Essays on Investor Inattention and Strategic Communication

by

Jinge Liu

Business Administration
Duke University

Date: ____________________
Approved:

____
S. Viswanathan, Co-supervisor

____
Simon Gervais, Co-supervisor

____
Luis Felipe Varas Greene

____
Arjada Bardhi

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Business Administration
in the Graduate School of Duke University

2022

Abstract

Essays on Investor Inattention and Strategic Communication

by

Jinge Liu

Business Administration
Duke University

Date: ____________________
Approved:

____
S. Viswanathan, Co-supervisor

____
Simon Gervais, Co-supervisor

____
Luis Felipe Varas Greene

____
Arjada Bardhi

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Business Administration
in the Graduate School of Duke University

2022

Abstract

This dissertation comprises two chapters studying how information transmitted in the economy and the financial markets becomes compressed in communications and its consequences. In chapter one, the compression is due to limited communication bandwidth, and in chapter two, it is due to limited attention on the receiver's end.

Chapter one discusses how information intermediaries selectively present evidence to serve financial decision-makers. Faced with a space limit for their communication reports, the information intermediaries present information selectively. I solve the model for the optimal messages in the intermediaries' communication with the decision-makers and investigate the relationship between the apparent messages and the inferred economic fundamentals. The two main findings are: (i) the model matches many stylized facts about content biases such as prior or extreme biases; (ii) I derive an analytical relationship between the messages and the inferred fundamentals in the asymptotics. This relationship can be conveniently used to interpret observed content biases and quantitatively analyze the effects of the context on the interpretation of contents. The theory also shows that content biases may improve rather than decrease welfare. The model relates to empirical content analysis using frequency-based proxies and can be used to analyze contextual effects on contents.

In chapter two, I develop and analyze a theoretical model that shows how investors allocate their limited attention resources to monitor a wide selection of target firms. An investor with limited attention demands information about the types of her portfolio firms before investing. The firms strategically supply good news and withhold bad news. The investor may press companies to reveal more information by allocating more costly attention to them. Because the benefit to attention is convex, the investor will optimally focus on a subset of firms and acquire complete information while giving up learning anything about the other firms. Firms in the scrutinized subset have low investigation costs and a high Expected Value of Perfect Information (EVPI), and they always receive an efficient amount of capital. The other firms are provided with an inefficient level of capital and suffer from extreme asymmetry in information transparency. The result rationalizes convertible debt as a socially optimal financing instrument for private firms. It can be applied to a venture capital context to analyze entrepreneurial investment relationships.

Contents

List of Figures

Acknowledgements

During my academic journey at Duke, I was very fortunate to receive support from my advisors as well as so many faculty members, colleagues, and friends that it is impossible to exhaustively list all of them. I want to use this opportunity to express my gratitude.

I am deeply indebted to my dissertation advisors, S. Viswanathan and Simon Gervais, who spent great time and effort to provide me with invaluable advice and support. Vish gave me much-needed support when I switched to theoretical research months before my preliminary exam and has been my advisor since I took the exam. He is very supportive of exploring novel ideas and directions of research, and I greatly appreciate the many conversations I have had with him to discuss my research progress. Simon provided me with invaluable advice and support in developing my research projects. I greatly appreciate the many meetings we had at Fuqua and over Zoom, as they helped me to develop, write, and present the theoretical ideas that I wanted to study.

I am deeply grateful to my committee members for their advice and support for my dissertation research. I first met Felipe Varas in my second-year corporate theory class. Later, I benefited a lot from my conversations with him to deepen my understanding of game theory in finance. I attended Arjada Bardhi in her seminars on continuous-time theory and information design. I have received invaluable advice from her on my dissertation research.

I am very grateful for the help I received throughout my academic journey at Duke. Lukas Schmid was the chair of my preliminary exam committee and offered me crucial help navigating the exam. Xu Jiang, from the accounting area, was a committee member for my preliminary exam and offered me instrumental guidance when I started my research in information disclosure. Ravi Bansal provided invaluable help when I explored my interest in asset pricing. I am grateful to Lukas Schmid, Andrew Patton, and Michael Weber from Chicago Booth for their support of my Ph.D. application. Throughout my years at Duke, I served as a teaching assistant for many professors, most notably Douglas Breeden and Ming Yang. I not only learned a lot about how to design an intriguing and compelling course but also benefited greatly from our interactions in deepening my understanding of various topics of finance.

I am fortunate to have many good friends and colleagues who accompanied me during my years of Ph.D. studies. I thank Feifan Zhang, Feng Tian, Huidi Lin, Yang Zhou, Chenyu Wang, Xiaoyu Chen, Chao Yang, Wenxi Liao, Joy Tong, Jing Huang, Yenan Wang, Prabhava Upadrashta, Morad Elsaify, Hao Pang, Lin Zhao, and many more that I cannot list here.

Finally, I am deeply indebted to my parents, Haiyi Liu and Zhaohui Qiao, for their unconditional love and support. No words can express my gratitude.

Chapter 1

Content Bias and Information Compression

An information presenter faces a physical limit of information transmission. She selects and reports a given number of signal realizations from a large set to maximize the decision maker's utility. The observed content deviates from the implied substance, as the picture of selected signals looks systematically different from the full fundamental picture. Economic contexts, including prior belief, utility shapes, and payoff relevance, drive the deviation. Apparent reporting biases found in such contents, including slanting to the prior belief or extremes, can be explained by the presenter's selective coverage to elaborate on the more valuable fundamental realizations against the side of prior or extremes, effectively appearing to generate reports by recalibrating fundamentals. Such biases improve welfare. An asymptotic and analytical mapping from fundamentals to report contents is derived to clarify the interpretation of content data and facilitate their analysis. The model relates to empirical content analysis using frequency-based proxies and can be used to analyze contextual effects on contents.

1.1 Introduction

Consider the problem of an information intermediary who holds an abundant quantity of information that can improve the choices of a decision-maker. The information intermediary seeks to guide the decision-maker as better decisions today will on average lead to an increased demand for her services in the future. For example, financial newspapers or magazines that successfully steer investors towards the best investment decisions may increase subscriptions over time. The decision maker expects the information intermediary to present some information to him. The problem, in many circumstances, is that the information intermediary cannot present all the information due to the existence of some physical constraint. In the example, all reported information must fit the space of a newspaper. The information intermediary must find an optimal way to selectively present news to the decision maker up to the physical limit and to maximize the welfare improvements that her communication will prompt.

How does the information intermediary select and present information, and what will the contents look like? An instinctive answer may be that a faithful information intermediary communicates a miniature of all information. Generally, however, that is not optimal. Because different information has different value for decision making, an information intermediary who engages in material selection will use the limited space to elaborate on more valuable information, even if that means to deliberately create biases in its reports. The rational audience can infer the fundamental information from such biases, understanding that they shall not be interpreted literally and shall be properly contextualized. In short, the literary content is biased and deviates from the fundamental substance, and that deviation depends on the context.

This paper models such material selection procedure by examining how the information intermediary designs the information structure used for reporting. The theory has two aims. First, it intends to examine how content biases arise from a novel perspective. Indeed, the physical capacity of information transmission is a common yet unexplored channel that shapes contents. To financial media, for instance, such capacity can be the space on a newspaper, the room for the first three swipes of a mobile app, or time slots for broadcast. The space limit forces the sender to cover selectively and this alone generates content biases, without the need to introduce any persuasion motives of the information intermediary. Notably, the model provides a novel explanation for well-documented media biases including the tendencies to report on the prior and the extremes.

As for the second aim, the model seeks to speak to the issue of contextualization in the bourgeoning empirical content analysis literature in economics and finance. Better computational power and data accessibility prompt more financial economists to examine contents of communications directly. For example, a literature focusing on financial media includes Antweiler and Frank (2004), Tetlock (2007), Tetlock et al. (2008), and Loughran and McDonald (2011). Many pioneering studies use word-count/frequency-based quantifications of contents and reduced-form regressions. Such quantifications are believed to contain information on financial fundamentals or market beliefs and are interpreted with “tones”, “sentiment”, or other relatively vague terms. The standing issue is that an interpretable structural approach that considers economic contexts is still lacking. To fill this gap, it is necessary to develop a theory that micro-founds content generation and speaks to economics and finance specifically.

To reach these goals, I solve for the information intermediary’s optimal information structure and derive an analytical expression of the asymptotic correspondence between the total signals and the reported signals under specific contexts including prior beliefs, payoff relevance and utility curvatures. The equilibrium features the separation between the literary and the substantial meanings of a report, and can generate stylized biases such as reporting on the prior or the extremes. With the analytical expression, it is possible to dissect the count/frequency-based measures used for quantifying contents, demystifying “tones” and “sentiments” and examining what such measures really say.

