of fundamentals and map fundamentals in each partition set to a report. Two questions follow: How many reports are in equilibrium? Which fundamentals are pooled for the same report?

The next proposition states that all reports must be used in equilibrium.

Proposition 2. *(Surjection) Under Assumption 1 and if* $\Pr(K|\theta = 1)/\Pr(K|\theta = 0)$ *is strictly monotone in* K*, for any given* k*, there exists* K *such that* $\sigma_{Kk} > 0$*.*

Intuitively, every distinctive report is valuable in separating fundamentals and should not be wasted. Because the sender benefits from recommending actions customized to scenarios with the help of reports, she enjoys having more reports at her disposal. An extra report is always a valuable addition to any information structure with pooling because the sender can at least take a partition set of fundamentals, break it into two sets, map one set to the old report and the other set to the extra one, strictly increasing expected utility. That implies the sender must use all $n+1$ reports and create $n+1$ partition sets.

The last proposition discusses which fundamentals are pooled. Intuitively, it is optimal to pool "similar" fundamentals to minimize compression loss. The proper way to measure similarity between fundamentals is by the closeness of their Bayes factors $\Lambda(K)$. For K,

$$\Lambda(K) := \frac{\Pr(K|\theta = 1)}{\Pr(K|\theta = 0)} = \left(\frac{p}{1-p}\right)^{2K-N}$$

under Assumption 2 and is strictly monotone in K. Proposition 3 explores this idea.

Proposition 3. *(Cutoff structure) Under Assumption 1, the optimal partition of the fundamentals for the reformulated problem has the following properties.*

(i) If $\Lambda(K)$ *is different for all* $K = 0, ..., N$*, then let* $(K_j)_{j=0}^{N}$ *be the permutation of* $(0, ..., N)$ *with* $\Lambda(K_0) < ... < \Lambda(K_N)$*. There exist cutoffs* $K_{i_1^*}, ..., K_{i_n^*}$ *so that the optimal partition* $\{B_0, ..., B_n\}$ *is* $B_0 = \{K_0, ..., K_{i_1^*}\}$*,* $B_1 = \{K_{i_1^*+1}, ..., K_{i_2^*}\}$*, ...,* $B_{k'} = \{K_{i_{k'}^*+1}, ..., K_{i_{k'+1}^*}\}$*, ...,* $B_n = \{K_{i_n^*+1}, ..., K_N\}$*.*

(ii) If Assumption 2 holds, then there exist cutoffs $\{K_1^*, ..., K_n^*\}$ *so that the optimal partition* $\{B_0, ..., B_n\}$ *is* $B_0 = \{0, ..., K_1^*\}$*,* $B_1 = \{K_1^* + 1, ..., K_2^*\}$*, ... ,* $B_{k'} = \{K_{k'}^* + 1, ..., K_{k'+1}^*\}$*, ... ,* $B_n = \{K_n^* + 1, ..., N\}$*.*

Proposition 3 states that the optimal solution to the reformulated problem has an *ordered partition* structure in which the sender pools K with similar Bayes factors. Under Assumption 2, the sender should pool adjacent K. The equilibrium is characterized by n cutoffs separating fundamentals into $n+1$ ordered partition sets.

The following example visually illustrates the intuition of Proposition 3.

Example 1. Let $u(a;\theta) = u_\theta \cos\left(\frac{\varpi}{2}|a-\theta|\right)$ by setting $h(a) = \sin\left(\frac{\varpi}{2}a\right)$, where ϖ is the mathematical constant *pi*. Under Assumption 2, the sender's objective is

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

where $\mathbf{v}_K = \left((1-\pi)u_0 C_N^K p^{N-K}(1-p)^K, \pi u_1 C_N^K p^K (1-p)^{N-K}\right) \in \mathbf{R}^2$, the norm is Euclidean, and B_i is a partition set in the partition $\{B_0, ..., B_n\}$.

Geometrically, each fundamental K can be represented by a vector $\mathbf{v}_K$. They are blue arrows in Fig. 1. The sender calculates her utility as follows: First, she partitions all fundamental vectors into $n+1$ sets. Then she calculates each partition set's representative vector, being the vector sum of all vectors in that set. The Euclidean length of each representative vector is the contribution of the fundamentals in that partition set to the sender's expected utility. Finally, she calculates the sum of all representative vectors' lengths which equals the expected utility.

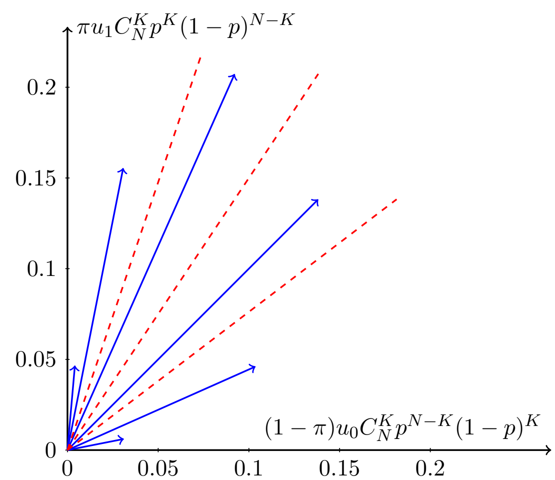

Figure 1: Fundamentals (arrows, $K = 0, ..., 5$ anticlockwise) and Partition (dashed) in Example 1 This figure is plotted under the following parameter values: $N = 5$, $n = 3$, $\pi = p = 0.6$, $u_1 = u_0 = 1$.

To maximize her expected utility, the sender should find the optimal partition that minimizes the loss of representative vector lengths caused by summing vectors in each partition set. Intuitively, the sender should pool vectors with similar angles. For $\mathbf{v_K}$, its angle is $\tilde{a}(K) = \arctan \frac{\pi}{1-\pi}\frac{u_1}{u_0}\Lambda(K)$, strictly monotone in $\Lambda(K)$ and hence K. Therefore, the sender should pool adjacent K. The partition is illustrated in Fig. 1.

Finally, it is worth noting that maximizing expected utility in the reformulated problem only involves solving for the optimal partition and does not involve how reports are attached to the partition sets, as long as a distinctive report is attached to each partition set. In that sense, multiple equilibria must exist because the sender can permute her report attachment arbitrarily. Report assignment is simply an act of labeling from a utility maximization point of view for now.

2.4 Bridging Reports and Substance

Content data are not arbitrary symbols. In practice, readers of newspapers or essays want more than blunt action recommendations. They are still interested in facts and rely on information intermediaries to know facts. They will not subscribe to information services delivering reports that look obviously detached from reality. Consequently, information intermediaries will not publish just any arbitrary symbols as the content, hoping the audience to somehow interpret and accept them. Instead, information intermediaries must frame those reports as some form of presentation of reality, hoping to convince the audience their reporting is fact-based or at least withstand some common-sense scrutiny from the audience.

To connect symbols with substance, I introduce two candidate criteria that reflect common expectations of an information intermediary on characters like trustworthiness, reliability, and professionalism. Such criteria do not constrain optimization.

The first criterion is *(verifiable) honesty*, or the ability to survive fact-checks. It captures the decision-maker's preference for facts. Specifically, to satisfy this criterion, the sender must report verifiably real signal realizations.

Criterion 1. *(Honesty) An information structure is honest if*

$$\{r_1, ..., r_n\} \subset \{s_1, ..., s_N\}.$$

The criterion's condition can be rewritten as the report of any K being no more than K and no smaller than $K - (N - n)$. Importantly, this interpretation of honesty only requires no constituent piece of the content is fabricated rather than a balanced coverage. Suppose $N = 100$ with fifty ones and fifty zeros in the full underlying information. With $n = 50$, an honest report can contain fifty ones and no zeros, being very biased but still surviving fact-checks. Hence, this criterion is not difficult to meet.

The second candidate criterion is *(logical) self-consistency*, or that the sender cannot contradict herself. That means the report should increase in the fundamental.

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

Under self-consistency, if one fundamental favors $\theta = 1$ over $\theta = 0$ at a higher degree than another fundamental, readers will expect the report on the former fundamental to 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 fundamental meaning implied from the information structure. Self-consistency requires the two meanings to move in the same direction. For instance, suppose $\theta = 1$ and 0 represent the good and bad states of the economy. When receiving better economic news from nature, a self-consistent sender should produce a more optimistic-looking report rather than a more pessimistic-looking one. Information intermediaries that are not logically self-consistent may be viewed as strange and untrustworthy.

Theorem 1 is this paper's first main result.

Theorem 1. *Under Assumption 1 and Assumption 2, there exists a solution to the reformulated problem that satisfies both Criteria 1 and 2. Additionally, any solution that satisfies Criterion 2 must also satisfy Criterion 1.*

The proof is straightforward: For any equilibrium, let the optimal ordered partition sets $B_0, ..., B_n$ map to reports $0, ..., n$ in order. This information structure is the only self-consistent one under the optimal partition. Also, it is obviously always honest.

I call this self-consistent optimal information structure the *content-generating* information structure. It must exist, but I have not yet characterized where are the cutoffs for the optimal ordered partition and hence what the content-generating information structure looks like. The analytical characterization is left for the next section, where I take the discrete baseline model to an asymptotic limit. For now, we can solve the discrete problem by exhaustively computing values of every ordered partition to search for the optimal cutoffs.

2.5 Illustrating Two Types of Biases

In the baseline model, *bias* refers to the difference between k/n and K/N. Although discreteness may create a minor wedge, it is not the point and will disappear in the asymptotic model later. Biases arise from the optimal strategic information compression with the content-generating information structure, as Example 2 illustrates.

Example 2. A newspaper reports to an investor who represents the target readers. The market state θ tomorrow is boom (1) or bust (0). The investor has access to two investment opportunities, 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 the position in the former asset when normalizing the short-selling constraint to 0 and budget to 1. The market belief for booms is π. The editor has $N = 5$ news stories about the economy, but the newspaper space only accommodates $n = 3$.

The content-generating information structure depends on the contextual economic variables, including π, u_1, u_0, p, and the utility shape. I focus on the effects of π, u_1, and u_0 and illustrate two cases under different parameter values.

Case 1 in Fig. 2 describes content bias that appeals to the audience, which becomes obvious as πu_1 greatly exceeds $(1 - \pi)u_0$. The left panel shows the content-generating information structure, and the right panel shows the *report curve*, defined as k/n against K/N. The optimal partition sets are $\{0\}$, $\{1\}$, $\{2\}$ and $\{3, 4, 5\}$ and are mapped to reports 0, 1, 2, and 3 in order. The policy can be equivalently described with K/N and k/n, respectively interpreted as the fundamental optimism level and the content's optimistic sentiment. The newspaper disproportionately omits bad stories, creates an upward-tilted k/n, and appears to cater to its audience who believe in booms or can make lucrative investments that will pay off in booms. For instance, when the fundamental is 20% optimistic with one good story and four bad stories, the editor hides two bad stories and makes the content 33% optimistic. When the fundamental is 60% optimistic with three good and two bad, the editor deletes anything bad and

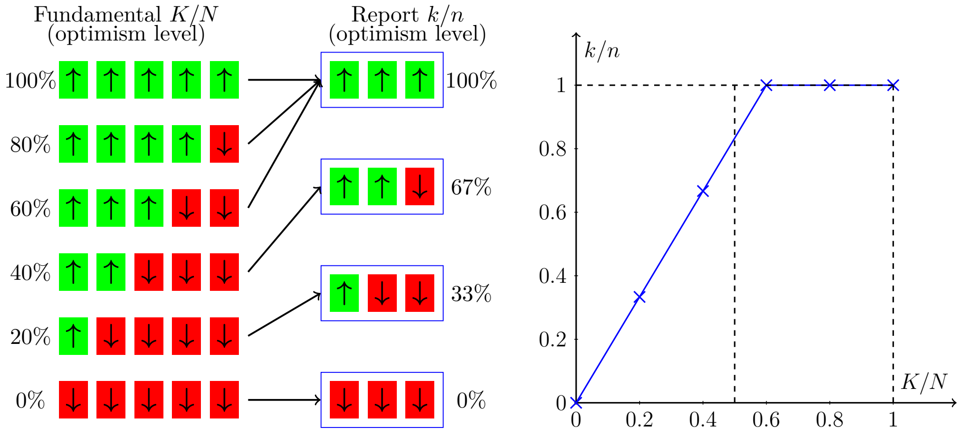

Figure 2: Appealing to the Audience, Information Structure (Left) and Interpolated Report Curve (Right)

This figure is plotted under the following parameter values: $N = 5$, $n = 3$, $\pi u_1 = 0.9$, $(1 - \pi)u_0 = 0.1$, $h(a) = \sin(\varpi a/2)$.

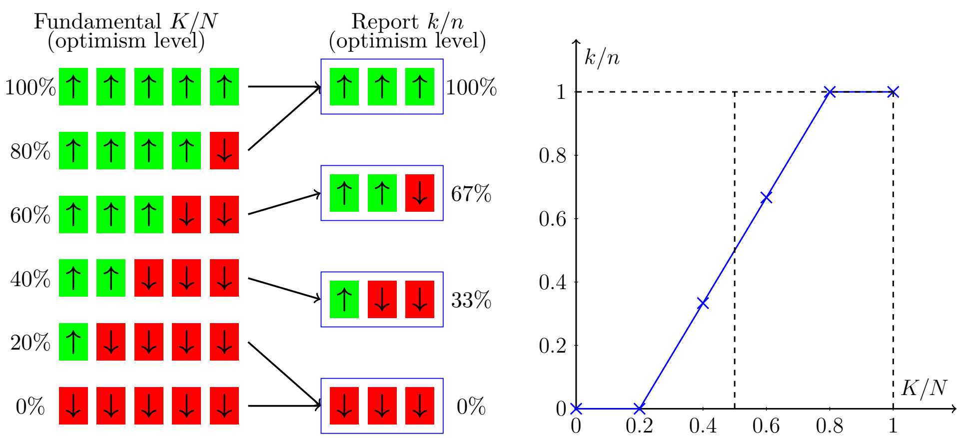

Figure 3: Sensationalism, Information Structure (Left) and Interpolated Report Curve (Right)

This figure is plotted under the following parameter values: $N = 5$, $n = 3$, $\pi u_1 = 0.6$, $(1 - \pi)u_0 = 0.4$, $h(a) = \sin(\varpi a/2)$.

reports a striking 100% optimism level.

Case 2 in Fig. 3 illustrates the bias of sensationalism under comparable levels of πu_1 and $(1 - \pi)u_0$. The content-generating information structure involves optimal partition sets $\{0, 1\}$, $\{2\}$, $\{3\}$ and $\{4, 5\}$ which are respectively mapped to reports 0, 1, 2, and 3. Apparently, the newspaper reports exaggeratedly in both optimistic and pessimistic directions by including disproportionately more stories that are aligned with the fundamental's overall direction to make the content look extreme. For example, when the fundamental is 80% (20%) optimistic, it reports 100% (0%). When the fundamental is 60% (40%) optimistic, it reports 67% (33%).