The benchmark model is set up as follows. The sender has access to binary signals that are informative about the binary state for truth, and the number of signals is bigger than the physical capacity, the number that can be transmitted. The sender picks a subset that fits the capacity and presents it to the receiver with the same preferences, who will then make a bet on the state and get a payoff that is increasing in the bet's closeness to the true state. This mirrors, for instance, an investor's problem of choosing the best portfolio whose payoff depends on the economy's state, after he learns information from his subscribed financial newspaper. The problem of the sender is to design an information structure that pins down the probability of reporting a particular subset of signals conditioning on each realization of all signals. A set of total signal realizations is called a fundamental, and a set of reported signals is called a report.

How should we approach this question? First of all, the fundamentals and reports can be simplified with their sufficient statistics. If assuming that the order of the reported signals does not carry information, then with ex-ante identical signals, the fundamentals and reports can be losslessly represented by the number of signals with a particular realization, or equivalently the proportion of a particular realization. For example, if a binary signal takes value from "bullish" or "bearish", then the fundamental can be represented by the number of bullish signals, or the proportion of bullish signals; the same for the report. In fact, the order of signals usually does not carry information. For instance, the order of stories in a newspaper usually does not matter; the space allocated for good or bad stories matters. Within texts, taking sentimental words as signals and ignoring the information carried by word order is the standard assumption of the bag-of-words approach in economic textual analysis. For example, Tetlock (2007) constructs a word-count based measure of market pessimism in the study of predictive relationships between the measure and returns. Hence, the model is applicable on both the story level and the text level.

At this stage, it shall be pointed out that information compression up to a physical constraint plays a central role in this model. Essentially, the information intermediary is recommending actions by presenting a set of signals. With access to an abundant quantity of information, the sender knows the best action for any fundamental, but she cannot deliver all those recommendations because the number of total available reports is smaller than that of possible fundamentals due to the physical capacity. This feature is preserved if the sufficient statistics are used. With count-based sufficient statistics, for instance, the fundamental space basically becomes a set of natural numbers no bigger than the total number of signals, and the report space becomes a set of natural numbers no bigger than the number of transmitted signals. The fundamental space is obviously larger than the report space. The problem of the information intermediary is how to best use the scarce resource of reports to transmit the most information and maximize utility. In fact, searching for the utility-maximizing information structure is equivalent to finding one that minimizes utility losses due to compression.

What does the equilibrium information structure look like? I characterize the conditions that an equilibrium must satisfy and narrow down the search for the equilibrium. Intuitively, a faithful information intermediary will only use pure strategies because it will not infuse any unnecessary noise in the communications. (This breaks down with preference discrepancies, discussed in Section 5.) Hence, the design of the information structure is choosing the optimal way to pool fundamentals into partition sets and map the fundamentals in the same partition set to the same report. Also, the number of pools should not be smaller than the number of possible reports as, otherwise, the information intermediary is throwing away valuable vehicles to transmit information. Lastly, the information intermediary shall pool similar fundamental values together in order to minimize the loss from compression. Here, the similarity refers to the closeness of the likelihood ratios about which is the true state. In the model, this is equivalent to how much fundamentals look alike. To sum up, the optimal information structure should feature an ordered partition of the fundamental space, with each partition set mapping to a distinct report and all reports used.

Using these characterizations, I proceed to pin down the equilibrium content of communication. The analysis above is based on utility maximization but, because the optimality of the information structure only involves how the fundamental space is partitioned, it is not enough for pinning down the looks of the reported contents. Reports are labels for the partition sets, and with the same ordered partitioning, report assignments do not matter as long as the sender and the audience agree on their interpretations. In practice, however, the report assignments cannot be arbitrary. After all, agents who rely on reports to make decisions still want to know the facts and are not willing to accept a baselessly fabricated news report. Some equilibrium selection criteria must be introduced to pin down the content, and these criteria should be the audience's common expectations for information intermediaries in practice.

Two good candidate criteria include honesty and self-consistency. An information structure is honest if all communicated signals are verifiably real. An information intermediary can be honest in this sense even if it is very one-sided in material selection. A self-consistent information structure requires the literary meaning of a report to be monotonically increasing in its corresponding substantial meaning; in other words, the information intermediary cannot appear to be self-contradictory.

The first key result of the paper that follows is the pinning down of the criteria-satisfying information structure and the equilibrium look of contents. I show that an honest and self-consistent information structure always exists, and in fact self-consistency implies honesty. Because of the preference alignment assumption, honesty and self-consistency do not enter as constraints to optimization; they are merely equilibrium selection criteria. The criteria-satisfying information structure is the subject of study for all subsequent analysis of contents in their biases, quantification, and contextualization.

With this information structure, it is clear how content biases are generated. Consider the aforementioned example where the signal realizations are "bullish" or “bearish”, and use the frequency-based sufficient statistics. In equilibrium, a fundamental with 100% bullish signals will map to a report with 100% bullish signals. The same is true for bearish signals. And indeed, the extreme fundamentals look the same as the extreme reports in the sense that they contain 100% of a certain kind of signal realizations. Suppose there is no literary biases, then the frequency of bullish signals in a report shall increase proportionally with that frequency in the corresponding fundamental. Generally, however, this is not the case, because the ordered partition sets of the sufficient statistic are not equal in sizes. For the fundamentals that are more critical for decision-making, the information intermediary will elaborate on them by finely partitioning those fundamentals and map the partition sets to distinct reports. Consequently, as the fundamental sufficient statistic changes, the report changes along with it sensitively; that is, the report statistic as a function of the fundamental statistic has a high slope. For the less newsworthy fundamentals, the reports are less sensitive to changes in corresponding fundamentals, and the slope is low. Sometimes reports could even be unaffected by changes in fundamentals when those fundamentals are all pooled together. Different report sensitivities at different levels of fundamental sufficient statistics induce non-linearity in the fundamental-report correspondence, causing the wedge between the literary and substantial meanings of a report, and generating literary biases.

The intuition helps to explain the apparent biases of slanting to the prior or the extremes. Decision-makers are more interested to know about evidence contradictory to their prior beliefs, because it will more likely sway decisions and generate value. Hence, the information intermediary allocates many distinctive reports to the contradictory fundamentals and few to the prior-leaning ones. Consequently, the intermediary appears to offer the audience what they already believe. On a side note, a similar bias of reporting towards the more payoff-relevant state can be explained in an analogous way. Likewise, decision-makers are not keen on distinguishing fundamentals at the extremes, because they have small probabilities and distinguishing them does not yield very different optimal actions. The information intermediary allocates few reports to them and many in the middle fundamentals, hence appearing to be sensationalist towards the extremes. Such biases are part of the most efficient communication protocol between the information intermediary and the audience that both parties tacitly agree upon and, contradictory to the conventional wisdom, actually improve welfare rather than compromise it.

The second key result of the paper is a tractable function that predicts the content of the reports given fundamentals and contexts. To fully interpret the relationship of fundamentals, reports and contexts, I derive an analytical expression of the reports as a function of the fundamentals by taking the model to its asymptotic limit. When there are a large number of signals, each with a low level of informativeness, a properly scaled sufficient statistic for the fundamentals is distributed as a Gaussian mixture. The aforementioned criteria-satisfying optimal information structure involves an ordered partition of the fundamental space, hence characterized by cutoffs distributed on the real line. When the reported information is sophisticated enough and can contain many signals, the ordered partition becomes dense, and a discrete mapping from fundamentals to the reports becomes an asymptotic and continuous one. This continuous mapping is a tractable and powerful result that clearly points out how information intermediaries report up to physical capacity and how economic contexts affect their reporting strategy.

The analytical mapping has many uses. First, it has a clear economic interpretation. The report for a fundamental is related heavily to the perfect-information action that the fundamental would induce, properly adjusted with the curvatures of the fundamental distribution and the utility function. Second, it highlights the value of different fundamentals to decision-making. If attributing the physical capacity to the receiver or the sender, then it characterizes the audience's attention allocation or the information intermediary's reporting effort allocation. Third, it makes econometric analyses convenient. It gives a clean-cut dissection of the count/frequency-based accounts for texts into different key factors including fundamentals, prior beliefs, payoff relevance, and utility curvatures. In an example, I show how the tractable mapping can contribute to the analysis of misspecification errors in linear regressions when the researcher fails to contextualize content data properly.