The first step to interpreting these biases is to interpret the partition. An information structure derives value from its ability to customize action recommendations to scenarios. Intuitively, reports are content space allocations between the two news types. The editor would like to ideally customize a different look for the newspaper to each scenario. Bounded by the space, however, the sender cannot be specific to every scenario, so she must only pick some scenarios to customize action recommendations more carefully by separating them more aggressively. Depending on the economic context, the sender assesses each scenario's potential to contribute to the expected utility once she customizes to it more carefully. Her assessment is comprehensive and covers each scenario's probability and action implications.

Specifically, in Case 1, with a high πu_1, the decision-maker strongly tends to bet close to 1 without the report. He does not place much value on new information that confirms $\theta = 1$ because even if he receives a report that accurately identifies a confirmatory scenario, he will not do something very different. What he truly finds valuable is accurate information about contradictory scenarios, which will help the decision-maker to rethink his action. Hence, the sender follows Fig. 2 and more aggressively pools information that favors $\theta = 1$.

In Case 2, absent strong audience-appealing effects, what becomes obvious is that the decision-maker is not very interested in having tail scenarios differentiated. Two reasons explain this. First, tail scenarios have slim probabilities and thus do not contribute much to expected utility. Second, even if the sender separate tail scenarios, the decision-maker will still choose some extreme actions similar to when those scenarios are not separated. Therefore, the sender follows Fig. 3 and more aggressively pools scenarios near extremes.

The second step is to translate the partition to the report. Two observations are helpful. First, the highest and lowest fundamentals are respectively mapped to the highest and lowest reports without bias. Crucially, unbiased reporting on end scenarios provides two anchors for analyzing bias in middle scenarios. Second, reports for those more newsworthy fundamentals, i.e., the moderate or contradictory ones are more sensitive to fundamentals than reports for less newsworthy scenarios. The same incremental increase in K/N may cause a smaller increase in k/n if such K/N is insignificant or a larger increase if otherwise. Such difference in sensitivities stems from pooling adjacent fundamentals at different intensities on an equidistant fundamental grid before mapping these pools to an also equidistant report grid monotonically.

We now analyze biases with these two observations. For some newsworthy fundamentals in a region, their range in terms of K/N is smaller than their report range in terms of k/n, due to their higher level of report sensitivity. Between the end scenarios with no bias, if such a fundamental region is contradictory, its report region may expand to a territory that looks not so contradictory or even confirmatory. If such a fundamental region is moderate, its report region may expand to a relatively more extreme territory. Furthermore, under a fixed 100% range of k/n, the report regions for less valuable fundamentals are squeezed to the side: Confirmatory fundamentals are forced to associate with even more confirmatory reports and near-extreme fundamentals with more extreme reports.

Two remarks are in order. (1) The report curve that depicts the relationship between reports and fundamentals is the quantitative representation of the equilibrium. (2) Interestingly, the report curve is also the curve of the cumulative counting frequencies of optimal cutoffs. It is naturally so because the report for K/N is its associated k, or equivalently k/n, which can also be interpreted as the percentage of cutoffs below K/N. This identity is intuitive: If the report curve has a high slope somewhere, then these fundamentals are getting separate reports and are the more valuable ones. Naturally, the sender inserts denser cutoffs around those fundamentals, leading to a higher cutoff concentration, or equivalently a higher slope in the cumulative cutoff distribution. This observation is useful in asymptotics.

2.6 Welfare

Contrary to the conventional wisdom that biases hurt welfare, this model shows biases may reflect the sender's efforts to maximize welfare. The proper welfare measure for both the decision-maker and the society is the sender's ex-ante maximized utility. Traditionally, the reasoning of why biases hurt welfare often rests on the presumption that biases result from agency problems. This paper, however, points out that biases may be independent of agency and simply the most efficient communication arrangement under a physical communication capacity. Consequently, biases do good for the welfare of the principal and society.

The result also implies that other communication policies are suboptimal for the decision-maker and social efficiency, notably including two policies that seem to be the widely accepted ethical standard: (1) The sender produces report content that best resembles the full underlying information. (2) The sender fully randomizes her reporting without any deliberate selection, with the ex-ante expected utility being the same as choosing n reported elements from n signals since the sender can simply report the first n of the N signals.

Parameters N and n affect welfare. The maximized utility strictly increases in N because the sender recommends better actions with richer fundamental information under the same communication capacity. The maximized utility also strictly increases 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 disposes of a report. Intuitively, welfare loss comes solely from compression and a bigger n implies a smaller loss.

3 The Asymptotic Model

The baseline model extends to an asymptotic model if N and n go to infinity. The extension has benefits: First, a large N better captures the rich and often overly abundant information in the real world, and a large n better describes the nuanced content used in practice. Second, asymptotics produces tractability and smoothness that facilitate a clear interpretation and convenient empirical analysis of the content.

3.1 Model Setup and Solution

The agents remain the same with utility that depends on $h(\cdot)$.

Assumption 3. *$h(\cdot)$ is fifth continuously differentiable; $h'(a) > 0$, $h''(a) < 0$ on $(0, 1)$.*

I compare Assumption 1 with Assumption 3. First, Assumption 3(1) requires more smoothness than Assumption 1(1) since asymptotics will involve higher order derivatives. Second, Assumption 1(2) is removed for asymptotics. Assumption 1(2) was included simply for convenience: Suppose otherwise some fundamentals trigger the same action of 1 (or 0) if separated, then the sender will pool them because no information will be lost. One can view such a pool as one fundamental, thus effectively reducing the total number of fundamentals. If that new total number is greater than n, Theorem 1 still holds since such a pool has the most extreme Bayes factor. If it is no greater than n, there is no information compression. I use Assumption 1(2) to avoid discussing those cases unnecessarily and losing the focus. In the asymptotic case, however, the discussion of some fundamentals generating the same extreme action is straightforward, so I forgo Assumption 1(2).

Now, let the binary signals s_i take values from $\{-\sigma/\sqrt{N}, \sigma/\sqrt{N}\}$ instead of $\{0, 1\}$, where $\sigma > 0$ is a parameter. Fixing N, I am simply reframing the signals using an affine transformation. The new information environment is equivalent to the previous one. The sum $K = \sum_{i=1}^{N} s_i$, still defined as the fundamental, is the sufficient statistic of the full underlying information. Noticeably, K is no longer a nonnegative integer, but is a real number that can be negative.

Let $\mu > 0$ be a parameter and set

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

where $\mu_1 = \mu$ and $\mu_0 = -\mu$. This is to relate p with N. Because the program is still the baseline model for a fixed pair of N and n, Proposition 1, Proposition 2 and Proposition 3 and Theorem 1 apply and the content-generating information structure features n optimal cutoffs for K.

Among possible paths of (N, n) reaching infinities, I consider when N greatly exceeds n by letting N go to infinity first and n next. It is a two-step process: In the first step, N goes to infinity for a fixed n. The sender produces a simple report on the rich and complex fundamental information. Since the content-generating information structure under (N, n) features n cutoffs for K, the information structure under N being taken to infinity will feature n cutoffs on the real line which becomes the domain of K. In the second step, n also goes to infinity as the sender adds sophistication in reporting. The content-generating information structure is characterized by not n cutoffs, but infinite cutoffs standardized to a unit measure that are continuously distributed on the real line, forming a cutoff density.

The first step starts with $N \to \infty$. By standard arguments of in-fill asymptotics,

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

where $\Rightarrow$ stands for convergence in law. The limiting fundamental K follows a Gaussian mixture distribution and is supported on $\mathbf{R}$. Assumption 2's analogue is thus the following Assumption 4(ii) (and (ii) implies (i)).

Assumption 4. *(i) $F_{K|\theta=1}$ and $F_{K|\theta=0}$ satisfy the following properties:*
(a) They are absolutely continuous and fifth continuously differentiable;
(b) Density $f_{K|\theta=1}(x) > 0$ iff. $x \in (\underline{K}^{(1)}, \bar{K}^{(1)})$, and $f_{K|\theta=0}(x) > 0$ iff. $x \in (\underline{K}^{(0)}, \bar{K}^{(0)})$, 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 $(1-\pi)u_0 h'(1-a)/\pi u_1 h'(a)$ 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) $K|\theta \sim N(\mu_\theta, \sigma^2)$.

The identity between the report curve and the cumulative counting frequencies of cutoffs still holds. Let $\kappa^*(n) = \{K_1^*, ..., K_n^*\}$ be the set of cutoffs such that the ordered partition $\{B_0, B_1, B_2, ..., B_n\}$ is $\{(-\infty, K_1^*), [K_1^*, K_2^*), [K_2^*, K_3^*), ..., (K_n^*, +\infty)\}$. Define

$$\beta_n(K) = \frac{1}{n} \sum_{K' \in \kappa^*(n)} \mathbf{1}_{K' \leq K} \tag{3}$$

as both the cumulative frequencies and the report curve. It fully characterizes the equilibrium in which $\kappa^*(n)$ maximizes the utility

$$\sum_{i=0}^{n} \left\{ \pi u_1 \Pr\left(K \in B_i | \theta = 1\right) h(a_i^*) - (1-\pi) u_0 \Pr\left(K \in B_i | \theta = 0\right) h(1 - a_i^*) \right\}. \tag{4}$$

For any i, a_i^* in Eq. (4) is the recommended action for B_i and hence 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 problem involves choosing cutoffs for the Gaussian-mixture fundamentals, thus avoiding the complexity brought by the discreteness of the fundamental space. As an extension of the baseline model, the equilibrium under $N \to \infty$ can be viewed as satisfying Proposition 1, Proposition 2, Proposition 3 as well as Theorem 1. The content-generating information structure is verifiably honest and logically self-consistent.

In the second step, the sender lets $n \to \infty$. The sender seeks to find $\beta_\infty(K) := \lim_{n\to\infty} \beta_n(K)$ that is both the limiting cumulative frequencies of cutoffs curve and the report curve. The limiting equilibrium is viewed as satisfying Proposition 1, Proposition 2, Proposition 3 and Theorem 1, and meets both criteria.

Obviously, $\beta_\infty(K)$ is a qualified cumulative distribution function by construction. It is increasing, right-continuous, and defined on $\mathbf{R}$ with 0 and 1 as limits at infinities. Hence, $\beta_\infty(K)$ is a distribution that induces a canonical probability space $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathbf{P}')$, with $\mathbf{P}'(S)$ of any $S \in \mathcal{B}(\mathbf{R})$ being the Lebesgue measure of $\beta_\infty(S)$. This probabilistic perspective to the report curve has the following two implications.

First, the communication capacity is a unit measure. In the baseline model, the communication capacity is the number of partition sets $n+1$ or, one can say, the n cutoffs that split those partition sets. As $n \to \infty$, the number of cutoffs goes to a probability of one, which becomes the proper asymptotic capacity measure.

Second, the sender may alternatively solve for the optimal $\beta'_\infty(K)$, the derivative of $\beta_\infty(K)$. Should $\beta_\infty(K)$ be absolutely continuous, $\beta'_\infty(K)$ is a density of cutoffs. Both $\beta_\infty(K)$ and $\beta'_\infty(K)$ equivalently characterize the equilibrium. I call $\beta'_\infty(K)$ the *newsworthiness curve* because it characterizes the fundamentals' worth for coverage: A bigger $\beta'_\infty(K)$ around some K means the sender chooses to insert more cutoffs around that K to separate it, which reflects the sender's assessment that such a scenario deserves more elaborate coverage. It echoes the discussion in the baseline model on how valuable each scenario is for differentiating, but in the asymptotic setup such a discussion can be conducted accurately with $\beta'_\infty(K)$.

To show this paper's connection to the limited attention research, I also name $\beta'_\infty(K)$ the *capacity allocation curve*, analogous to the attention allocation curve in the literature. The sender distributes the scarce resource of cutoffs up to a unit measure, which comes from the physical constraint binding any of the sender, the decision-maker, or the channel. In fact, if one interprets the constraint as facing the decision-maker like the attention literature, $\beta'_\infty(K)$ can be properly called the attention allocation curve. The concept of communication capacity or attention in this paper has a strong economic motivation, complementing the existing information theoretic literature.

What is the equilibrium $\beta'_\infty(K)$? To answer this question, I introduce the perfect information optimal action $\tilde{a}(K)$, the hypothetical best action assuming the decision-maker 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 $R(K) \leq h'(1)/h'(0)$, then $\tilde{a}(K) = 0$; if $R(K) \geq h'(0)/h'(1)$, then $\tilde{a}(K) = 1$. Under Assumption 4(i)(d), the range of $\tilde{a}(K)$ covers $(0,1)$. If fundamentals that induce extreme actions of 1 or 0 exist under a specific economic context, they should be pooled without loss and generate an extreme report of 1 or 0 due to their extreme Bayes factors. It is only necessary to pin down $\beta_\infty(K)$ or $\beta'_\infty(K)$ for those less radical fundamentals that induce interior $\tilde{a}(K)$. I denote the interval on which $\tilde{a}(K) \in (0,1)$ as $(\underline{K}, \bar{K})$.

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

Theorem 2. *(Asymptotic capacity allocation) Under Assumption 3:*

(1) Suppose Assumption 4(i) holds. Then on $(\underline{K}, \bar{K})$,

$$\beta'_\infty(K) \propto \lambda_h(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} \tilde{a}'(K)^{\frac{1}{2}} \tag{5}$$

if it is integrable, where

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

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

(2) Suppose Assumption 4(ii) holds. Then

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

Theorem 2 decomposes newsworthiness into three components: $\lambda_h(K)$ about the curvature of the utility function, $\lambda_F(K)$ about the curvature of the fundamental distribution, and $\tilde{a}'(K)$, the sensitivity of the hypothetical perfect information best action. Theorem 2 can be alternatively written in the log form as the log density of cutoffs being linear in the logs of $\lambda_h(K)$, $\lambda_F(K)$, and $\tilde{a}'(K)$, with loadings of 1/6, 1/6, and 1/2. The report curve is an antiderivative of $\beta'_\infty(K)$ scaled as a cumulative distribution function. The proof of Theorem 2 is in Appendix B.

Another way to present Eq. (5) 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. (5) is

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

With Theorem 2, $\beta_\infty'(K)$ under common utility functions can be conveniently calculated. Table 1 lists some examples. Some report curves have nice forms: With exponential utility, for instance, the report curve is a cumulative distribution function for a truncated $N(0, 3\sigma^2)$ distribution with asymmetric tail cutoffs. Also, $\lambda_h(K)^{\frac{1}{6}} \tilde{a}'(K)^{\frac{1}{2}}$ is proportional to a logistic density for the cosine difference preference, a hyperbolic secant density for the quadratic preference, and a hyperbolic secant density raised to certain powers for log and power preferences. In Fig. 4, I present the capacity allocation and report curves using the cosine difference preference as an example.

Table 1: Equilibrium With 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	$C_0 - C_1\exp(-Aa)$ $(A,\ C_1 > 0)$	$\begin{cases} 1 & \text{for} \quad -\frac{\sigma^2}{2\mu}\ln\left(\frac{\pi u_1}{(1-\pi)u_0}\right) - \frac{A\sigma^2}{2\mu} < K \ ... \\ & \quad ... < -\frac{\sigma^2}{2\mu}\ln\left(\frac{\pi u_1}{(1-\pi)u_0}\right) + \frac{A\sigma^2}{2\mu}; \\ 0 & \text{otherwise} \end{cases}$

Theorem 2 is the key asymptotic result. Importantly, it is not about revelation, although it characterizes a relationship with no compression loss. In fact, with continuous fundamental and report values, any 1-to-1 correspondence can be a loss-free report curve. The real important question answered in Theorem 2 is which 1-to-1 correspondence we will end up with in the limit of either the baseline problem or the asymptotic problem in Eq. (4) that are economically meaningful. The analytical report policy in Theorem 2 serves as a tractable asymptotic approximation to the baseline or Eq. (4)'s equilibrium information structures.

Posterior. Let β_∞^{-1} be the inverse of β_∞ on $(\underline{K}, \bar{K})$ and Φ be the standard Gaussian cumulative distribution function. Under Assumption 4(ii), the decision-maker's

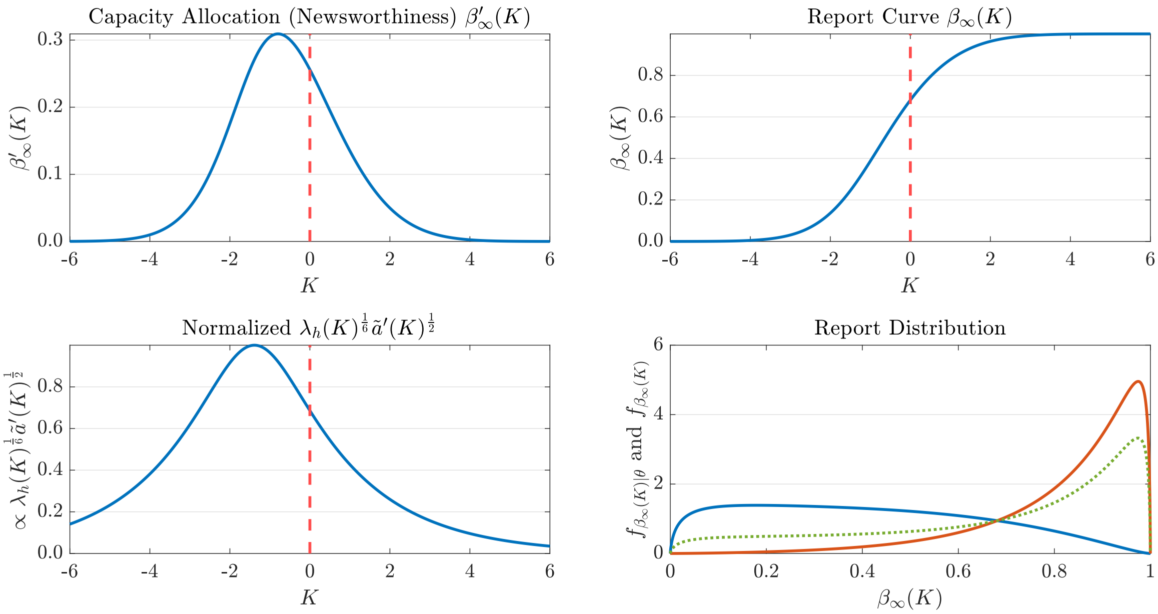

Figure 4: Illustration of Theorem 2 with Cosine Difference Preference
This figure is plotted under the following parameter values: $\mu = 0.5, \sigma = 1, u_1/u_0 = 2, \pi = 2/3$. The value of normalized $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ at $-\sigma^2 \ln(\pi u_1/(1-\pi)u_0)/2\mu$ is set to 1. The report distributions in blue, red, and dotted lines are respectively the conditional density on $\theta = 0$, on $\theta = 1$, and the unconditional density.

posterior belief on seeing a report $\rho \in (0, 1)$ is the same as knowing $K = \beta_\infty^{-1}(\rho)$, 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

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

3.2 Determinants of Capacity Allocation

The three components in Eq. (5) originate from the problem's higher-order curvatures and fall into two groups. The first group is $\lambda_F(K)^{\frac{1}{6}}$, which solely depends on the fundamental's conditional distributions and intuitively describes the direct effect of likelihood on newsworthiness. It comes from the optimality condition for a_i^*

$$\frac{F_{K|\theta=1}(K_{i+1}^*) - F_{K|\theta=1}(K_i^*)}{F_{K|\theta=0}(K_{i+1}^*) - F_{K|\theta=0}(K_i^*)} = \frac{(1-\pi)u_0 h'(1-a_i^*)}{\pi u_1 h'(a_i^*)} \equiv \frac{f_{K|\theta=1}(\tilde{K}(a_i^*))}{f_{K|\theta=0}(\tilde{K}(a_i^*))}. \tag{6}$$

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 set of fundamentals inducing action a. The second equality in Eq. (6) follows from the definition of $\tilde{a}(K)$. This condition implies the relative location of $\tilde{K}(a_i^*)$ within $[K_i^*, K_{i+1}^*]$ is

$$\frac{\tilde{K}(a_i^*) - K_i^*}{K_{i+1}^* - K_i^*} \approx \frac{1}{2} + \frac{1}{24}\left(\left.\frac{d\ln\lambda_F(K)}{dK}\right|_{K=\frac{K_i^*+K_{i+1}^*}{2}}\right)(K_{i+1}^* - K_i^*).$$

The second group is $\lambda_H(K)^{\frac{1}{6}} = \lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$. Intuitively, $\tilde{a}'(K)$ describes how sensitive the optimal action is to fundamentals. And $\lambda_h(K)$, which only depends on K through $\tilde{a}(K)$, describes how such action sensitivity translates to utility sensitivity. 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 K_i^* within $[\tilde{K}(a_{i-1}^*), \tilde{K}(a_i^*)]$, is

$$\frac{K_i^* - \tilde{K}(a_{i-1}^*)}{\tilde{K}(a_i^*) - \tilde{K}(a_{i-1}^*)} \approx \frac{1}{2} + \frac{1}{24}\left(\left.\frac{d\ln\lambda_H(K)}{dK}\right|_{K=\frac{\tilde{K}(a_{i-1}^*)+\tilde{K}(a_i^*)}{2}}\right)(\tilde{K}(a_i^*) - \tilde{K}(a_{i-1}^*)).$$

Intuitively, with totally one unit of cutoffs, studying their allocation among localities equals studying the relative cutoff allocation between any two different localities. Locally, for a cutoff K_i^* to draw nearby cutoffs, it may attract $\tilde{K}(a_i^*)$ more than its neighbor K_{i+1}^* does, and meanwhile such $\tilde{K}(a_i^*)$ may attract K_{i+1}^* more than the more distant $\tilde{K}(a_{i+1}^*)$ does. 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 in a neighborhood. Then for two localities K_1 and K_2 apart, their relative cutoff allocation depends on properly aggregating such local relative concentration characterizations between K_1 and K_2. That eventually leads to Theorem 2.

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

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

That is, $\lambda_F(K)$ increases in the two conditional likelihoods and the elasticity of the likelihood ratio. And $\lambda_H(K)$ increases in the two conditional value sensitivities and the elasticity of the value sensitivity ratio, each of which can be further decomposed into a utility sensitivity term and an action sensitivity term $\tilde{a}'(K)$ with the chain rule.

These terms fully describes what makes a fundamental newsworthy.

3.3 Characterizing Bias in the Report Curve

Bias is a report curve property because the curve captures the report policy and measures how the report distribution distorts the fundamental distribution.

Appealing to the Audience. To identify the audience appealing bias in the report policy, I compare the location of $\beta'_\infty(K)$ with the midpoint of K. The proper midpoint for K is zero since it is the case under any finite N. Compared with zero, the more $\beta'_\infty(K)$ leans against the direction dictated by the sign of $\frac{\pi u_1}{(1-\pi)u_0} - 1$, the more the report policy should be viewed as biased toward that direction.

Let $K_{1/2}$ denote $-\frac{\sigma^2}{2\mu} \ln \frac{\pi u_1}{(1-\pi)u_0}$ which solves $\tilde{a}(K) = 1/2$ under Assumption 4(ii). Obviously, its sign is opposite to $\frac{\pi u_1}{(1-\pi)u_0} - 1$. Here is a bias characterization.

Definition 1. *$\beta_\infty(K)$ 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. Suppose a strongly appealing $\beta_\infty(K)$ is the distribution of some random variable, then its mean (if well-defined) and median, both being commonly used location notions of distributions, and the average of the upper and lower $\alpha\%$ quantiles for any α are all on the same side of zero as $K_{1/2}$.

What is the source of audience appealing? 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, such bias is from $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$. Proposition 4 describes some shape properties of $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$.

Proposition 4. *Under Assumption 3 and Assumption 4(ii),*
(1) $\tilde{a}(K)$ is symmetric about $(K_{1/2}, \frac{1}{2})$, and $\tilde{a}'(K)$ is symmetric about $K = K_{1/2}$;
(2) $\lambda_h(K)$ is symmetric about $K = K_{1/2}$.

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

Proposition 5. *Under Assumption 3 and Assumption 4(ii),*
(1) $\beta_\infty(K)$ is strongly appealing if $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is hump-shaped, i.e., it increases on

$(\underline{K}, K_{1/2})$ *and decreases on* $(K_{1/2}, \bar{K})$*;*

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

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

$$\frac{d}{da}\left(\frac{h'(a)^3 h'(1-a)^3}{(-h'(1-a)h''(\tilde{a}) - h''(1-a)h'(a))^2}\right) \geq (\leq)\ 0; \tag{7}$$

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

Proposition 5 only involves h and is simple to verify. Many common utilities, including cosine difference, quadratic, log, power (with $\gamma \leq 2$), and exponential preferences fit Proposition 5(1), and all above except for log and power fit Proposition 5(3).

Intuitively, Proposition 5 states it is sufficient for the report curve to be strongly appealing if customizing action recommendations for extreme scenarios does not matter too much to overall utility. Notice that $\tilde{a}'(K)$ is hump-shaped with both tails tending to zero, implying the action recommendation is not sensitive to fundamentals near the action boundaries. 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. That is, $\lambda_h(K)$ can be hump-shaped, meaning customizing extreme action recommendations is inconsequential. Or $\lambda_h(K)$ can explode but only at a rate controllable by $\tilde{a}'(K)$ such that $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is still hump-shaped, implying action customization for extreme scenarios is highly valued with the excessive marginal utility, but still has limited newsworthiness implications for the fundamentals since actions are not sensitive enough to fundamentals. Example 3 illustrates this intuition.

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

Fig. 5 shows $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ and $\beta'_\infty(K)$ for various γ. The tails of $\lambda_h(K)^{\frac{1}{6}}$ behave properly for $\gamma \leq 1$ but explode for $\gamma > 1$. When $\gamma < 2$, $\tilde{a}'(K)^{\frac{1}{2}}$ prevails despite the exploding tails of $\lambda_h(K)^{\frac{1}{6}}$, and the combined $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ fits Proposition 5(1), 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(1*). Take $\beta'_\infty(K)$ as $\lambda_F(K)$ scaled by $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$. Because $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ is symmetric around $K_{1/2}$ and is bigger for fundamentals more distant from $K_{1/2}$, it scales up $\lambda_F(K)$ opposite to $K_{1/2}$ more and eventually causes the capacity allocation to focus on confirmative K values, leading to an interesting alarmist bias for high γ values.

Ultimately, the bias type depends on Eq. (7), which can be rewritten as

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

The first two terms are the elasticities for conditional marginal utilities, a certain type of marginal utility level sensitivity. And the third term is twice the elasticity of the elasticity of the conditional marginal utility ratio, a certain type of marginal utility ratio sensitivity. For a small a, under most preferences, the marginal utility level "sensitivity" dominates the marginal utility ratio "sensitivity", leading to the strongly appealing property, whereas under CRRA utility with $\gamma > 2$, it is the opposite which causes the strongly alarmist property.

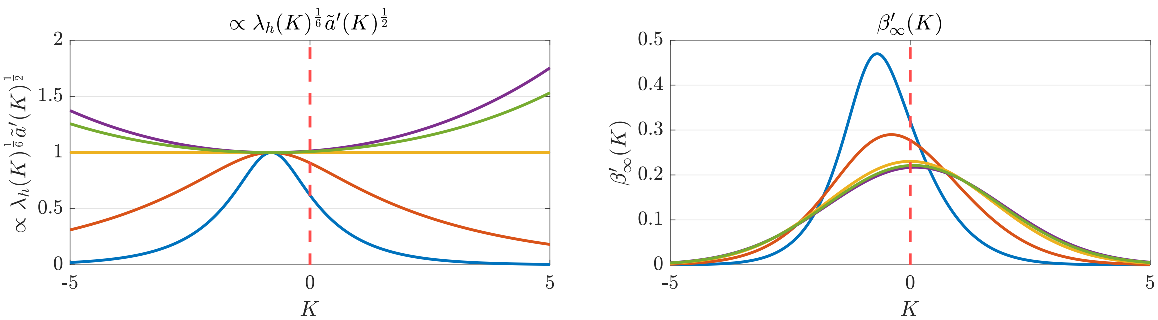

Figure 5: Illustration of Example 3 with Appealing versus Alarmist Biases
This figure is plotted under the following parameter values: $\mu = 1, \sigma = 1, \pi u_1/(1-\pi)u_0 = 5, \pi = 2/3$. The value of normalized $\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ at $-\sigma^2\ln(\pi u_1/(1-\pi)u_0)/2\mu$ is set to 1. The curves in blue, red, yellow, green, and purple are with $\gamma = 0.5, 1, 2, 5, 10$.

Sensationalism. Sensationalism can be described by the sender dumping two vast regions of tail scenarios into two small bins of extreme reports, or rigorously by the small slopes of the report curve when K tends to infinities. Under Assumption 4(ii) and with integrability in Theorem 2, sensationalism is predicted to be inevitable.