The scope of model applications covers economic scenarios for which the assumptions are satisfied beyond the canonical example of financial media. Consider, for instance, a business consultant presenting her research report to a corporate manager, an intelligence officer presenting events in a briefing to a policymaker, a writer sketching an argument by presenting supporting materials, or a doctor explaining the results of physical examination to her patient. In all these examples, the purpose of information delivery is to facilitate decision-making. Furthermore, the communicator needs to present supportive information rather than directly give a blunt summary or recommendation. And like in the financial media, the physical capacity is a real concern that constrains the amount of information that can be conveyed in presentation. It takes form of the required or conventional length of an article, a report or a presentation. It is usually relatively stable and may be determined in a bigger equilibrium. Given the physical capacity, this model can be applied to analyze the relationship of fundamentals, reports, and contexts such as prior beliefs, payoff relevance and utility functions.

1.2 Literature

Media bias. The results of this paper can be applied to understand media content biases, thus contributing to the burgeoning literature of media slant studies. In the theory literature, the demand-side stories for slants focus on how media cater to the audience when their payoffs depend on the audience, including Suen (2004), Mullainathan and Shleifer (2005), and Gentzkow and Shapiro (2006). Some work along this line also features some form of inattention (Che and Mierendorff (2019), Perego and Yuksel (2020)). These models assume some particular form of preference discrepancy among the players. This paper departs from the existing literature by identifying a novel mechanism associated with information compression alone without the need to introduce any persuasive effects or market segmentation. Hence, my paper identifies a common channel, complementing the existing media bias literature.

This paper is also connected to the theoretical study of financial media. Financial media has been shown to be instrumental to investors' decision-making. Goldman et al. (2021) models financial journalists in a market setup and studies how they affect market outcomes, including the quality of announcements and prices. More broadly, this paper is connected to the literature of informed parties as information intermediaries and market manipulation (e.g., Bénabou and Laroque (1992)). Besides assuming preference alignment in the benchmark model, my paper does not rely on the availability of markets and prices. It aims to fundamentally understand what will be found in the specific content of media reports.

Empirical content analysis. Another area of application is the increasingly popular field of empirical content analysis. How to contextualize contents in a way specific to economics is a long standing question, and this paper attempts to provide an analytical answer to this question. Because of modeling resemblance, this paper is particularly closely related to the strand of research that adopts the token-count or frequency measures of texts, often treated as proxies of "tones", to describe popular belief between competing hypotheses (see survey of Gentzkow et al. (2019)). This strand of the literature includes research based on counts of tokens (e.g. Antweiler and Frank (2004), Tetlock (2007), Tetlock et al. (2008), Loughran and McDonald (2011)) and counts of articles and covered events (e.g. Gentzkow and Shapiro (2010), Baker et al. (2016)), and aimed at understanding, for example, sentiments on boom versus bust, political left versus right, or stability versus instability. These papers, however, are silent about how different contextual parameters influence the contents. This paper, featuring information compression, proposes a framework of studying contextual determinants, fundamentals, and reports together. It also derives a tractable 1-to-1 correspondence between fundamentals and report appearances, shaped by preference curvatures, payoff relevance, and prior beliefs. It can help us to better understand contents and to further improve the theoretical rigor of relevant analysis.

Communication and limited attention. This paper contributes to the literature of limited attention in the study of communication games. Limited attention is an economic interpretation of strategic data compression. A communication game with limited attention usually exhibits two features: the information is strategically compressed, and sometimes the demand side can potentially purchase more capacity. In this paper, the data transmission capacity is fixed. Either media firms or audience members cannot buy more bandwidth. Hence, although the model does not necessarily intend to attribute the channel constraint to either side, the model can be interpreted as some form of inattention if attributing the bandwidth limit to the audience's limited information-processing ability.

A popular approach to modeling attention is to borrow the information-theoretic tool of mutual information (Sims (2003)). Following this tradition, Gentzkow and Kamenica (2014) list mutual information costs as an example to solve Bayesian persuasion (Kamenica and Gentzkow (2011)). Bloedel and Segal (2020) use a disclosure setup with commitment and studies attention allocation up to mutual information costs. Mutual information can be interpreted as the change in entropy before and after learning a signal. By Shannon's source coding theorem, entropy is the lower bound of the expected length of the binary codewords for encoding the fundamental space (Shannon (1948); see Chapter 5 of Cover and Thomas (2006)). Without further structures, such shortest binary encoding is neither designed to solve games nor to be economically meaningful. Hence, while mutual information is a convenient and useful mathematical tool, its micro-foundations do not always optimally fit specific economic setups. Rather than directly calling results from the conventional information theory, this paper attempts to build from micro-foundations that fit the desired analysis directly. It proposes a new way of looking at the data compression effect in communication games.

Another contribution is to propose a method of directly modeling the content of reports, recognizing that content cannot be arbitrary nor naively truthful in practice. Because this paper assumes alignment of preferences, it does not have the time-inconsistency issue for optimal communication and hence relates to both commitment and no-commitment models. While no-commitment models may generate the wedge between the literal and substantial understandings of a signal (e.g., Gentzkow and Shapiro (2006)), many past works with commitment cannot do the same thing. That is because they abstract from explicitly modeling the looks of reports and only analyze the induced posteriors and actions (Kamenica and Gentzkow (2011), Bergemann and Morris (2019)). While such analysis works in optimization, it cannot separate content from substance. This paper fills the gap by proposing an approach of modeling content for commitment models with rational inattention, thereby contributing to the discussion of the content bias, separate from the discussion of the biases of beliefs or actions.

Hard evidence. This paper relates to the hard evidence literature that examines games where information is partially verifiable. The sender announces a set of states that contains the true state, as in communication games such as Milgrom and Roberts (1986) and Dye (1985), and mechanism design such as Green and Laffont (1986).

The solution of the games involves finding the optimal subset containing the true state for each given state. This paper uses "honesty" or partial verifiability as a candidate equilibrium selection criterion and hence is related to the category of the hard evidence games. If honesty is satisfied, then the media firm reports a subset of true fundamental signals, which can be alternatively understood as reporting the biggest subset of fundamentals that contain those reported signals. Hence, the media firm does not only in essence reports a partition set containing the true fundamental, but also in appearance reports a subset of the fundamental space containing the true fundamental, hence related to the hard evidence or partial verifiability literature.

1.3 A Baseline Model of Contents

Agents

A single media firm reports to a homogeneous decision-making audience, whose utility is contingent on the underlying binary state of nature $\theta \in \{0, 1\}$. Decision-makers have prior belief $\Pr(\theta = 1) = \pi \in (0, 1)$ and make bets $a \in [0, 1]$ whoes payoff depends on the state. The payoff on the bet is

$$u(a; \theta) = u_\theta h(1 - |a - \theta|) = \begin{cases} u_1 h(a) & \text{if } \theta = 1 \\ u_0 h(1 - a) & \text{if } \theta = 0 \end{cases} \tag{1.1}$$

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 inverse distance between the true state and the bet. Intuitively, the decision-maker is better off if his bet is closer to the true state. I make the following assumption.

Assumption 1.3.1. *$h(\cdot)$ has the following properties:*

(1) h is twice continuously differentiable, and $h'(a) > 0$ and $h''(a) < 0$ on $(0, 1)$;

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

The first-order assumption says the decision-maker will get a higher payoff if the bet is closer to the true state. The second-order assumption aims to capture the diversification benefit of betting on an interior number between 0 and 1. In fact, if there is no diversification benefit and $h''(a) \geqslant 0$, the decision-makers will simply choose to bet on either 0 or 1, and the problem becomes trivial, as will be discussed in Section 5. Hence, I consider the interesting case in this benchmark model where the decision-makers can optimally choose an interior bet. I further assume that $h(\cdot)$ ensures the optimal action to be always only in the interior. An example of such assumptions is $h'(1) = 0$.

The beliefs and preferences of the media outlet are the same as the decision-makers'. This assumption can be a consequence of the profit-driven motives of the media. To see this, consider media outlet m with prior belief π_m and preference $u_m(a;\theta) = u_{m\theta} h(1 - |a_m - \theta|)$. Assume it is purely profit-driven, and the π_m and $u_{m\theta}$'s are choices of how outlets position themselves. Suppose the profit of m is increasing in the ex-ante value of information to the audience member who pays for the news service and subscribes selectively. In that case, m can choose to align its utility function and belief to those of the audience so that m's own optimality as a constraint does not bind. Surely, it is also possible that information intermediaries are simply faithful and offer a service that maximizes audiences' utility. This specification shows a clear-cut information compression effect that produces literal biases even without resorting to the persuasion effect due to preference discrepancies.