Essentially, the source of sensationalism is the structural assumption that N greatly exceeds n on their path to infinities. This assumption also determines the proper scaling of fundamentals and reports for the asymptotics. It captures the reality in many situations that the underlying information is complex in nature and rich in details, far beyond the power of an information intermediary to present no matter how much nuanced she can make the content.

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

3.4 The Contextual Effects

Report Distribution. The observable data include the report and may not include latent K. Under Assumption 4(ii), the conditional distribution of the report ρ is

$$\begin{cases} \Pr(\rho = 0|\theta) = \Phi\left(\frac{\underline{K}-\mu_\theta}{\sigma}\right) & \text{if } \rho = 0; \\ \rho|\theta \sim \Phi\left(\frac{\beta_\infty^{-1}(\rho)-\mu_\theta}{\sigma}\right) & \text{if } \rho \in (0,1); \\ \Pr(\rho = 1|\theta) = 1 - \Phi\left(\frac{\bar{K}-\mu_\theta}{\sigma}\right) & \text{if } \rho = 1, \end{cases} \tag{8}$$

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

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

Definition 2 compares the degree of audience appealing.

Definition 2. *$\beta_\infty^{(1)}(K)$ "leans more positively" than $\beta_\infty^{(2)}(K)$ 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 affects the report distribution through the report curve. Suppose $u_1^{(1)}/u_0^{(1)} > u_1^{(2)}/u_0^{(2)}$ 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" $\beta_\infty^{(1)\prime}(K)/\beta_\infty^{(2)\prime}(K)$ strictly decreases, implying $\beta_\infty^{(2)}(K)$ has FOSD over $\beta_\infty^{(1)}(K)$ and thus $\beta_\infty^{(1)}(K)$ leans more positively. Trivially for exponential utility, $\beta_\infty^{(1)}(K)$ leans more positively. For power utility with $\gamma > 2$, $\beta_\infty^{(1)\prime}(K)/\beta_\infty^{(2)\prime}(K)$ strictly increases and hence $\beta_\infty^{(1)}(K)$ leans less positively. In usual situations, higher payoff relevance of a state leads to the report policy leaning toward that state more.

The conditional and unconditional report distribution for $u_1^{(1)}/u_0^{(1)}$, when $\beta_\infty^{(1)}(K)$ leans more positively, has FOSD over the corresponding report distribution for $u_1^{(2)}/u_0^{(2)}$. That is because we have $\beta_\infty^{(1)}(K) \geq \beta_\infty^{(2)}(K)$ 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. (8).

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 is the mixing probability. The second effect implies π and u_1/u_0 are separately identifiable in report data.

If for $\pi^{(1)} > \pi^{(2)}$, $\beta_\infty^{(1)}(K)$ leans more positively than $\beta_\infty^{(2)}(K)$, then the conditional report distributions under $\pi^{(1)}$ has FOSD over under $\pi^{(2)}$, following an analogous reasoning for u_1/u_0. The unconditional report distribution under $\pi^{(1)}$ also has FOSD over under $\pi^{(2)}$, because besides higher conditional distributions, more weight is given to the higher $\theta = 1$ conditional distribution in the mixture.

Informativeness μ/σ. The identifiable parameter in report data for the information structure is μ/σ. Parameters μ and σ are not separately identifiable. To see this, suppose $\mu^{(2)} = C\mu^{(1)}$ and $\sigma^{(2)} = C\sigma^{(1)}$. Then K in setting (1) is the identical fundamental to CK in (2). That is, (i) $\frac{2\mu}{\sigma^2}K$ and hence the reports at K and CK are the same, and (ii) the conditional cumulative probabilities are the same. Therefore, the conditional and unconditional report distributions in (1) and (2) are the same. Essentially, relabeling K with affine transformations does not change the problem; hence, what matters is the problem for the standardized fundamental K/μ. For the same reason, if we want to back out the fundamental $K = \beta_\infty^{-1}(\rho)$ for some observed $\rho \in (0, 1)$, we must first assume either μ or σ without loss of generality.

Informativeness contributes to sensationalism. For instance, let h satisfy Proposition 5(1) and let σ tends to zero, fixing μ. Both hump-shaped $\lambda_H(K)^{\frac{1}{6}}$ and $\tilde{a}'(K)^{\frac{1}{2}}$ horizontally stretch toward zero and obviously, the limiting $\beta_\infty(K)$ is 0 for negative K and 1 for nonnegative K. This extremely sensationalist policy captures the vital importance of separating middle fundamentals when the information is very clear as the action and utility sensitivities are wildly high. Consequently, the conditional report distributions will tend to a mass of one at 0 and 1 respectively, and the unconditional distribution is highly dispersed as their mix.

Preference. Proposition 5 and Example 3 outline the effect of preference curvatures: Under different utilities, the shape of $\lambda_H(K)$ may change, leading to different bias implications for other contextual parameters.

4 Implications: Intuition and Methodology

This model brings new insights in two ways. First, it proposes a novel perspective to understanding content bias, an important topic in business, politics, and everyday life, with a new and common information selection mechanism.

Second, it micro-founds content analysis. In the literature, content analysis often follows a reduced-form or a data-mining approach. It is difficult to see beyond the literal meaning of content data and understand them in the original environment in which they were produced. It is also challenging to interpret the methodology and the results rigorously. This paper fills in by proposing a tractable content model.

As a structural content model, it has three major advantages. First, the model is not a black box, but is founded on the common editorial practice. Second, it treats the literal meaning and the underlying true meaning separately while linking them with the context. This allows us, the third-party observers, to not take content literally but put ourselves in the shoes of the people in that economic environment at the time, thinking how they see the content. Third, contextual effects are modeled clearly.

This model is particularly fitting to study sentiment, as is discussed subsequently.

The Scope of Applicability This model applies to diverse forms of content with the following characteristics. (1) The content facilitates decision-making by presenting information and evaluating two competing hypotheses; (2) there is reason to believe the content length is not very flexible due to requirements or conventions, necessitating information selection; and (3) the primary focus of the analysis is not placed on the heterogeneity within the target audience or agency issues.

Media content is an obvious example. Other examples may include briefings for a busy decision-maker, consultancy reports, filings publishing information, and essays with selectively presented evidence for argumentation.

4.1 Media Bias

Media outlets often form narratives that skew reality. Demand-side theories (e.g., Suen (2004), Mullainathan and Shleifer (2005), Gentzkow and Shapiro (2006)) attribute this to the incentive to satisfy the audience, while they all require some form of belief or preference discrepancy among relevant parties.

This model proposes a novel demand-side framework based on information selection and independent of the discrepancy perspective. Media narratives, which are formed of the often biased editorial selection of covered events, emerge because the media outlets know their target audience's preferences and beliefs and are simply trying to convey the most useful information under the newspaper pages, broadcast time, or website front page size. While the media firms appear to feed their audience with more information believed or liked and sensationalize any content, such slants, biases, and cherrypicking are only manifestations of the most efficient communication protocol that both the media and the audience tacitly agree on.

For instance, θ may represent conflicting alternative facts claimed by the left and right-wing politicians. Households must figure out the fact to choose the right saving plans or policies to support, often with the help of news outlets that target an audience with a political tendency. If a household has confidence in the rightist view or more stakes on the situation that the rightist version of reality is true, then it optimally matches a right-wing media that positions its preference and belief to such audience's. The media will efficiently communicate by apparently filling the newspaper or air time with more evidence supporting the rightist view and to make any narrative more extreme. Same with the leftist.

Should we worry about slants? The model says no, but surely it assumes full rationality that may be too strong. Also, it is not the only bias channel and does not consider sociological and cultural aspects. Nonetheless, it provides a perspective why slants may be reasonable.

4.2 Empirical Implications

Data and Model Preparations In practice, the reports are usually alternative data and need to be formatted into the quantified form in the model. This process is tokenization, which refers to dividing the full content into basic tokens, such as events, pieces of evidence, or sentences and words that correspond to the reported elements in the model. Each token should support one competing hypothesis over the other. In doing so, we obtain the content measure k/n, i.e., ρ.

One feature of such tokenization is the irrelevance of the token order. The model assumes the location of covered events or pieces of evidence carries no information, as long as they are placed in the same newspaper cover page or main body of an essay. This assumption becomes strong if we perform linguistic analysis within a text and take sentences or words as signals. Even so, this assumption is not stronger than what is often used in the popular bag-of-words approach to vectorize a text. Compared to the relevant empirical literature, the proposed method is an improvement.

Results like Theorem 2 and Eq. (8) can prepare the structural model if we set up a problem. Example 4 is for demonstration.

Example 4. Consider an investor with CRRA utility $u(w) = \frac{1}{1-\gamma} w^{1-\gamma}$ on his portfolio worth w allocating one dollar between two categories of assets: The A category with gross return $R_{A,1}$ upon booms and $R_{A,0}$ upon busts, and the B category with $R_{B,1}$ upon booms and $R_{B,0}$ upon busts. To avoid a dominant asset, let $R_{A,1} > R_{B,1}$ and $R_{A,0} < R_{B,0}$. These returns are calibratable parameters from data, for instance, with mean returns for universes of assets that pay off more in booms or busts. The investor shares the market belief for booms π. A financial newspaper targets such investors. To match setups, set

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

where

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

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

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

the truncated distribution of the power case in Table 1 on $(\underline{K}, \bar{K})$. That gives $\beta_\infty(K)$, $\beta_\infty^{-1}(K)$, and the structural model for the report distribution.

Identification with Reports With report data only, the identifiable parameters include π, u_1/u_0, μ/σ, and the identifiable parameter that governs the shape of the utility given the utility form. Assuming μ or σ, we can extract the standardized fundamental for a given report value. If given the joint data of the report and some proxies of other parameters or of the fundamental information, then we may have theoretical access to not just the unconditional marginal report distribution but the joint distribution. In that case, results like Theorem 2 and Eq. (8) help build a parametric model accounting for changes in the context.

Sentiment Analysis and Model (Mis)specification This model is particularly relevant for sentiment analysis. Conventional sentiment research proceeds in two steps. In the first step, it constructs a textual frequency measure such as the ratio of negative words in all words, or of positive stories in all stories. That measure is often named sentiment, pessimism, tones, attitude, or tendencies, and believed to proxy variables like beliefs, preferences, or the fundamentals. In the second step, it uses the measure in regressions or data mining. Its issue is the lack of clarity in what information is embedded in such a measure and how, which compromises the proper interpretation and use of sentiment. This paper speaks to this issue with its report definition k/n matching the sentiment construct.

The content model provides an empirical model specification. It predicts the sentiment measure is a nonlinear mix of fundamentals, preference, and beliefs, following Theorem 2 or its discrete version. It is proposed that an intermediate step be inserted to the two-step process: That is, to use the model to distill the specific information needed for analysis from this mix. For instance, if a researcher wants to study textual-based K, then after assuming μ without loss of generality, she should use the report data possibly joint with other data and estimate a full model, which predicts K for subsequent analysis or possibly includes subsequent analysis in the model as well.

Missing the intermediate step predicts misspecification. For instance, if a researcher confuses k/n with the fundamental and correlates it with an economic variable, obtaining statistical significance, it is uncertain whether the significance comes from the fundamental or from the fact that k/n exaggerates the fundamental, making it easier to obtain significance or possibly amplifying the effect. Or from the fact that k/n includes bias towards market belief or preference, which is what actually correlates with the economic variable. The model can further evaluate the severity of potential misspecification errors quantitatively.

Textual frequencies evaluating two competing hypothesis are constructs similar to the sentiment and can also be studied with the same approach.

5 Beyond Content: An Analysis of Ratings Data

The model can further study bias in other data forms, such as a product's star ratings.

Consider the problem of a consumer (*she*) who rates her customer experience of a product on a scale of five stars ($k = 0, ..., 4$) for a later shopper (*he*). The product's type θ is good (1) or bad (0). The customer experience K is a random variable with its conditional distributions on θ satisfying the monotone likelihood ratio property. The later shopper browses the rating and decides on a, which stands for the probability or the amount of purchase. Suppose everyone aligns in preferences and beliefs.

While the star rating is not a compilation of reported elements, the problem shares a similar information compression structure and can be solved by a slight twist of the model. The customer experience K is much more elaborate than the report k and should be represented by more than five equidistant integers or even continuous real numbers. Like in the baseline model, I start from simplicity and let K be integers.

Importantly, the proofs of Proposition 1, Proposition 2 and Proposition 3 do not depend on the specific distributions of s_i or even K, but only rests on the strictly monotone likelihood ratio property for $K|\theta$. The information structure as the rating strategy is a pure strategy, a surjection, and characterized by cutoffs. To find the optimal rating-generating information structure, we can trim the equilibria with self-consistency, which is natural as customers give higher ratings for a better experience in practice. Next, let K become dense and satisfy (i) or (ii) of Assumption 4. The report curve is obtained by numerically solving Eq. (4), or approximated by Theorem 2.

This paper explains why the ratings we observe in practice often look skewed. A product rated with many five and four stars but relatively fewer low stars may reflect that customers expect its quality to be good or benefit more from purchasing a good product than missing a bad one. In that case, they are keen on separating bad experience scenarios with various low ratings while squeezing relatively fine experiences in four and five stars. A dispersed ratings distribution may reflect high informativeness of experience on quality.

The Model's Essence The extension reveals the model's essential mathematical structure. The model has two equidistant grids: a fundamental grid with more elements and a report grid. The problem is to find a mapping from the fundamental grid to the report grid as the information structure with the highest two grid values joined and the lowest as well. That mapping must involve some monotone pooling, and the problem is what pooling is optimal in maximizing the expected utility for a parsimonious decision problem. Such pooling creates curvature for the mapping, which manifests as various interesting phenomena in practice.

Similar extensions include, for example, studying exam scores on an equidistant score table that reflect the exam-takers' skill, supposing the score serves a decision.

6 Concluding Remarks

This paper identifies information selection under the physical communication capacity as a reason for content bias such as appealing to the audience and sensationalism, and provides asymptotic characterizations. The bias stems from the sender optimally compressing fundamental information based on its newsworthiness. It is apparent and improves welfare.

The modeled content generation channel is independent of any preference or payoff discrepancies. If such discrepancies are the primary concern, other persuasion-related channels may also be in place. The information structure may involve mixed strategies and the two criteria may constrain optimization and become at odds with each other. This paper's channel remains relevant if information selection is also a concern.

The model can be applied widely to settings involving information selection. In these applications, the model separates the content's literal meaning and inferred fundamental meaning, while connecting the two in a tractable and smooth function involving contextual economic parameters. This paper lists several examples: The model brings a new perspective on media slants. It can lay an economic foundation for certain content analysis involving textual frequencies, such as sentiment analysis. Its variation can explain ratings data. It has empirical potential in interpreting content data in context and parameterizing a model of such data for estimation that is relieved from misspecification errors.

References

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

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

Benabou, R. and Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *The Quarterly Journal of Economics*, 107(3):921–958.

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

Bloedel, A. and Segal, I. (2020). Persuasion with rational inattention. *Working Paper*.

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

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley & Sons.

Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research*, pages 123–145.

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

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–74.

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

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

Goldman, E., Martel, J., and Schneemeier, J. (2021). A theory of financial media. *Journal of Financial Economics*.

Green, J. R. and Laffont, J.-J. (1986). Partially verifiable information and mechanism design. *The Review of Economic Studies*, 53(3):447–456.

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

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Milgrom, P. and Roberts, J. (1986). Relying on the information of interested parties. *The RAND Journal of Economics*, pages 18–32.

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

Perego, J. and Yuksel, S. (2020). Media competition and social disagreement. *Working Paper*.

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

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American Economic Review*, 49(3):253–283.

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

Stein, J. C. (1989). Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *The quarterly journal of economics*, 104(4):655–669.

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

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

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

Appendix A. Propositions 1, 2, and 3

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

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

where $a^*(k)$ is $\arg\max_a \pi'_k u_1 h(a) + (1 - \pi'_k) u_0 h(1 - a)$ if $\sum_K \sigma_{Kk} > 0$ for that k and is 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 it has a maximum.

Proof of Propositions 1, 2. Prove Prop 1' → 2 → 1 in turn, where Prop 1' is a weaker version of Prop 1:

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

Proof. (Proposition 1') For σ_{Kk},

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

The optimality of $a^*(k)$ for such k that $\sum_K \sigma_{Kk} > 0$ is characterized by the F.O.C.: $0 = \pi'_k u_1 h'(a) - (1 - \pi'_k) u_0 h'(1 - a)$. 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,

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

and ≥ 0 by $h''(\cdot) < 0$. 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 and hence Eq. (A.2) and Eq. (A.3) hold. Write U as $U(\sigma_{Kk_1}, \sigma_{Kk_2})$. Discuss the following scenarios:

(i). If $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} \neq \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$, w.l.o.g. $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} > \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$, then for a small ε, $\sigma^{**}_{Kk_1} = \sigma^*_{Kk_1} + \varepsilon$ and $\sigma^{**}_{Kk_2} = \sigma^*_{Kk_2} - \varepsilon$ will improve U. That is because $\frac{\partial U}{\partial \sigma_{Kk_1}}$ is continuous in σ_{Kk_1} and so is $\frac{\partial U}{\partial \sigma_{Kk_2}}$ in σ_{Kk_2}, and hence $\exists \varepsilon > 0$ such that $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}+\varepsilon'} > \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}-\varepsilon'}$, $\forall \varepsilon' \in (0, \varepsilon]$; the increase in utility after changing to $\sigma^{**}_{Kk_1}$ and $\sigma^{**}_{Kk_2}$ is $U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2}) = U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) + U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2})$, and by Eq. (A.3) is $\int_0^\varepsilon \frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}+\varepsilon'} d\varepsilon' - \int_0^\varepsilon \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}-\varepsilon''} d\varepsilon'' > 0$. It is contradictory to optimality;

(ii). If $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} = \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$ and $\nexists \delta > 0$ s.t. 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 $(\sigma^*_{Kk_2} - \delta, \sigma^*_{Kk_2} + \delta)$, and w.l.o.g. assume $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} > 0$ 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 the reason: The utility increase is $U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2}) = U(\sigma^{**}_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) + U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2})$, and by Eq. (A.3) is $U(\sigma^*_{Kk_1}, \sigma^{**}_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2}) + U(\sigma^{**}_{Kk_1}, \sigma^*_{Kk_2}) - U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2})$. Define a continuous function $f(x) := U(\sigma^*_{Kk_1}, \sigma^*_{Kk_2} + x)$ if $x \in [-\delta_0, 0]$ and $:= U(\sigma^*_{Kk_1} + x, \sigma^*_{Kk_2})$ if $x \in (0, \delta_0]$. $f(x)$ is convex on $[-\delta_0, \delta_0]$ and strictly convex on $(0, \delta_0]$. The utility increase is $f(\delta_0) + f(-\delta_0) - 2f(0)$, positive by Jensen's Inequality. It is contradictory to optimality;

(iii). If $\frac{\partial U}{\partial \sigma_{Kk_1}}|_{\sigma^*_{Kk_1}} = \frac{\partial U}{\partial \sigma_{Kk_2}}|_{\sigma^*_{Kk_2}}$ and $\exists \delta > 0$ s.t. 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 $(\sigma^*_{Kk_2} - \delta, \sigma^*_{Kk_2} + \delta)$, one of the two following cases will occur.

(iii-a). If $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} > 0$ or $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} > 0$ somewhere on $(0, \sigma^*_{Kk_1} + \sigma^*_{Kk_2})$, and w.l.o.g. assume 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 $\sigma_{Kk_2} \in (\sigma^*_{Kk_2} - (\sigma^0 - \sigma^*_{Kk_1}), \sigma^*_{Kk_2})$, then let $\sigma^0_{Kk_1} = \sigma^0$ and $\sigma^0_{Kk_2} = \sigma^*_{Kk_2} - (\sigma^0 - \sigma^*_{Kk_1})$. We get a new information structure falling under scenario (ii) with the same utility. Resort to the reasoning of (ii) by letting $\sigma^0_{Kk_1}$ and $\sigma^0_{Kk_2}$ be the new $\sigma^*_{Kk_1}$ and $\sigma^*_{Kk_2}$ to find a contradiction to optimality;

(iii-b). If $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} = 0$ 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 $\sigma^{**}_{Kk_2} = 0$.

(iii-b-1). If $\sigma^*_{Kk_1} + \sigma^*_{Kk_2} = 1$, then the new strategy has the same utility and does not involve mixing for K. If no other K' involves mixing, then we have found a pure strategy that delivers the same utility as the optimal mixed strategy. If some other K' involves mixing, then let K' be the new K and iterate the discussion of (i)(ii)(iii);

(iii-b-2). If $\sigma_{Kk_1^*} + \sigma_{Kk_2^*} < 1$, then $\exists k_3$ s.t. $\sigma^*_{Kk_3} \in (0, 1)$. Let k_3 be the new k_2 and iterate the discussion of (i)(ii)(iii).

Hence, among all possible scenarios, the existence of mixing is only plausible under (iii-b-1). A mixed strategy exists in equilibrium only when it delivers the same utility as a pure strategy. Therefore, a pure strategy equilibrium always exists. □

Lemma 1. *Under Assumption 1 and with p_K/q_K strictly increasing in K, 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. (Lemma 1) From the proof of Proposition 1', the scenario for mixing to possibly occur in equilibrium is (iii-b-1) when $\sigma^*_{Kk_1} + \sigma^*_{Kk_2} = 1$. In that scenario, $\partial U / \partial \sigma_{Kk_1}$ is a constant, implying that $a^*(k_1)$ is constant with respect to σ_{Kk_1} by Eq. (A.1). Hence, by the F.O.C., σ_{Kk_1} does not affect $\pi'_{k_1}/(1 - \pi'_{k_1})$. Since p_K/q_K differ across K, it can be inferred that $\sigma_{K'k_1} = 0$, $\forall K' \neq K$. The same argument goes for k_2. Namely, K is the only fundamental value that maps to k_1 or k_2. The utility-equivalent pure strategy must involve either $\sum_K \sigma^{**}_{Kk_1} = 0$ or $\sum_K \sigma^{**}_{Kk_2} = 0$, i.e., some k ends up unused. □

Proof. (Proposition 2) By Proposition 1' and Lemma 1, I only need to consider pure strategy equilibria. Prove 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 there exist a report associated with $m \geq 2$ fundamentals. Let this set of fundamentals be $\{K^{(1)}, ..., K^{(m)}\}$ and their report be k_2. The contribution to utility by k_2 is U_{k_2} and the optimal action for k_2 is $a^*(k_2)$. Consider an alternative pure strategy with $K^{(1)}$ mapped to k_1, $\{K^{(2)}, ..., K^{(m)}\}$ mapped to k_2, and the rest of the strategy the same as the old strategy. The contributions to utility by k_1 and k_2 are denoted as U'_{k_1} and U'_{k_2}, and the optimal actions as $a'^*(k_1)$ and $a'^*(k_2)$. Therefore,

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

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 be equal by strictly monotone p_K/q_K. It is contradictory to optimality. □

Proof. (Proposition 1) By Proposition 2, the mixed strategy possibility in Lemma 1 is ruled out. The equilibrium only involves pure strategies. □

Proof of Proposition 3. Prove Proposition 3(i), and 3(ii) will naturally follow. For fundamental K, denote $x_K = \pi u_1 p_K$ and $y_K = (1-\pi)u_0 q_K$. Each fundamental can be fully characterized by (x_K, y_K). Let $\eta_K := y_K/x_K = \frac{\pi}{1-\pi}\frac{u_1}{u_0}\Lambda(K)$.

Lemma 2. *Let Assumption 1 hold, the fundamental space be $\{K_0, ..., K_N\}$ with unequal η_K, the optimal partition be $\{B_k^*\}_{k=0}^n$, and B_n^* consist of $m \geq 2$ fundamentals (by Prop 2, $m \leq N - n + 1$). Define fundamental v by $x_v = \sum_{K \in B_n} x_K$ and $y_v = \sum_{K \in B_n} y_K$. For the alternative problem with the 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. (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 fundamental probabilities 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 with all elements of B_n in the partition set where v belongs among $\{K_0, ..., K_N\} \cup \{v\}\backslash B_n$. Namely, the range of utility for partitioning $\{K_0, ..., K_N\} \cup \{v\}\backslash B_n$ is a subset of the range for partitioning $\{K_0, ..., K_N\}$. Therefore, if $\{B_k^*\}_{k=0}^n$ is the optimal partition of $\{K_0, ..., K_N\}$, 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. (Proposition 3) Prove by induction. Let $P(N, n)$ be the problem of choosing n reported elements from N signals, and the goal is to show $P(N, n)$ has the property that its solution involves an ordered partition of the fundamental space. I proceed by first showing the property for $P(N, 1)$ by induction starting from $P(2, 1)$ and $P(3, 1)$, and second showing it for $P(N, n)$. W.l.o.g., let $\Lambda(0) < \Lambda(1) < ... < \Lambda(K)$.

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

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

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

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

Need to show $s = t = 0$ is a suboptimal strategy. By the Envelope Theorem,

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

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

Consider two scenarios:

(i). If $a \neq b$ at $s = t = 0$, then as long as one of the two partial derivatives is positive at $s = 0$ or $t = 0$, and for instance say it is $\frac{\partial U}{\partial s}|_{s=0} > 0$, the pure strategy at $s = t = 0$

is strictly worse than a strategy with a small and positive s. That is a mixed strategy and thus by Proposition 1 is suboptimal. Therefore, I only need to show Eq. (A.4) or Eq. (A.5) is positive.

Show by contradiction. Suppose both Eq. (A.4) and Eq. (A.5) are nonpositive at $s = t = 0$. Then on one hand, $\eta_1 = \frac{h'(b)}{h'(1-b)}$ by the optimality of b. On the other hand, however, if $a > b$,

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

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

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

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

(ii). If $a = b$ 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, $\frac{y_0+y_2}{x_0+x_2} = \frac{h'(a)}{h'(1-a)} = \frac{h'(b)}{h'(1-b)} = \frac{y_1}{x_1}$, and hence $\frac{h'(a)}{h'(1-a)} = \frac{h'(b)}{h'(1-b)} = \frac{y_0+y_1+y_2}{x_0+x_1+x_2}$. Therefore, for $s = t = 1$, the optimal action is also a. The utility $U|_{s=t=1} = U|_{s=t=0}$. 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. Q.E.D. for Step 1.1.

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

Prove by contradiction. Possible non-ordered partitions $\{B_k\}_{k=0}^{1}$ are $\{\{1\}, \{0, 2, 3\}\}$, $\{\{2\}, \{0, 1, 3\}\}$, $\{\{0, 3\}, \{1, 2\}\}$ and $\{\{0, 2\}, \{1, 3\}\}$. I discuss why they are suboptimal:

(i). For $\{\{1\}, \{0, 2, 3\}\}$ and $\{\{2\} , \{0, 1, 3\}\}$: Suppose w.l.o.g. the optimal partition is $\{\{1\} , \{0, 2, 3\}\}$. 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's result, $\{\{1\}, \{0, v\}\}$ cannot be optimal because it is not a ordered partition; a contradiction.

(ii). For $\{\{0, 3\}, \{1, 2\}\}$: Suppose it 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 a $P(2, 1)$, and since $\eta_0 < \eta_v < \eta_3$, by Step 1.1's result, $\{\{v\}, \{0, 3\}\}$ cannot be optimal because it is not an ordered partition; a contradiction.

(iii). For $\{\{0, 2\}, \{1, 3\}\}$: Suppose it is optimal. Consider two scenarios.

(iii-a). If $a > b$, then consider the following strategy: (1). the fundamentals that map to k_0 are $(0, 1, 2, 3)$ with probabilities $(1, s, 1 - t, 0)$ respectively; the optimal action is a; and (2). the fundamentals that map to k_1 are $(0, 1, 2, 3)$ with probabilities $(0, 1 - s, t, 1)$ respectively; the optimal action is b. Thus,

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

Need to show $s = t = 0$ is a suboptimal strategy. By the Envelope Theorem,

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

Analogous to the argument in Step 1.1, I only need to show one of the two partial derivatives is positive at $s = 0$ or $t = 0$. Suppose both are nonpositive, then $\eta_2 \leq \frac{h(a)-h(b)}{h(1-b)-h(1-a)} \leq \eta_1$, contradictory to the assumption that $\eta_2 > \eta_1$.

(iii-b). If $a < b$, then consider the following strategy: (1). the fundamentals that map to k_0 are $(0, 1, 2, 3)$ with probabilities $(1 - t, 0, 1, s)$ respectively; the optimal action is a; and (2). the fundamentals that map to k_1 are $(0, 1, 2, 3)$ with probabilities $(t, 1, 0, 1 - s)$ respectively; the optimal action is b. Thus,

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

Need to show $s = t = 0$ is a suboptimal strategy. By the Envelope Theorem,

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

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

(iii-c). If $a = b$, 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 bundling $\{\{0,1,2,3\}, \varnothing\}$ delivers the same optimal actions as $s = t = 0$ and the same utility. However, the alternative strategy does not use both reports and thus is suboptimal by Proposition 2. Hence, the strategy at $s = t = 0$ is also suboptimal. Q.E.D. for 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 ($N \geq 4$).