Information and strategy

In the beginning, the media outlet receives N symmetric binary signals $s_1, ..., s_N$ from Nature. I make the following assumption regarding the distribution of signals.

Assumption 1.3.2. *Conditioning on θ, $s_1, ..., s_N$ are independent and* $\Pr(s_i = \theta|\theta) = p > 0.5$.

These signals represent the entire information about the truth that the financial media has at its disposition and can use in its news report. I call a set of realizations for $\mathbf{s} = (s_1, ..., s_N) \in \{0,1\}^N$ a "fundamental". The media firm, however, does not transmit all its information due to limited physical capacity. The measure of physical capacity is an integer $n \leqslant N$. The media outlet chooses to report n binary tokens $r_1, ..., r_n$, each supposedly taking the form of some signal realization. A news report is a set of such tokens $\mathbf{r} = (r_1, ..., r_n) \in \{0,1\}^n$. I assume that the order of the reports does not convey information; all reports with the same number of zeros and ones are equivalent signals.

The strategy of the audience is to choose the best bet $a^*(\mathbf{r})$, conditioning on the information sent by the media outlet. That action is essentially recommended to the audience by the media outlet. The audience's problem in the subgame of $\mathbf{r}$ is

$$\max_{a \in [0,1]} \mathrm{E}[u(a;\theta)|\mathbf{r}] \tag{1.2}$$

Knowing how the audience will respond to $\mathbf{r}$, the media outlet chooses the information structure $\{\sigma_{\mathbf{sr}}\}_{\mathbf{s}\in\{0,1\}^N, \mathbf{r}\in\{0,1\}^n}$, where $\sigma_{\mathbf{sr}} := \Pr(\mathbf{r}|\mathbf{s})$. The problem of the media outlet is as follows:

$$\max_{\{\sigma_{\mathbf{sr}}\}} U = \mathrm{E}[u(a^*(\mathbf{r};\theta))] \qquad \text{(Problem [P1])}$$

$$s.t. \quad \sum_{\mathbf{r}} \sigma_{\mathbf{sr}} = 1, \forall \mathbf{s} \in \{0,1\}^N$$

$$\sigma_{\mathbf{sr}} \geqslant 0, \forall \mathbf{s} \in \{0,1\}^N, \mathbf{r} \in \{0,1\}^n$$

$$\sigma_{\mathbf{sr_1}} = \sigma_{\mathbf{sr_2}}, \text{ for } \mathbf{r_1}'\mathbf{1}_{(n\times 1)} = \mathbf{r_2}'\mathbf{1}_{(n\times 1)} \qquad \text{(orderless reports)}$$

Dimension Reduction

The problem can be simplified by reducing the dimensions of the fundamental and report spaces. Since I only consider ex-ante identical signals, a sufficient statistic for a fundamental is its sum $K = \sum_{i=1}^{N} s_i$, the count for the fragmented signals with realization being 1. Thus, the fundamental space becomes $\{0, 1, ..., N\}$. The distribution of K is a binomial mixture, with $K|\theta$ following $Bi(N, p)$ if $\theta = 1$ and $Bi(N, 1-p)$ if $\theta = 0$. Meanwhile, since the reports are assumed to be without order, I can summarize any report with its sum $k = \sum_{i=1}^{n} r_i$. The report space is hence $\{0, 1, ..., n\}$. These are the "codewords" that media can use for information delivery.

Notice that all affine transformations of K and k also serve as legit variables that summarize the fundamentals and the reports. Specifically, K/N and k/n, the proportions of tokens that equal to 1 respectively in the fundamental signals and the reports, are equivalent K and k. Such affine transformations are legit because they preserve the equidistant nature of fundamental grids or report grids, hence they do not affect the analysis of biases.

The strategy of the media is to choose the information structure $\{\sigma_{Kk}\}_{0 \leqslant K \leqslant N, 0 \leqslant k \leqslant n}$, where $\sigma_{Kk} \in [0, 1]$ is the probability of producing news report k when the fundamental is K. After receiving the signals from nature, the media outlet will produce a news report k. Then the decision-maker will choose a bet $a^*(k)$ as the optimal action.

Formally, the media's problem can be rewritten as follows:

$$\max_{\{\sigma_{Kk}\}} \; U = E[u(a^*(k);\theta)] \qquad \text{(Problem [P1])}$$

$$s.t. \quad \sum_k \sigma_{Kk} = 1, \forall K \in \{0,1,...,N\}$$

$$\sigma_{Kk} \geqslant 0, \forall K \in \{0,1,...,N\}, k \in \{0,1,...,n\}$$

At this point, I do not require these codewords k to have any substance in terms of content. Hence, the report codeword space can be any set of $n+1$ distinctive labels other than $\{0,...,n\}$ and it will make no difference to the optimal value of the information structure. For now, the reports have no intrinsic meanings. Later in the paper, I will show how to link content to the labels and give them substance.

Minimizing Compression Loss

Since the media's problem involves convex optimization over a compact set, the optimal solution must exist. To locate the equilibrium, I use the necessary conditions that the equilibrium must satisfy to narrow down the search and provide characterizations.

The task for the media firm is to compress the complex data obtained from nature into a simpler report that fits the physical capacity. To achieve this, it needs to minimize the loss of information from compression and fully use the capacity. The first condition that the media outlet must meet is presented in Proposition 1. Because the preferences are aligned, the media outlet must not infuse any unnecessary noise in the reports by adopting mixed strategies in its optimal information structure. Consequently, it should only adopt pure strategies.

Proposition 1.3.3. *(Pure strategy) Under Assumption 1.3.1, [P1]'s optimal solution* $\sigma^*_{Kk} \in \{0,1\}$*,*$\forall K \in \{0,1,...,N\}$*,*$k \in \{0,1,...,n\}$*.*

The proof is in the Appendix. From here on, I will use $C(K)$ to denote the optimal codeword for K.

By Proposition 1.3.3, the optimal strategy for the media outlet is to pool the fundamentals into sets, and for each set, to produce the same report. There are two questions to be answered: (i) how many different types of reports are present in the equilibrium; (ii) which fundamentals are pooled for producing the same report. The first question is answered by Proposition 1.3.4.

Proposition 1.3.4. *(Surjection) Under Assumption 1.3.1,* $\forall k \in \{0, 1, ..., n\}$*,* $\exists K \in \{0, 1, ..., N\}$ *such that* $\sigma_{Kk} > 0$*.*

See the Appendix for proof. Intuitively, the media outlet wants to do as little pooling as possible. Because each report codeword is an opportunity for the media to convey some information, it must spare no reports. To be specific, each report codeword is a vehicle for the media to make a distinct recommendation. Since the value of information comes from making different recommendations customized to different circumstances, the media firm would like to have as many vehicles as possible. An additional vehicle is always helpful because the media outlet can at least break a group of fundamentals into two groups, one mapping to the old codeword and the other mapping to the otherwise unused one, hence strictly increasing value. Therefore, the media firm must use all codewords in equilibrium.

The second question has an intuitive answer: it is optimal to pool "similar" fundamentals together in order to minimize the information loss from compression. The proper way to measure "similarity" between fundamentals is by the closeness of their Bayes factors $\Lambda(K)$. Under the binomial structure of the fundamentals, the

Bayes factor of a fundamental $\Lambda(K)$ is

$$\Lambda(K) := \frac{\Pr(K|\theta = 1)}{\Pr(K|\theta = 0)} = (\frac{p}{1-p})^{2K-N} \tag{1.3}$$

and is strictly monotone in K. An equivalent way to describe similarity is using the optimal action $a^*(K)$ associated with fundamental K assuming that the listeners observe K perfectly. That is because $a^*(K) = \arg\max_a (1-\pi_K)u_0 h(1-a) + \pi_K u_1 h(a)$ where $\pi_K = \Pr(\theta = 1|K)$, and hence is strictly monotonic in $\Lambda(K)$. Proposition 1.3.5 explores this idea and presents a necessary condition for the optimal information structure.