Prove by contradiction. Suppose the optimal partition $\{B_0, B_1\}$ for $P(N, 1)$ is not an ordered partition, discuss two cases:

(i). If there exist neighboring fundamentals $i, i+1 \in B_0$ (or B_1, here pick B_0 w.l.o.g.), then let v be a fundamental and $(x_v, y_v) = (x_i + x_{i+1}, y_i + y_{i+1})$. Hence $\Lambda(i) < \Lambda(v) < \Lambda(i+1)$. Consider the $P(N-1, 1)$ problem of grouping $\{0, 1, ..., i-1, v, i+2, ..., N\}$. By the optimality of $\{B_0, B_1\}$, the solution has to be $\{B'_0, B'_1\}$ by Lemma 2, with $B'_0 = B_0 \cup \{v\} \backslash \{i, i+1\}$ and $B'_1 = B_1$. However, this is not a ordered partition; a contradiction.

(ii). It there do not exist patterns in (i), i.e. $B_0 = \{0, 2, 4, ...\}$ and $B_1 = \{1, 3, 5, ...\}$, then let v be a fundamental and $(x_v, y_v) = (x_0 + x_2, y_0 + y_2)$. Hence $\Lambda(v) < \Lambda(2)$. Consider the $P(N-1, 1)$ problem of grouping $\{v, 1, 3, 4, ..., N\}$. By the optimality of $\{B_0, B_1\}$, the solution has to be $\{B'_0, B'_1\}$ by Lemma 2, with $B'_0 = B_0 \cup \{v\} \backslash \{0, 2\}$ and $B'_1 = B_1$. However, $\Lambda(v) < \Lambda(3) < \Lambda(4)$ and therefore 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 and let the solution be $\{B_0,...,B_n\}$. For any $0 \le i,j \le n$ $(i \ne j)$, get $B_i = \{K_1^i, K_2^i, ..., K_{m_i}^i\}$ and $B_j = \{K_1^j, K_2^j\ , ..., K_{m_j}^j\}$ and consider the problem $P(m_i+m_j,1)$ with the fundamental space $\{K_1^i, K_2^i, ..., K_{m_i}^i, K_1^j, K_2^j, ..., K_{m_j}^j\}$. By the assumed optimality of $\{B_0,...,B_n\}$, the solution has to be $\{B_i, B_j\}$ to avoid contradiction. By the result of Step 1, it is an ordered partition. Therefore, for any two partition sets in $B_0,...,B_n$, there is a cutoff with each set of fundamentals on a different side. The only way possible is that $\{B_0,...,B_n\}$ is an ordered partition. □

Appendix B. Theorem 2

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

(i) $\tilde{a}(K)$ is strictly increasing on $(\underline{K}, \bar{K})$.

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

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

Proof. (i) $\tilde{a}(K)$ is determined by

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

The LHS strictly increases in K by Assumption 4(i)(c). The RHS strictly increases in a by Assumption 3. Hence, $\tilde{a}(K)$ is strictly increasing.

(ii) For 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 RHS strictly increases in a. The LHS is in $(\frac{f_{K|\theta=1}(K_1)}{f_{K|\theta=0}(K_1)}, \frac{f_{K|\theta=1}(K_2)}{f_{K|\theta=0}(K_2)})$ by Assumption 4(i)(c). Hence, $\tilde{a}(K_1) < a^* < \tilde{a}(K_2)$. Apply this result to all partition sets.

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

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

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

and (ii)

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

Proof. (i) Prove by contradiction. Suppose $\exists \delta > 0$ such that for a subsequence $\{n_j\}$,

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

With slight abuse of notation, denote the cutoffs that achieve the max for n_j as K_{i+1}^*, K_i^*. Then an alternative information structure with cutoffs $\kappa^*(n) \cup \{(K_{i+1}^* + K_i^*)/2\}$ delivers a higher utility than $\kappa^*(n)$ by at least w, where $w := \min_{K', K'+\delta \in [\underline{K}, \bar{K}]} \Pr(K \in [K, K+\delta])$ $\left\{\mathrm{E}\left[\max_{a\in[0,1]} \mathrm{E}[u(a;\theta)|\mathbf{1}_{K<K'+\delta/2}]|K \in [K', K'+\delta]\right] - \max_{a\in[0,1]} \mathrm{E}\left[u(a;\theta)|K \in [K', K'+\delta]\right]\right\}$. Because of Assumption 4(i)(c) on $(\underline{K}, \bar{K})$, the optimal actions on $(K', K'+\delta/2)$ and $(K'+\delta/2, K'+\delta)$ cannot coincide for any K' such that $K', K'+\delta \in (\underline{K}, \bar{K})$. Hence, $w > 0$. Here, w is a constant.

Meanwhile, let $\hat{\kappa}_1(n_1), \hat{\kappa}_2(n_2) \subset [\underline{K}, \bar{K}]$ be two sets of cutoffs and $\hat{\kappa}_1(n_1) \subset \hat{\kappa}_2(n_2)$. They respectively deliver expected utility levels $\hat{U}_1$ and $\hat{U}_2$. Obviously, $\hat{U}_1 \le \hat{U}_2$. Also note that $\hat{U}_2 \le M := h(1)\max\{u_0, u_1\} < +\infty$ and $\hat{U}_1 \ge m := \min_{a\in[0,1]} \pi u_1 h(a) + (1-\pi)u_0 h(1-a) > -\infty$. Hence, $\hat{U}_2 - \hat{U}_1 \le M - m$. Here, M, m are constants.

For n_j, let $\hat{\kappa}_2(n_2)$ be $\kappa^*(n_j)$ and $\hat{\kappa}_1(n_1)$ be $\{K_i^* \in \kappa^*(n_j) | i \text{ is odd}\}$. Then, $\hat{U}_2 - \hat{U}_1 = \sum_{t=1,\dots,\lfloor\frac{n_j+1}{2}\rfloor} \nu_t$, where $\nu_t = \Pr(B_{2t-2} \cup B_{2t-1}) \times$

$$\left\{\mathrm{E}[\max_{a\in[0,1]} E[u(a;\theta)|\mathbf{1}_{B_{2t-2}}]|B_{2t-2} \cup B_{2t-1}] - \max_{a\in[0,1]} \mathrm{E}[u(a;\theta)|B_{2t-2} \cup B_{2t-1}]\right\}$$

for $t < \lfloor\frac{n_j+1}{2}\rfloor$ and $\nu_t = \Pr\left(\cup_{t\ge 2\lfloor\frac{n_j+1}{2}\rfloor} B_t\right) \times$

$$\left\{\mathrm{E}[\max_{a\in[0,1]} E[u(a;\theta)|\mathbf{1}_{B_{2t-2}} + \mathbf{1}_{B_{2t-2}\cup B_{2t-1}}]| \cup_{t\ge 2\lfloor\frac{n_j+1}{2}\rfloor} B_t] - \max_{a\in[0,1]} \mathrm{E}[u(a;\theta)| \cup_{t\ge 2\lfloor\frac{n_j+1}{2}\rfloor} B_t]\right\}$$

for $t = \lfloor\frac{n_j+1}{2}\rfloor$. Each $\nu_t \ge 0$ and, to avoid contradiction to the upper bound $M - m$,

$\min_t \nu_t \leq 2(M-m)/n_j$. As $n_j \to \infty$, $2(M-m)/n_j \to 0$, and hence $\min_t \nu_t \to 0$. Therefore, given w, there exists a large enough n_j such that $\min_t \nu_t < w$. Let this min be achieved by $\tilde{t}$. Thus, the cutoffs $\kappa^*(n_j) \cup \{(K_i^* + K_{i+1}^*)/2\} \backslash \{K_{2\tilde{t}-1}^*\}$ pin down an alternative information structure better than the cutoffs $\kappa^*(n_j)$ by at least $w - \min_t \nu_t > 0$, a contradiction.

(ii) Prove by contradiction. Suppose $\exists M > \underline{K}$ such that $K_1^* > M$, $\forall n$. Then define K^{**} as $(\inf\{K_1^* : n \geq 1\} + \underline{K})/2$ if $\underline{K} > -\infty$ and $\inf\{K_1^* : n \geq 1\} - 1$ if $\underline{K} = -\infty$. Then an alternative information structure with cutoffs $\kappa^*(n) \cup \{K^{**}\}$ delivers a higher utility by at least $w > 0$. Meanwhile, use the same construction of μ_t, it can be found that $\kappa^*(n_j) \cup \{K^{**}\} \backslash \{K_{2\tilde{t}-1}^*\}$ is a better information structure, a contradiction. Analogous reasoning applies to $K_n^* \to \bar{K}$. □

Proof of Theorem 2.

Proof. (Theorem 2) Prove in three steps.

Step 1. Reformulating the Problem. Let

$$\delta_n(a) := \frac{1}{n} \sum_{K' \in \kappa^*(n)} \mathbf{1}_{\{\tilde{a}(K') \leq a\}}. \tag{B.1}$$

Because $\tilde{a}(K)$ is strictly increasing on $(\underline{K}, \bar{K})$ and $\kappa^*(n) \subset (\underline{K}, \bar{K})$ for all n by Lemma 3, $\beta_n(K) = \delta_n(\tilde{a}(K))$. Therefore, to show $\beta_n(K)$ converges, it suffices to first show $\delta_n(a)$ converges. By Lemma 4, it suffices to show that the function sequence

$$\hat{\delta}_n(a) = \begin{cases} \delta_n(\tilde{a}(K_i^*)) + \frac{\delta_n(a(K_{i+1}^*)) - \delta_n(\tilde{a}(K_i^*))}{\tilde{a}(K_{i+1}^*) - \tilde{a}(K_i^*)}(\tilde{a}(K) - \tilde{a}(K_i^*)), & \text{if } a \in [\tilde{a}(K_i^*), \tilde{a}(K_{i+1}^*)); \\ \frac{\delta_n(\tilde{a}(K_1^*))}{\tilde{a}(K_1^*)}\tilde{a}(K), & \text{if } a \in (0, \tilde{a}(K_1^*)); \\ 0, \quad \text{if } a \leq 0; \qquad 1, & \text{if } a \geq \tilde{a}(K_n^*) \end{cases}$$

converges. Functions $\hat{\delta}_n(a)$ are absolutely continuous cumulative distribution functions.

By Scheffé's Lemma (see Theorem in Scheffé (1947)), it suffices to show its density

$$d_n(a) = \begin{cases} \frac{\delta_n(\tilde{a}(K^*_{i+1}))-\delta_n(\tilde{a}(K^*_i))}{\tilde{a}(K^*_{i+1})-\tilde{a}(K^*_i)}, & \text{if } a \in [\tilde{a}(K^*_i), \tilde{a}(K^*_{i+1})); \\ \frac{\delta_n(\tilde{a}(K^*_1))}{\tilde{a}(K^*_1)}, & \text{if } a < \tilde{a}(K^*_1); \\ 0, & \text{otherwise} \end{cases}$$

pointwise converges to some limiting density almost everywhere.

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 notations, let $g_1(a) = F_1(\tilde{K}(a))$, $g_0(a) = F_0(\tilde{K}(a))$, $t_1(a) = h(a)$, $t_0(a) = h(1-a)$, $I_i = \tilde{a}(K^*_{i+1}) - \tilde{a}(K^*_i)$, and $J_i = a^*_i - a^*_{i-1}$. Then the F.O.C. for a^*_i is

$$\frac{g_1(\tilde{a}(K^*_{i+1})) - g_1(\tilde{a}(K^*_i))}{g_0(\tilde{a}(K^*_{i+1})) - g_0(\tilde{a}(K^*_i))} = \frac{g'_1(a^*_i)}{g'_0(a^*_i)}, \tag{B.2}$$

and for K^*_i is

$$\frac{t_1(a^*_i) - t_1(a^*_{i-1})}{t_0(a^*_i) - t_0(a^*_{i-1})} = \frac{t'_1(\tilde{a}(K^*_i))}{t'_0(\tilde{a}(K^*_i))}. \tag{B.3}$$

In Eq. (B.2), Taylor-expand $g_1(\tilde{a}(K^*_{i+1}))$, $g_1(\tilde{a}(K^*_i))$, $g_0(\tilde{a}(K^*_{i+1}))$, and $g_0(\tilde{a}(K^*_i))$ at $\bar{a}_i := (\tilde{a}(K^*_{i+1}) + \tilde{a}(K^*_i))/2$ to the fourth order and $g'_1(a^*_i)$ and $g'_0(a^*_i)$ at $\bar{a}_i$ to the first order, getting

$$\begin{aligned} a^*_i - \bar{a}_i =& \frac{1}{24}\frac{g'''_0(\bar{a}_i)g'_1(\bar{a}_i) - g'''_1(\bar{a}_i)g'_0(\bar{a}_i)}{g'_1(\bar{a}_i)g''_0(\bar{a}_i) - g'_0(\bar{a}_i)g''_1(\bar{a}_i)} I_i^2 + \frac{R_1 + R_2 + R_3 + R_4 + R_5 + R_6}{g'_1(\bar{a}_i)g''_0(\bar{a}_i) - g'_0(\bar{a}_i)g''_1(\bar{a}_i)} \\ :=& \Gamma(\bar{a}_i) I_i^2 + Rem^a_i, \end{aligned} \tag{B.4}$$

where

$$\begin{aligned} &R_1 = C_1(g'_0(\bar{a}_i)g'''_1(a_d) - g'_1(\bar{a}_i)g'''_0(a'_d)) && (a^*_i - \bar{a}_i)^2 \\ &R_2 = C_2(g'''_0(\bar{a}_i)g''_1(\bar{a}_i) - g'''_1(\bar{a}_i)g''_0(\bar{a}_i)) && I_i^2(a^*_i - \bar{a}_i) \\ &R_3 = C_3(g'''_0(\bar{a}_i)g'''_1(a_d) - g'''_1(\bar{a}_i)g'''_0(a'_d)) && I^2(a^*_i - \bar{a}_i)^2 \\ &R_4 = C_4((g^{(5)}_0(a''_c) - g^{(5)}_0(a'''_c))g'_1(\bar{a}_i) - (g^{(5)}_1(a_c) - g^{(5)}_1(a'_c))g'_0(\bar{a}_i)) && I_i^4 \\ &R_5 = C_5((g^{(5)}_0(a''_c) - g^{(5)}_0(a'''_c))g''_1(\bar{a}_i) - (g^{(5)}_1(a_c) - g^{(5)}_1(a'_c))g''_0(\bar{a}_i)) && I_i^4(a^*_i - \bar{a}_i) \\ &R_6 = C_6((g^{(5)}_0(a''_c) - g^{(5)}_0(a'''_c))g'''_1(a_d) - (g^{(5)}_1(a_c) - g^{(5)}_1(a'_c))g'''_0(a'_d)) && I_i^4(a^*_i - \bar{a}_i)^2. \end{aligned}$$

Here, a_d and a'_d are between a_i^* and $\bar{a}_i$ and are respectively in the Taylor remainders of $g'_1(a_i^*)$ and $g'_0(a_i^*)$. Also, a_c and a''_c are between $\tilde{a}(K_{i+1}^*)$ and $\bar{a}_i$ and are respectively in the Taylor remainders of $g_1(\tilde{a}(K_{i+1}^*))$ and $g_0(\tilde{a}(K_{i+1}^*))$. a'_c and a'''_c are between $\bar{a}_i$ and $\tilde{a}(K_i^*)$ and are respectively in the Taylor remainders of $g_1(\tilde{a}(K_i^*))$ and $g_0(\tilde{a}(K_i^*))$. With slight abuse of notations, subscripts i for all aforementioned notations are omitted. C_1, ..., C_6 are constants. Differentiation is possible by smoothness in Assumption 4 and Lemma 3. The denominator $g'_1(\bar{a}_i)g''_0(\bar{a}_i) - g'_0(\bar{a}_i)g''_1(\bar{a}_i) > 0$ when $\bar{a}_i \in (0,1)$ because of $(F'_1(K)/F'_0(K))' > 0$ in Assumption 4(i) and $\tilde{K}'(a) > 0$ in Lemma 3.

In Eq. (B.3), Taylor-expand $t_1(a_i^*)$, $t_1(a_{i-1}^*)$, $t_0(a_i^*)$, and $t_0(a_{i-1}^*)$ at $\bar{\bar{a}}_i := (a_i^* + a_{i-1}^*)/2$ to the fourth order and $t'_1(\tilde{a}(K_i^*))$ and $t'_0(\tilde{a}(K_i^*))$ at $\bar{\bar{a}}_i$ to the first order, getting

$$\begin{aligned}\tilde{a}(K_i^*) - \bar{\bar{a}}_i =& \frac{1}{24}\frac{t'''_0(\bar{\bar{a}}_i)t'_1(\bar{\bar{a}}_i) - t'''_1(\bar{\bar{a}}_i)t'_0(\bar{\bar{a}}_i)}{t'_1(\bar{\bar{a}}_i)t''_0(\bar{\bar{a}}_i) - t'_0(\bar{\bar{a}}_i)t''_1(\bar{\bar{a}}_i)}J_i^2 + \frac{S_1+S_2+S_3+S_4+S_5+S_6}{t'_1(\bar{\bar{a}}_i)t''_0(\bar{\bar{a}}_i) - t'_0(\bar{\bar{a}}_i)t''_1(\bar{\bar{a}}_i)} \\ :=& T(\bar{\bar{a}}_i)(a_i^* - a_{i-1}^*)^2 + Rem_i^K,\end{aligned} \tag{B.5}$$

where

$$\begin{aligned}
S_1 &= C_1(t'_0(\bar{\bar{a}}_i)t'''_1(a_{dd}) - t'_1(\bar{\bar{a}}_i)t'''_0(a'_{dd})) && (\tilde{a}(K_i^*) - \bar{\bar{a}}_i)^2 \\
S_2 &= C_2(t'''_0(\bar{\bar{a}}_i)t''_1(\bar{\bar{a}}_i) - t'''_1(\bar{\bar{a}}_i)t''_0(\bar{\bar{a}}_i)) && J_i^2(\tilde{a}(K_i^*) - \bar{\bar{a}}_i) \\
S_3 &= C_3(t'''_0(\bar{\bar{a}}_i)t'''_1(a_{dd}) - t'''_1(\bar{\bar{a}}_i)t'''_0(a'_{dd})) && J_i^2(\tilde{a}(K_i^*) - \bar{\bar{a}}_i)^2 \\
S_4 &= C_4((t_0^{(5)}(a''_{cc}) - t_0^{(5)}(a'''_{cc}))t'_1(\bar{\bar{a}}_i) - (t_1^{(5)}(a_{cc}) - t_1^{(5)}(a'_{cc}))t'_0(\bar{\bar{a}}_i)) && J_i^4 \\
S_5 &= C_5((t_0^{(5)}(a''_{cc}) - t_0^{(5)}(a'''_{cc}))t''_1(\bar{a}_i) - (t_1^{(5)}(a_{cc}) - t_1^{(5)}(a'_{cc}))t''_0(\bar{\bar{a}}_i)) && J_i^4(\tilde{a}(K_i^*) - \bar{\bar{a}}_i) \\
S_6 &= C_6((t_0^{(5)}(a''_{cc}) - t_0^{(5)}(a'''_{cc}))t'''_1(a_{dd}) - (t_1^{(5)}(a_{cc}) - t_1^{(5)}(a'_{cc}))t'''_0(a'_{dd})) && J_i^4(\tilde{a}(K_i^*) - \bar{\bar{a}}_i)^2.
\end{aligned}$$

Here, a_{dd} and a'_{dd} are between $\tilde{a}(K_i^*)$ and $\bar{\bar{a}}_i$ and are respectively in the Taylor remainders of $t'_1(\tilde{a}(K_i^*))$ and $t'_0(\tilde{a}(K_i^*))$. a_{cc} and a''_{cc} are between a_i^* and $\bar{\bar{a}}_i$ and are respectively in the Taylor remainders of $t_1(a_i^*)$ and $t_0(a_i^*)$. a'_{cc} and a'''_{cc} are between $\bar{\bar{a}}_i$ and $\tilde{a}(K_i^*)$ and are respectively in the Taylor remainders of $t_1(a_{i-1}^*)$ and $t_0(a_{i-1}^*)$. Again with slight abuse of notations, subscripts i for all aforementioned notations are omitted. Differentiation is possible by smoothness in Assumption 3. The denominator $t'_1(\bar{\bar{a}}_i)t''_0(\bar{\bar{a}}_i) - t'_0(\bar{\bar{a}}_i)t''_1(\bar{\bar{a}}_i) < 0$ when $\bar{\bar{a}}_i \in (0,1)$ by Assumption 3.

Then, we can get

$$I_i - I_{i-1} = -2\Gamma(\bar{a}_i)I_i^2 - 2\Gamma(\bar{a}_{i-1})I_{i-1}^2 - 4T(\bar{\bar{a}}_i)J_i^2 - 2Rem_i^a - 2Rem_{i-1}^a - 4Rem_i^K, \quad (B.6)$$

$$J_i - I_{i-1} = -2\Gamma(\bar{a}_{i-1})I_{i-1}^2 - 2T(\bar{\bar{a}}_i)J_i^2 - 2Rem_{i-1}^a - 2Rem_i^K. \quad (B.7)$$

Step 2. Investigating $d_n(x)/d_n(y)$ on $(0,1)$. Next, choose any small and positive number α and examine the ratio $d_n(x)/d_n(y)$, $\forall x, y \in [\alpha, 1-\alpha]$ $(x < y)$. Define $i_z := \max\{i : \tilde{a}(K_i^*) \leq z\}$ $(z \in (0,1))$. By Lemma 4, $\lim_{n\to\infty} K_1^* = \underline{K}$, so $d_n(x)$ $(x \in [\alpha, 1-\alpha])$ should be expressed for a large n as

$$\frac{\delta_n(\tilde{a}(K_{i_x+1}^*)) - \delta_n(\tilde{a}(K_{i_x}^*))}{\tilde{a}(K_{i_x+1}^*) - \tilde{a}(K_{i_x}^*)} = \frac{1/n}{I_{i_x}}.$$

Thus, for any given large n,

$$\frac{d_n(x)}{d_n(y)} = \frac{(1/n)/I_{i_x}}{(1/n)/I_{i_y}} = \frac{I_{i_y}}{I_{i_x}} = \frac{I_{i_y}}{I_{i_y-1}} \times \ldots \times \frac{I_{i_x+1}}{I_{i_x}} = \exp\left(\ln\left(\frac{I_{i_y}}{I_{i_y-1}}\right) + \ldots + \ln\left(\frac{I_{i_x+1}}{I_{i_x}}\right)\right). \quad (B.8)$$

And since Taylor-expanding $\ln\left(\frac{I_i}{I_{i-1}}\right)$ at 1 to the first order gives

$$\ln\left(\frac{I_i}{I_{i-1}}\right) = \frac{I_i}{I_{i-1}} - 1 - \frac{1}{2t_i}\left(\frac{I_i}{I_{i-1}} - 1\right)^2$$

where t_i is between $\frac{I_i}{I_{i-1}}$ and 1, Eq. (B.8) becomes

$$\frac{d_n(x)}{d_n(y)} = \exp\left(\sum_{i=i_x+1}^{i_y}\left(\frac{I_i}{I_{i-1}} - 1\right)\right) / \exp\left(\sum_{i=i_x+1}^{i_y}\frac{1}{2t_i}\left(\frac{I_i}{I_{i-1}} - 1\right)^2\right), \quad (B.9)$$

where I_i/I_{i-1} is backed out from Eq. (B.6) to be

$$\frac{I_i}{I_{i-1}} = 1 - 2\Gamma(\bar{a}_i)I_i\frac{I_i}{I_{i-1}} - 2\Gamma(\bar{a}_{i-1})I_{i-1} - 4T(\bar{\bar{a}}_i)J_i\frac{J_i}{I_{i-1}} - \frac{2Rem_i^a + 2Rem_{i-1}^a + 4Rem_i^K}{I_{i-1}}. \quad (B.10)$$

In that expression, J_i/I_{i-1} is given by Eq. (B.7) as

$$\frac{J_i}{I_{i-1}} = 1 - 2\Gamma(\bar{a}_{i-1})I_{i-1} - 2T(\bar{\bar{a}}_i)J_i\frac{J_i}{I_{i-1}} - \frac{2Rem^a_{i-1} + 2Rem^K_i}{I_{i-1}}. \tag{B.11}$$

Now, discuss Eq. (B.9). First, I show three useful results for later analysis.

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

To show Result 1, for brevity let $X_i = I_i/I_{i-1}$ and $Y_i = J_i/I_{i-1}$. Eq. (B.10) and Eq. (B.11) can be rearranged into the following system of two equations

$$X_i = 1 - 2\Gamma(\bar{a}_i)I_iX_i - 2\Gamma(\bar{a}_{i-1})I_{i-1} - 4T(\bar{\bar{a}}_i)J_iY_i - 2\frac{Rem^a_i/I_{i-1}}{X_i}X_i - 2Rem^a_{i-1}/I_{i-1} - \frac{4Rem^K_i/I_{i-1}}{Y_i}Y_i,$$

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

Solving the system of equations can get us

$$X_i = \frac{1 - 2\Gamma(\bar{a}_{i-1})I_{i-1} - 2Rem^a_{i-1}/I_{i-1} - (4T(\bar{\bar{a}}_i)J_i + \frac{4Rem^K_i/I_{i-1}}{Y_i})Y_i}{1 + 2\Gamma(\bar{a}_i)I_i + 2\frac{Rem^a_i/I_{i-1}}{X_i}}; \tag{B.12}$$

$$Y_i = \frac{1 - 2\Gamma(\bar{a}_{i-1})I_{i-1} - 2Rem^a_{i-1}/I_{i-1}}{1 + 2T(\bar{\bar{a}}_i)J_i + \frac{2Rem^K_i/I_{i-1}}{Y_i}}. \tag{B.13}$$

Here, Γ, T and all functions appearing in Rem^a and Rem^K consisting of higher order derivatives of g and t are continuous on $[\alpha, 1-\alpha]$ by the continuous differentiability assumption and therefore bounded. Denote the uniform upper bound of their absolute values as $M > 0$.

Notice that in Rem^a_{i-1}, $|a^*_{i-1} - \bar{a}_{i-1}| \leq I_{i-1}$ since both $a^*_{i-1}, \bar{a}_{i-1} \in [\tilde{a}(K^*_{i-1}), \tilde{a}(K^*_i)]$ by Lemma 3(ii). Hence, by Triangular Inequality, $|Rem^a_{i-1}| \leq M(I^2_{i-1} + I^3_{i-1} + I^4_{i-1} + I^4_{i-1} + I^5_{i-1} + I^6_{i-1})$, and hence $|Rem^a_{i-1}/I_{i-1}| \leq M(I_{i-1} + I^2_{i-1} + 2I^3_{i-1} + I^4_{i-1} + I^5_{i-1})$. By $\max_i I_{i-1} \to 0$ in Lemma 4, $\max_i |Rem^a_{i-1}/I_{i-1}| \to 0$.

Also notice that in Rem^K_i, $|\tilde{a}(K^*_i) - \bar{\bar{a}}_i| \leq J_i$ since both $\tilde{a}(K^*_i), \bar{\bar{a}}_i \in [a^*_{i-1}, a^*_i]$. Hence, $|Rem^K_i| \leq M(J^2_i + J^3_i + J^4_i + J^4_i + J^5_i + J^6_i)$, and hence $|(Rem^K_i/I_{i-1})/Y_i| = |Rem^K_i/J_i| \leq M(J_i + J^2_i + 2J^3_i + J^4_i + J^5_i)$. By $\max_i J_i \to 0$ in Lemma 4, $\max_i |(Rem^K_i/I_{i-1})/Y_i| \to 0$.

Hence in Eq. (B.13), $|\Gamma|, |T| \leq M$, and all indexed terms uniformly converge to 0.

Thus, $Y_i \to 1$ uniformly over $[\alpha, 1-\alpha]$.

Further notice that in Rem_i^a, $|a_i^* - \bar{a}_i| \leq I_i$ since both $a_i^*, \bar{a}_i \in [\tilde{a}(K_i^*), \tilde{a}(K_{i+1}^*)]$. 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 hence $|(Rem_i^a/I_{i-1})/X_i| = |Rem_i^a/I_i| \leq M(I_i + I_i^2 + 2I_i^3 + I_i^4 + I_i^5)$. By $\max_i I_i \to 0$ in Lemma 4, $\max_i |(Rem_i^a/I_{i-1})/X_i| \to 0$.

Hence in Eq. (B.12), $|\Gamma|, |T| \leq M$, Y_i uniformly converges to 1 and all other indexed terms uniformly converge to 0. Thus, $X_i \to 1$ uniformly over $[\alpha, 1-\alpha]$. Therefore, Result 1 holds.

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.

Result 2 is a direct corollary of Result 1.

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 $z = \min(a_m + b_m + c_m)_{m=1}^M$. Define $o_i(z)$ as the notation of $\sum_{m=1}^M I_{i-1}^{a_m} I_i^{b_m} J_i^{c_m}$.