Proposition 1.3.5. *(Cutoff structure) Under Assumption 1.3.1, equilibria for [P1] must feature "ordered partitions":*

(i) In an equilibrium with any distribution for $K|\theta$ with unequal Bayes factors for fundamentals, let $(K^{(j)})_{j=0}^N$ be a permutation of $(0, 1, ..., N)$ with $\Lambda(K^{(0)}) < \Lambda(K^{(1)}) < ... < \Lambda(K^{(N)})$. Then there exists optimal cutoffs of fundamentals $\{K^{(i_1)}, ..., K^{(i_n)}\}$ such that $B_0 = \{K^{(0)}, K^{(1)}, ..., K^{(i_1)}\}$, $B_1 = \{K^{(i_1)+1}, ..., K^{(i_2)}\}, ..., B_{k'} = \{K^{(i_{k'})+1}, ..., K^{(i_{k'+1})}\}, ..., B_n = \{K^{(i_n)+1}, ..., K^{(N)}\}$. And $\forall K \in B_{k'}, C(K) = k_{k'}$, where $(k_{k'})_{k'=0}^n$ is any permutation of $(0, 1, ..., n)$, i.e. $\{k_0, ..., k_n\} = \{0, ..., n\}$.

(ii) Under Assumption 1.3.2, i.e. with $K|\theta \sim Bi(N, \theta p + (1-\theta)(1-p))$, there exists optimal cutoffs of fundamentals $\{K_1^*, ..., K_n^*\}$ such that $B_0 = \{0, ..., K_1^*\}$, $B_1 = \{K_1^* + 1, ..., K_2^*\}$, ... , $B_{k'} = \{K_{k'}^* + 1, ..., K_{k'+1}^*\}$, ... , $B_n = \{K_n^* + 1, ..., N\}$, and $\forall K \in B_{k'}$, $C(K) = k_{k'}$, where $(k_{k'})_{k'=0}^n$ is any permutation of $(0, 1, ..., n)$, i.e. $\{k_0, ..., k_n\} = \{0, ..., n\}$.

The proof is in the Appendix. Proposition 1.3.5 states that, in equilibrium, the optimal pooling is to cluster adjacent K. The equilibrium is characterized by n cutoffs that separate the fundamentals into $n + 1$ ordered partition sets. The fundamentals in the same ordered partition set will map to the same distinctive report codeword. Proposition 3 also implies that there must be multiple equilibria. Because the codewords do not carry meanings, all permutations of the codewords are essentially equivalent for optimization.

To illustrate Proposition 1.3.5 more clearly, consider the following example. Consider $u(a;\theta) = u_\theta \cos(\frac{\bar{\pi}}{2}|a-\theta|)$ by setting $h(a) = \sin(\frac{\bar{\pi}}{2}a)$, where $\bar{\pi}$ is used for 3.14... rather than π. It can be shown that the objective of (2) becomes

$$\max_{\{B_i\}_{i=1}^n} U = \sum_{0 \leqslant i \leqslant n} \| \sum_{\vec{v}_K \in B_i} \vec{v}_K \| \tag{1.4}$$

where $\vec{v}_K = \big((1-\pi)u_0 C_N^K p^{N-K}(1-p)^K, \pi u_1 C_N^K p^K (1-p)^{N-K}\big) \in \mathbb{R}^2$ and the norm is Euclidean. $\{B_i\}_{i=1}^n$ is a partition of $\{0, ..., N\}$. Geometrically, each fundamental is represented by a vector. The contribution to the value function by each group of reports is represented by the Euclidean length of the vectorial sum of the included fundamental vectors. The media firm's objective is to pool vectors into groups to minimize the harm done to the total length of vectorial sums. Intuitively, the media outlet needs to pool fundamentals with similar vectorial angles in order to minimize length losses caused by vectorial summation. For $\vec{v}_K$, its angle is $a^*(K) = \arctan \frac{\pi}{1-\pi}\frac{u_1}{u_0}\Lambda(K)$, strictly monotone in $\Lambda(K)$ and henceforth K, and therefore the media firm would pool fundamentals with similar values of K. The geometric representation is shown in Figure 1.1.[1] Details of the calculation are in the Appendix.

Bridging Reports and Content

Up until this point, the "news reports" are only meaningless labels and do not possess any content. In practice, however, readers of news reports do not simply want

[1] Example 1 with $u(a;\theta) = u_\theta \cos(\frac{\bar{\pi}}{2}|a-\theta|)$, $N = 5$, $n = 3$, $\pi = p = 0.6$, $u_1 = u_0 = 1$. The blue vectors represent fundamentals, and the dotted line cut fundamentals into optimal pools.

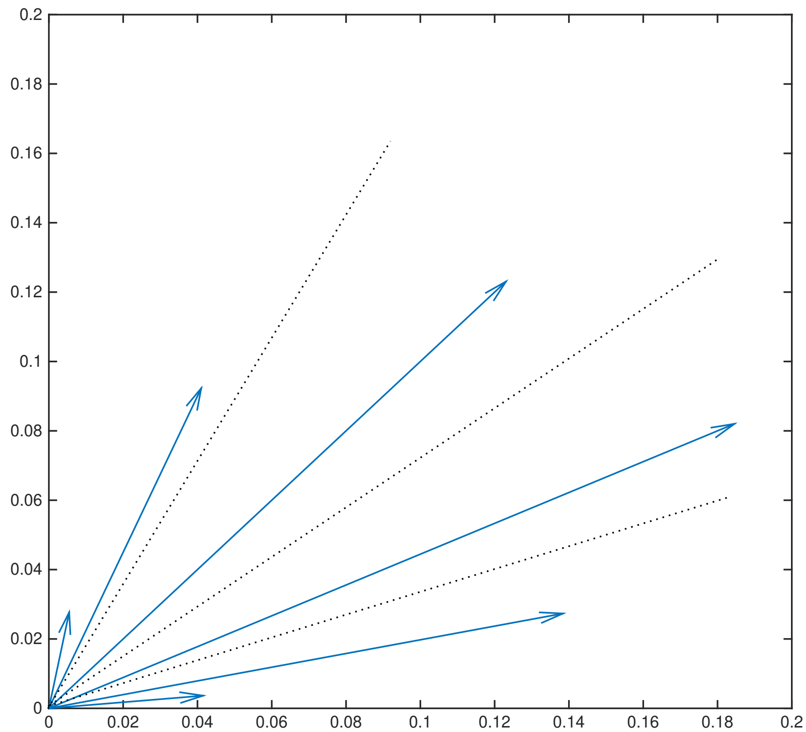

Figure 1.1: **Example 1**

blunt recommendations of actions. They are still interested in facts and rely on news to know it. They will not subscribe to a media outlet delivering reports that are obviously detached from reality. Consequently, the media will not simply publish any arbitrary symbols, hoping the audience to accept them as long as the media and readers agree on their interpretations. They must frame those reports in the form of a presentation of reality that convincingly and verifiably reflects the facts, or at least can pass some casual inspection by the audience.

To attach content to the symbols, I introduce two candidate criteria that are common in practice in determining media's trustworthiness. Notice that such criteria do not constrain optimization; they are only used for equilibrium selection.

The first criterion is honesty, or the ability to survive fact-checks, capturing readers' preference for facts. Specifically, the reported signals need to be verifiable as genuine signal realizations.

Criterion 1. (Honesty) An information structure $\{\sigma_{Kk}\}$ is *honest* if $\{r_1, r_2, ..., r_n\} \subset \{s_1, s_2, ..., s_N\}$.

And using sufficient statistics, Criterion 1 can be rewritten as $K - (N - n) \leqslant C(K) \leqslant K$. Notice that this is a broad interpretation of honesty: a report is considered honest as long as no constituent piece is fabricated. It does not require the coverage to be balanced. Suppose $N = 100$ and 50 of the signal realizations are ones and the other 50 are zeros. Suppose $n = 50$ and a media firm decides to fill all the capacity with realizations of ones. In that case, it may seem very biased but still survives the fact-checks and satisfies the honesty criterion. Therefore, this criterion is conceptually not difficult to achieve.

The second candidate criterion is self-consistency.

Criterion 2. (Self-consistency) An information structure $\{\sigma_{Kk}\}$ is *self-consistent* if $\forall\ K_1$,$K_2 = 0, 1, ..., N$ and $K_1 < K_2$ with $\sigma_{K_1k_1} > 0$ and $\sigma_{K_2k_2} > 0$, there is $k_1 \leqslant k_2$.

Readers expect positive monotonicity of the reports in the fundamentals. For instance, if the underlying facts favor $\theta = 1$ over $\theta = 0$, readers will expect the report to exhibit favor for $\theta = 1$ as well. Intuitively, the reports look like a collection of signals and have literal meanings. They also have implied fundamental meanings, represented by the corresponding set of fundamentals. Self-consistency requires positive monotonicity of literal meanings in fundamental meanings. This criterion captures the demand for "consistency" from media outlets: those with such information structures are more likely viewed as objective and trustworthy, while those without them may be seen as self-contradictory and playing games.

A natural question is whether there exists an equilibrium that satisfies these two criteria. Theorem 1.3.6 addresses this question and is also the starting point for analyzing the relationship between the fundamentals and the reports.

Theorem 1.3.6. *Under Assumption 1.3.1 and 1.3.2, there exists an equilibrium* $\{\sigma_{Kk}\}$ *of [P1] that satisfies both Criterion 1 and Criterion 2. Additionally, any equilibrium that satisfies Criterion 2 must also satisfy Criterion 1.*

Proof. Theorem 1.3.6 is an automatic consequence of Propositions 1.3.3, 1.3.4, and 1.3.5. For the proof, it is sufficient to show there exist equilibria that satisfy Criterion 2 and that these equilibria satisfy Criterion 1.

Since any equilibrium satisfies Proposition 1.3.5, then simply let the permutation in Proposition 1.3.5 be $k_0 = 0$, $k_1 = 1$, ..., $k_n = n$. Then the equilibrium satisfies Criterion 2. Such equilibria form the set of self-consistent equilibria. It is obvious that these equilibria satisfy Criterion 1, too. □

Theorem 1.3.6 shows that a self-consistent optimal information structure exists and is always honest. The media outlet may simply choose a self-consistent information structure if that is what the audience prefers. Henceforth, I focus on the self-consistent information structures and examine how they pin down the relationship between the fundamentals and the reports.

1.4 Content Bias in Strategic Contexts

According to Theorem 1.3.6, media outlets may appear to "slant" in equilibrium, but only because this is what the audience prefers. Instinctively, one may expect faithful media to produce a set of reports whose frequency of positive signals is as close as possible to that of the fundamentals rather than to produce a biased report strategically. Instead, however, Theorem 1.3.6 establishes that it is the media outlet's faithfulness that leads it to slant. With the optimal information structure, the reports will not look close to the fundamentals in general. Apparently, the media outlet is engaged in cherry-picking the facts, twisting the tones, or elaborating on some fundamentals over others in a strategic way. It will eventually produce a seemingly biased representation of the full picture. That, however, is simply the outcome of the most efficient communication protocol that the audience and the media tacitly agree upon.

The natural questions that arise after Theorem 1.3.6 are where the cutoffs are located, what is the report associated with each fundamental, and how the report distribution will deviate from the fundamental distribution. This section will address these questions. To begin with, I use the following example for an intuitive illustration of the biases.

1.4.1 An Illustrative Example for Two Types of Biases

Example 2. Consider an investor who reads a financial newspaper for information to bet on the state of the economy. The economy will be either boom ($\theta = 1$) or bust ($\theta = 0$). The action variable $a \in (0, 1)$ is bounded. The bigger a is, the more stocks the investor would like to buy, and without loss of generality, normalize the short sale constraint to 0 and budget constraint to 1, as they are the lower and upper bounds of a respectively. The news outlet has access to $N = 5$ informative stories and will choose $n = 3$ to publish. The stories are either positive ($s_i = 1$) or negative ($s_i = 0$) on the health of the economy. To satisfy its customer, the media firm will make the information that it communicates as informative as possible. Depending on the investor's prior beliefs and payoff relevance for different states of the economy, the communication arrangement will be different. Next, I show examples for different biases with different parameters.

First, consider the bias of herding on the prior or the more payoff-relevant state, as is exhibited in Figure 1.2 and 1.3. The prior for a booming economy is strong ($\pi = 0.9$). Therefore, if the signals are not too pessimistic, the investor will invest heavily in stocks regardless and will not greatly value precise information. Since there are only four distinctive codewords, the media firm feels no need to assign too many of them for relatively bullish fundamentals. However, if the signals are against the prior, they become valuable because they carry information for re-optimizing. Hence it is better to assign distinctive codewords for each of the bearish fundamentals.

The most notable observation from the optimal information structure is that the reports will seem upward biased. For instance, if the journalists find 1 positive story and 4 negative stories, they will report the positive story and only 3 of the 4 negative stories. The proportion of positive news in the media report will rise to 33% from 20%, meaning the content sentiment will seem more bullish than what the fundamental would indicate. Furthermore, if the journalists find no more than 2 negative stories, they will hide every one of them. The newspaper will be full of bullish stories (100%) compared to 60%, 80% or 100% of bullish stories in the fundamentals. This information structure will transform any distribution on the fundamentals to a more upward-tilted distribution on the content, hence generating an apparent herding-on-the-prior bias.

Analogously, one can also illustrate the bias of reporting heavily on the more

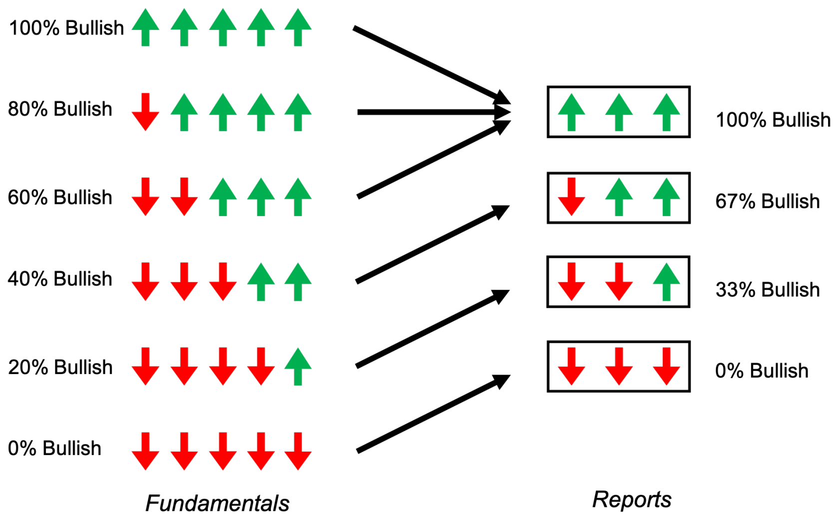

Figure 1.2: Example 2, Information Structure for Prior Bias

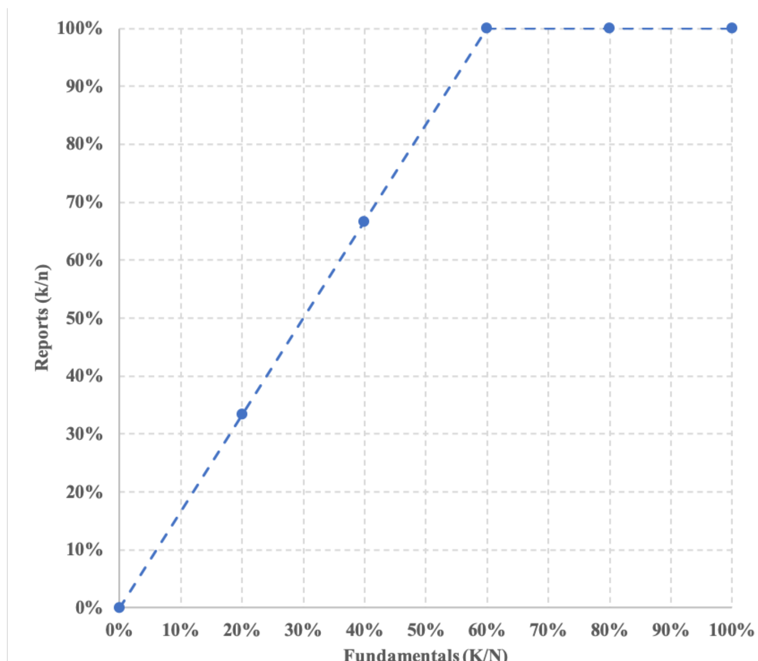

Figure 1.3: Example 2, Report Curve for Prior Bias

payoff-relevant state. Suppose that, in the investment opportunities accessible to the investor, the "boom" opportunities, or the assets that pay off when $\theta = 1$, are more profitable than the "bust" opportunities (that pay off when $\theta = 0$). Effectively, the case is analogous to when the prior belief for $\theta = 1$ is strong, because $\pi/(1-\pi)$

and u_1/u_0 function similarly in the formulas. The bias mechanism is identical to the slant-to-the-prior bias.

Second, consider the extreme bias. Figure 1.4 and 1.5 presents the optimal information structure under a setup where there is no strong prior ($\pi = 0.6$) for exhibiting an obvious herding-on-the-prior phenomenon. For example, consider the case when journalists find 2 positive stories and 3 negative ones, or 40% bullish fundamentals. In that event, they will report 1 positive story and 2 negative stories, and therefore the bullish rate in the reports will be down to 33%. The same argument for slanting to the boom can be applied to the fundamental for the 60% bullish fundamental. The effect is more evident in the extremes, since if the journalists find no more than a story of any sentiment, they will throw it away and report 100% bullish or bearish. This information structure will transform any distribution on the fundamentals to a more heavily tailed distribution on the content, hence generating an apparent extreme bias.