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

To show Result 3, let $a, b, c \geq 0$ be integers and only need to show $\sum_{i_x+1}^{i_y} I_{i-1}^a I_i^b J_i^c \to 0$ for $x, y \in [\alpha, 1-\alpha]$ if $a + b + c \geq 2$. W.l.o.g. let $a > 1$, then $0 \leq \sum_{i_x+1}^{i_y} I_{i-1}^a I_i^b J_i^c \leq (\sum_{i_x+1}^{i_y} I_{i-1})(\max_i I_{i-1})^{a-1}(\max_i I_i)^b(\max_i J_i)^c = (y-x)(\max_i I_{i-1})^{a-1}(\max_i I_i)^b(\max_i J_i)^c \to 0$. The proof also works if letting $b > 1$ or $c > 1$.

With Results 1, 2, and 3, go examine Eq. (B.9). We need to understand $I_i/I_{i-1} - 1$. Substitute Eq. (B.10) and Eq. (B.11) into the RHS of Eq. (B.10), and deduct one on both sides, getting

$$\frac{I_i}{I_{i-1}} - 1 = -2\Gamma(\bar{a}_i)I_i - 2\Gamma(\bar{a}_{i-1})I_{i-1} - 4T(\bar{\bar{a}}_i)J_i + Residual_i, \tag{B.14}$$

where $Residual_i =$

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

$$-4T(\bar{\bar{a}}_i)J_i\left(-2\Gamma(\bar{a}_{i-1})I_{i-1} - 2T(\bar{\bar{a}}_i)J_i\frac{J_i}{I_{i-1}} - \frac{2Rem_{i-1}^a + 2Rem_i^K}{I_{i-1}}\right)$$

$$-\frac{2Rem_i^a + 2Rem_{i-1}^a + 4Rem_i^K}{I_{i-1}}.$$

In $Residual_i$, first examine Rem_i^a. For terms R_2, R_3, R_4, R_5 and R_6 in Rem_i^a, because $|a_i^* - \bar{a}_i| \leq I_i$ as shown in Result 1's proof, each term scaled by its denominator should have its absolute value respectively not exceeding some constant times I_i^3, I_i^4, I_i^4, I_i^5, and I_i^6. For the term R_1, firstly we know scaled $|R_1|$ is bounded by a constant times I_i^2, and secondly we can substitute Eq. (B.4) into the term and get $R_1 = C_1(g_0'(\bar{a}_i)g_1'''(a_d) - g_1'(\bar{a}_i)g_0'''(a_d'))(\Gamma(\bar{a}_i)I_i^2 + Rem_i^a)^2$, which contains terms with absolute values not exceeding some constant times I_i^4, $I_i^2 Rem_i^a$ (which does not exceed some 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 does not exceed some constant times I_i^4,...,I_i^{12}), so scaled $|R_1|$ is actually bounded by some constant times I_i^4 plus higher orders. Hence, $|Rem_i^a|$ is bounded by $\frac{M_1}{1+\eta}(I_i^3 + ... + I_i^{12})$, where M_1 is a constant. By Result 2, $I_i/I_{i-1} \leq 1+\eta$ for large n, and hence 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, if examine Rem_{i-1}^a in the same way, we can get $|\frac{Rem_{i-1}^a}{I_{i-1}}| \leq M_2(I_{i-1}^2 + ... + I_{i-1}^{11}) \leq M_2 o_i(2)$.

Then examine Rem_i^K. For terms S_2, S_3, S_4, S_5 and S_6 in Rem_i^K, because $|\tilde{a}_{K_i^*} - \bar{\bar{a}}_i| \leq J_i$ as shown in Result 1's proof, each term scaled by its denominator should have its absolute value respectively not exceeding some constant times J_i^3, J_i^4, J_i^4, J_i^5, and J_i^6. For the term S_1, firstly we know scaled $|S_1|$ is bounded by a constant times J_i^2, and secondly we can substitute Eq. (B.5) into the term and get $S_1 = C_1(t_0'(\bar{\bar{a}}_i)t_1'''(a_{dd}) - t_1'(\bar{\bar{a}}_i)t_0'''(a_{dd}'))(T(\bar{\bar{a}}_i)J_i^2 + Rem_i^K)^2$, which contains terms with absolute values not exceeding some constant times J_i^4, $J_i^2 Rem_i^K$ (which does not exceed some constant times $J_i^2(J_i^2 + J_i^3 + J_i^4 + J_i^4 + J_i^5 + J_i^6)$), and $(Rem_i^K)^2$ (which does not exceed some constant times J_i^4,...,J_i^{12}), so scaled $|S_1|$ is actually bounded by some constant times J_i^4 plus higher orders. Hence, $|Rem_i^K|$ is bounded by $\frac{M_3}{1+\eta}(J_i^3 + ... + J_i^{12})$, where M_3 is a constant. By Result 2, $J_i/I_{i-1} \leq 1+\eta$ for large n, and hence for large n, $|\frac{Rem_i^K}{I_{i-1}}| \leq \frac{M_3}{1+\eta}\frac{J_i}{I_{i-1}}(J_i^2 + ... + J_i^{11}) \leq M_3 o_i(2)$.

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

$$|Residual_i| \leq M_4 o_i(2) \tag{B.15}$$

on $[\alpha, 1-\alpha]$. Hence, with Result 3, we know

$$\sum_{i=i_x+1}^{i_y} Residual_i \to 0 \tag{B.16}$$

by $0 \leq |\sum_{i=i_x+1}^{i_y} Residual_i| \leq \sum_{i=i_x+1}^{i_y} |Residual_i| \leq \sum_{i=i_x+1}^{i_y} M_4 o_i(2) \to 0$. Taking the square of Eq. (B.14) and using Eq. (B.15), we can get for large n and some constant M_5, $|\frac{I_i}{I_{i-1}} - 1|^2 \leq M_5 o_i(2)$. Hence we know the denominator of Eq. (B.9)

$$\exp\left(\sum_{i=i_x+1}^{i_y} \frac{1}{2t_i}(\frac{I_i}{I_{i-1}} - 1)^2\right) \to 1, \tag{B.17}$$

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 examine Eq. (B.9). Substitute Eq. (B.14) in Eq. (B.9), and rewrite Eq. (B.9) as

$$\frac{d_n(x)}{d_n(y)} = \exp\left(\sum_{i=i_x+1}^{i_y} (-2\Gamma(\bar{a}_i)I_i - 2\Gamma(\bar{a}_{i-1})I_{i-1} - 4T(\bar{\bar{a}}_i)J_i)\right) \times Rest(n; [x,y]),$$

or equivalently

$$\frac{d_n(x)}{d_n(y)} = \exp\left(RS(-4\Gamma; [x,y]) + RS(-4T; [x,y])\right) \times Rest(n; [x,y]), \tag{B.18}$$

where $Rest(n;[x,y]) \to 1$ by Eq. (B.16) and Eq. (B.17), and $RS(f,[x,y])$ refers to a Riemann sum of the function f on $[x,y]$. Because $\int_x^y f'(a)/f(a)da = \ln(|f(a)|)|_x^y$, we get when $x, y \in [\alpha, 1-\alpha]$,

$$\begin{aligned}
\frac{d_n(x)}{d_n(y)} &\to \exp\left(-4\int_x^y \Gamma(a)da - 4\int_x^y T(a)da\right) \\
&= \left|\frac{g_1'(x)g_0''(x) - g_0'(x)g_1''(x)}{g_1'(y)g_0''(y) - g_0'(y)g_1''(y)}\right|^{\frac{1}{6}} \left|\frac{t_1'(x)t_0''(x) - t_0'(x)t_1''(x)}{t_1'(y)t_0''(y) - t_0'(y)t_1''(y)}\right|^{\frac{1}{6}} \\
&= \left|\frac{F_1'(\tilde{K}(x))F_0''(\tilde{K}(x)) - F_0'(\tilde{K}(x))F_1''(\tilde{K}(x))}{F_1'(\tilde{K}(y))F_0''(\tilde{K}(y)) - F_0'(\tilde{K}(y))F_1''(\tilde{K}(y))}\right|^{\frac{1}{6}} \left|\frac{h'(x)h''(1-x) + h'(1-x)h''(x)}{h'(y)h''(1-y) + h'(1-y)h''(y)}\right|^{\frac{1}{6}} \left|\frac{\tilde{K}'(x)}{\tilde{K}'(y)}\right|^{\frac{1}{2}}.
\end{aligned}$$

Under Assumption 3, $|h'(x)h''(1-x)+h'(1-x)h''(x)| = -(h'(x)h''(1-x)+h'(1-x)h''(x))$. Under the monotone likelihood ratio property in Assumption 4(i), $|F_1'(\tilde{K}(x))F_0''(\tilde{K}(x)) - F_0'(\tilde{K}(x))F_1''(\tilde{K}(x))| = F_0'(\tilde{K}(x))F_1''(\tilde{K}(x)) - F_1'(\tilde{K}(x))F_0''(\tilde{K}(x))$. By the integrability assumption in Theorem 2, the denominator as a function of y is integrable on $(0,1)$. Let $m(y)>0$ denote the denominator expression scaled by its integral on $(0,1)$ so that $\int_0^1 m(y)dy=1$. Hence, pointwise for $x,y\in[\alpha,1-\alpha]$,

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

For any $x,y\in(0,1)$, Eq. (B.19) holds because one can always find a small α such that $x,y\in[\alpha,1-\alpha]$.

Step 3. Closing the Proof. In this step, the first objective is to show for any $y\in(0,1)$,

$$\lim_{n\to\infty} d_n(y)=m(y). \tag{B.20}$$

To show Eq. (B.20), notice that by Eq. (B.19),

$$\frac{1}{m(y)}=\int_0^1\frac{m(x)}{m(y)}dx=\int_0^1\lim_{n\to\infty}\frac{d_n(x)}{d_n(y)}dx\leq\liminf_{n\to\infty}\int_0^1\frac{d_n(x)}{d_n(y)}dx=\liminf_{n\to\infty}\frac{1}{d_n(y)}=\frac{1}{\limsup\limits_{n\to\infty}d_n(y)},$$

where the inequality is by Fatou's Lemma and the following equality comes from $\int_0^1 d_n(x)dx=1$ for any n. Hence,

$$m(y)\geq\limsup_{n\to\infty}d_n(y)>0. \tag{B.21}$$

Therefore, $\limsup_{n\to\infty}d_n(y)$ and $d_n(y)$ are dominated by $m(y)$ which is integrable on $(0,1)$. By the Fatou-Lebesgue Theorem,

$$1=\limsup_{n\to\infty}1=\limsup_{n\to\infty}\int_0^1 d_n(y)dy\leq\int_0^1\limsup_{n\to\infty}d_n(y)dy\leq\int_0^1 m(y)dy=1$$

and hence $\int_0^1\limsup_{n\to\infty}d_n(y)dy=\int_0^1 m(y)dy$. With Eq. (B.21), this implies almost everywhere

$$\limsup_{n\to\infty}d_n(y)=m(y). \tag{B.22}$$

Next, I show $\liminf_{n\to\infty}d_n(y)=m(y)$ by contradiction. Suppose otherwise, then considering Eq. (B.22), there exists $y_0 \in (0,1)$ such that

$$\liminf_{n\to\infty} d_n(y_0) = m(y_0) - \delta \in [0, m(y_0)).$$

Hence there exists a convergent subsequence $\{d_{n_k}(y_0)\}$ such that

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

Because of Eq. (B.19), for any $x \in (0,1)$,

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

Then on the one hand,

$$\int_0^1 \lim_{k\to\infty} d_{n_k}(x)dx = \frac{m(y_0) - \delta}{m(y_0)},$$

but on the other hand, because

$$d_{n_k}(x) \le \limsup_{k\to\infty} d_{n_k}(x) \le \limsup_{n\to\infty} d_n(x) \le m(x),$$

by the Dominated Convergence Theorem,

$$\int_0^1 \lim_{k\to\infty} d_{n_k}(x)dx = \lim_{k\to\infty} \int_0^1 d_{n_k}(x)dx = \lim_{k\to\infty} 1 = 1,$$

a contradiction. Hence Eq. (B.20) is proven.

Finally, by the arguments in the beginning of Step 1, Eq. (B.20) implies $\delta_n(a) \to \int_0^a m(y)dy$. Hence,

$$\beta_\infty(K) := \lim_{n\to\infty} \beta_n(K) = \lim_{n\to\infty} \delta_n(\tilde{a}(K)) = \int_0^{\tilde{a}(K)} m(y)dy$$

and $\beta'_\infty(K) = m(\tilde{a}(K))\tilde{a}'(K) = \lambda_h(K)^{\frac{1}{6}} \lambda_F(K)^{\frac{1}{6}} \tilde{a}'(K)^{\frac{1}{2}}$. □

Appendix C. Propositions 4 and 5

Proof. (Proposition 4) (1). By the definition of $\tilde{a}(K)$,

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

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 $(\underline{K}, \bar{K})$. The symmetry obviously holds outside $(\underline{K}, \bar{K})$. The symmetry of $\tilde{a}(K)$ implies $\tilde{a}'(K)$'s symmetry about $K = K_{1/2}$.

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

Proof. (Proposition 5) (1). 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 $KK_{1/2} > 0$, $\beta'_\infty(K) = \exp\Big(-\frac{K^2}{6\sigma^2}\Big)\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}} \geq \exp\Big(-\frac{K^2}{6\sigma^2}\Big)\lambda_h(-K)^{\frac{1}{6}}\tilde{a}'(-K)^{\frac{1}{2}} = \beta'_\infty(-K)$. The inequality is because the distance between $K_{1/2}$ and K is closer than between $K_{1/2}$ and $-K$.

(1*) Analogous to (1)'s arguments, 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 $KK_{1/2} > 0$, $\beta'_\infty(K) = \exp\Big(-\frac{K^2}{6\sigma^2}\Big)\lambda_h(K)^{\frac{1}{6}}\tilde{a}'(K)^{\frac{1}{2}}$ $\leq \exp\Big(-\frac{K^2}{6\sigma^2}\Big)\lambda_h(-K)^{\frac{1}{6}}\tilde{a}'(-K)^{\frac{1}{2}} = \beta'_\infty(-K)$.

(2). Differentiate Eq. (C.1) on both sides with respect to K, rearrange terms, and get

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

Meanwhile, differentiate the multiplicative inverse of both sides in the same way and get

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

The product of the 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, $\lambda_h^{\frac{1}{6}}(\tilde{a}')^{\frac{1}{2}}$ increases (decreases) iff.

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

i.e.,

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

This needs to hold iff. $a < \frac{1}{2}$ for $\lambda_h^{\frac{1}{6}}(\tilde{a}')^{\frac{1}{2}}$ to be hump-shaped in K.

(3). A sufficient condition of Eq. (C.3) ($\geq$) is for both $\frac{h''(a)}{h'(a)}$ and $\frac{h'''(a)}{h'(a)}$ to 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 Eq. (C.3) ($\geq$) holds. □