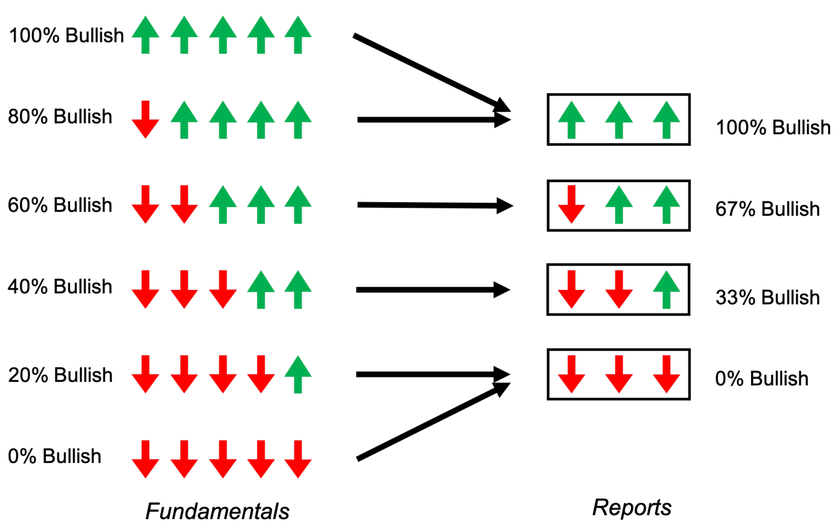

Figure 1.4: Example 2, Information Structure for Extreme Bias

To sum up, the mechanism generating the apparent bias is "elaboration and recalibration". There are two observations to consider. First, the adjacent report values

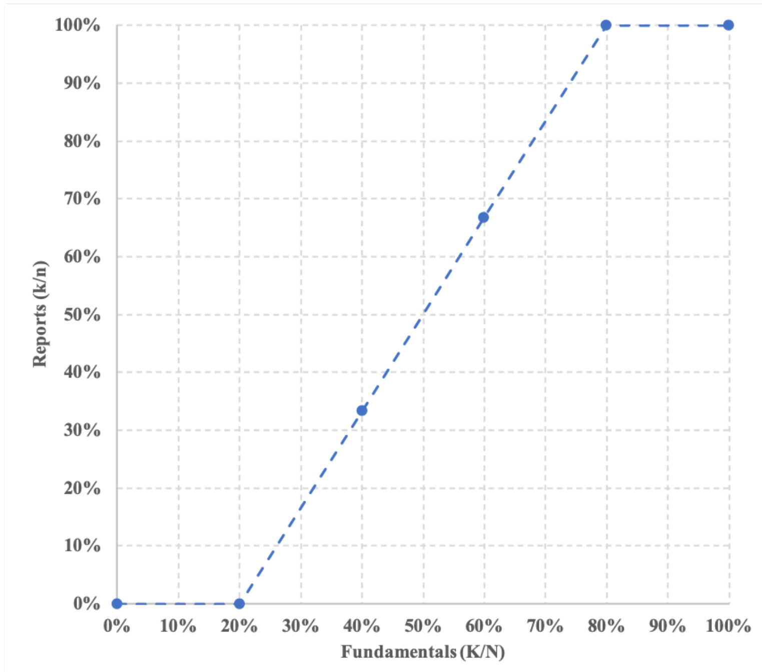

Figure 1.5: **Example 2, Report Curve for Extreme Bias**

are equidistant due to the equal size for each reported signal. Similarly, adjacent fundamental values are also equidistant. Second, the extreme reports and fundamentals look alike, comprising 0 or 100% of certain type of signals. With these observations, it is clear how apparent biases are generated. If the media firm chooses to elaborate around some fundamentals by using finer cutoffs, then the reports will change sensitively with respect to the fundamentals. In other places, the fundamentals are squeezed to pool into fewer partition sets, and hence the reports are less sensitive. Because the extreme fundamentals and reports are calibrated to look the same, the different sensitivities between two extremes lead to a recalibration between the fundamentals and the reports, hence generating apparent biases and a wedge between content and substance.

1.4.2 An Asymptotic Result: Biases and Attention Allocation

The discrete setting does not yield convenient analytical solutions, making comparative statics essentially impossible to derive. For tractability, I take the model to its limit and look at the asymptotic case where the distributions are continuous. It must be stressed that, theoretically speaking, the discrete setting discussed in the previous section is sufficient for any analysis of biases. Following Theorem 1.3.6, one can exhaustively search all ordered partitions for the optimal information structure and discuss its effects. The continuous case to be discussed should be understood as an asymptotic approximation of the discrete case.

To prepare for the asymptotic analysis, I introduce a set of assumptions that are analogous to Assumption 1.3.1:

Assumption 1.4.1. *$h(\cdot)$ has the following properties:*
(1) h is fourth continuously differentiable, and $h'(a) > 0$, $h''(a) < 0$ on $(0, 1)$;
(2) regularity conditions hold: for $a \in (0, 1)$, $|h^{(4)}(a)| < +\infty$, $|h'''(a)h'(1 - a) + h'(a)h'''(a)| < +\infty$.

Compared to Assumption 1.3.1, I add more regularity conditions for differentiability at Assumption 1.4.1(1) and get rid of the interior requirement of Assumption 1.3.1(2). That is because, with the distribution assumption that is introduced later, the fundamentals with a first-best action recommendation at 1 or 0 can be simply pooled together and mapped to the extreme report. It will become clear as I proceed to the model solutions for this part. The newly added Assumption 1.4.1(2) is a technical condition that helps with the analysis and is easily satisfied by common functions.

Now, let each binary fundamental signal s_i take values from $\{\frac{\sigma}{\sqrt{N}}, -\frac{\sigma}{\sqrt{N}}\}$ instead of $\{1, 0\}$, where $\sigma > 0$ is a parameter. For a fixed N, I am simply reframing the game using an affine transformation. The new information environment is completely equivalent to the previous one. The sum $K = \sum_{i=1}^{N} s_i$ is still the sufficient statistic, although it is no longer nonnegative. Let $\mu > 0$ be a parameter and let $\Pr(s_i = \frac{\sigma}{\sqrt{N}}|\theta) = \frac{1}{2}(1 + \frac{\mu_\theta}{\sigma}\frac{1}{\sqrt{N}})$, where $\mu_\theta = \mu$ when $\theta = 1$ and $\mu_\theta = -\mu$ when $\theta = 0$. Define $p_N = \Pr(s_i = \frac{\sigma}{\sqrt{N}}|\theta)$. By standard arguments of in-fill asymptotics,

$$F_{K|\theta=1}(x) \Rightarrow \Phi(\frac{x-\mu}{\sigma}), F_{K|\theta=0}(x) \Rightarrow \Phi(\frac{x+\mu}{\sigma}) \tag{1.5}$$

as $N \to \infty$, where Φ is the standard Gaussian cumulative distribution function and $\Rightarrow$ stands for convergence in law.

For each fixed pair of N and n, the structure of the problem is the same as that in the previous section, and Theorem 1.3.6 applies. One can solve the optimal information structure given this discrete problem. The question is what will be the limit of the information structure if N and n converge to infinity. Specifically, I consider the case where I first take N to infinity, and then take n to infinity. Intuitively, this means the attention capacity is always small compared to the nuanced fundamentals. This simplifies the problem in that, when choosing the optimal ordered partition, the media outlet only needs to find optimal cutoffs for the Gaussian-mixture fundamentals, thus avoiding the complexity brought by the discreteness of the fundamental space. Formally, the analogue to Assumption 1.3.2 is as follows:

Assumption 1.4.2. $F_{K|\theta}(x) = \Phi(\frac{x+(1-2\theta)\mu}{\sigma})$

Because the information structure is characterized by cutoffs, it is equivalently characterized by a distribution of cutoffs. Specifically, let $\kappa^*(n) = \{K_1^*, ..., K_n^*\}$ be the set of cutoffs on $\mathbb{R}$ such that the ordered partition is $\{(-\infty, K_1^*), [K_1^*, K_2^*), ...$,

$(K_n^*, +\infty)\}$. Then, following Theorem 1.3.6, the main problem [P1] becomes

$$\max_{\substack{\{-\infty=K_0^*<K_1^*<\ldots\\<K_n^*<K_{n+1}^*=+\infty\}}} U = \sum_{i=0}^{n} \Big\{ \pi u_1 \left(F_{K|\theta=1}(K_{i+1}^*) - F_{K|\theta=1}(K_i^*) \right) h(a_i^*) + \ldots$$

$$+ (1-\pi) u_0 \left(F_{K|\theta=0}(K_{i+1}^*) - F_{K|\theta=0}(K_i^*) \right) h(1-a_i^*) \Big\} \qquad \text{(Problem [P1'])}$$

s.t.

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

And then define

$$\beta_n(K) = \frac{1}{n} \int_{K' \leqslant K} d\mathbb{1}_{\{K' \in \kappa^*(n)\}}$$

which can fully characterize the equilibrium of [P1'] at n. At this point, it is worth mentioning that the solutions to [P1'] also satisfies Criteria 1 and 2. In other words, the media outlet is honest and only reports a subset of what it receives; it is simply engaged in material selection. The media outlet is obviously self-consistent, if assigning the report i to the partition set (K_i^*, K_{i+1}^*). Because I care about analytical results in asymptotics, the question is to solve for $\beta_\infty(K) := \lim_{n\to\infty} \beta_n(K)$.

The meanings of the function $\beta_\infty(K)$ and its derivative (with respective to K) $\beta'_\infty(K)$ are worthy of some discussion. First, $\beta_\infty(K)$ is more than the limiting cumulative frequencies for the cutoffs; it is also the limiting mapping from the fundamental K to the reports β_∞. That is because for the discrete case, the report for K is its corresponding k, which is equivalent to the scaled k/n or simply $\beta_n(K; F_K)$. Hence, I refer to $\beta_\infty(K)$ as the *report curve*. It tells us what the limiting report is for a fundamental. It is also a transformation that helps to pin down the probability distribution of the reports, given the distribution of the fundamentals.

Second, should $\beta_\infty(K)$ be absolutely continuous, its differentiation $\beta'_\infty(K)$ stands for the "density" of the cutoffs and is an equivalent characterization of the cutoff distribution. There are two ways of looking at this. From a journalist's point of view, the bigger $\beta'_\infty(K)$ is around some K, the more cutoffs the media would like to insert around that K, meaning the media outlet optimally elaborates on that fundamental more. As different fundamentals have different worth for coverage, $\beta'_\infty(K)$ is a good measure of the limiting newsworthiness of the fundamentals, and hence $\beta'_\infty(K)$ can be referred to as the *newsworthiness curve*.

Another way of understanding $\beta'_\infty(K)$ is that it represents the optimal attention allocation of the audience. Attention allocation is a traditional topic in rational inattention research. In this model, the attention resource is the number of available codewords or partition sets, or one plus the number of cutoffs. Denser cutoffs around a fundamental means the audience allocating more scarce attention resource to that fundamental. Therefore, the newsworthiness curve is also the *attention allocation curve*, hence connecting to the attention allocation literature (e.g., Bloedel and Segal (2020)).

There is a strong relationship between the limiting report function $\beta_\infty(K)$ and the first-best optimal action, an analogy to the Bayes factors in the discrete case. To capture this relationship, it is necessary to introduce this first-best action function $\tilde{a}(K)$ $(K \in \mathbb{R})$ as an auxiliary function, defined by

$$\frac{\pi u_1 F'_{K|\theta=1}(K)}{(1-\pi)u_0 F'_{K|\theta=0}(K)} = \frac{h'(1-\tilde{a})}{h'(\tilde{a})} \tag{1.6}$$

when $\tilde{a} \in (0,1)$, and in this setting $F_1(K) = \Phi(\frac{K-\mu}{\sigma})$ and $F_{K|\theta=0}(K) = \Phi(\frac{K+\mu}{\sigma})$. Intuitively, if assuming that the fundamentals are absolutely-continuously distributed, and that the audience directly observes K, then $\tilde{a}(K)$ is the optimal action they should take. It is also the optimal action that the media outlet wants to recommend if the fundamental is K. The definitional formula is simply the first order condition for that optimization. If under $F_{K|\theta}$, $\tilde{a}$ is strictly monotone in K on some interval, then it has an inverse function $\tilde{K}(a) := \tilde{a}^{-1}(a)$ when it is well defined.

Notice, however, that this definition applies to fundamentals with interior optimal actions only. For very big or very small tail fundamentals, the first-best actions may be corner solutions. Specifically, for K such that

$$\frac{h'(1)}{h'(0)} \geqslant \frac{\pi u_1}{(1-\pi)u_0}\frac{f_1(K)}{f_0(K)} = \frac{\pi u_1}{(1-\pi)u_0}\exp(\frac{2\mu K}{\sigma^2}) \tag{1.7}$$

The optimal action recommendation is 0, hence define $\tilde{a}(K) = 0$. If they exist, they should be mapped to the same report. By Criteria 1 and 2, the report should obviously be the minimal report, and in this setup it is 0. Analogously, for K such that

$$\frac{h'(0)}{h'(1)} \leqslant \frac{\pi u_1}{(1-\pi)u_0}\frac{f_1(K)}{f_0(K)} = \frac{\pi u_1}{(1-\pi)u_0}\exp(\frac{2\mu K}{\sigma^2}) \tag{1.8}$$

they should be pooled together with the corresponding report being 1, if they exist at all. Hence, define $\tilde{a}(K) = 1$. Therefore, for such possibly existent fundamentals at the extremes, the report curve is trivial. It is only necessary to pin down the report curve for the less radical fundamentals.

Theorem 1.4.3 below is the main result that links the limit distribution for cutoffs and the auxiliary function for a general $F_{K|\theta}$.

Theorem 1.4.3. *(Asymptotic attention allocation and report curve)*
(1) Assume Assumption 1.4.1. Let the absolutely continuous $F_{K|\theta}$ ($\theta = 0, 1$) have bounded densities with monotone likelihood ratio, so that $\tilde{a}(K)$ is absolutely continuous, $\tilde{a}'(K) < +\infty$, strictly monotone increasing on a real interval $I \subset \mathbb{R}$, and ensures $\tilde{a}(K) \in (0,1) \Rightarrow K \in I$. Then the derivative of $\beta_\infty(K) = \lim_{n\to\infty}\beta_n(K)$ satisfies

$$\beta'_\infty(K) \propto (\lambda_h(K))^{\frac{1}{6}}(\lambda_F(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}} \tag{1.9}$$

for $K \in I$, where

$$\lambda_h(K) = \underbrace{-\left(h'(\tilde{a}(K))h''(1-\tilde{a}(K)) + h'(1-\tilde{a}(K))h''(\tilde{a}(K))\right)}_{\text{curvature of utility function}} \tag{1.10}$$

$$\lambda_F(K) = \underbrace{|F'_{K|\theta=0}(K)F''_{K|\theta=1}(K) - F'_{K|\theta=1}(K)F''_{K|\theta=0}(K)|}_{\text{curvature of fundamental distribution}} \tag{1.11}$$

(2). Assume Assumption 1.4.1 and 1.4.2. Then set $F_{K|\theta} = \Phi(\frac{K+(1-2\theta)\mu}{\sigma})$ and therefore

$$\lambda_F(K) \propto \exp(-\frac{K^2}{\sigma^2}) \tag{1.12}$$

Theorem 1.4.3 presents a sharp dissection of the cutoff distribution (or the report curve, or the attention allocation curve) into three parts: the first-best action function $\tilde{a}'(K)$, the preference curvature adjustment $\lambda_h(K)$, and the distributional curvature adjustment $\lambda_F(K)$. The proof of Theorem 1.4.3 is in the Appendix.

Theorem 1.4.3 is the key result that characterizes the locations of the optimal cutoffs and the look of the optimal information structure. Notice that Theorem 1.4.3 is not a revelation result. Actually, with a continuum of codewords and a continuum of fundamentals, full revelation of all information is easily guaranteed in the limit and can be achieved by any 1-to-1 correspondence between K and β_∞. The real question is which 1-to-1 correspondence will be achieved in the limit. Theorem 1.4.3 answers that question strongly by pinning down a uniquely correct correspondence, which turns out to link the appearance of the reports directly to the action recommendations and curvatures of utility and distribution functions.

To study the shape of β_∞, it is necessary to examine respectively $\tilde{a}$, λ_h and λ_F. First zoom in on $\tilde{a}(K)$. The obvious takeaway from Theorem 1.4.3 is that the optimal reports are closely related to the first-best optimal recommendations. Remark 1 presents some properties of $\tilde{a}(K)$:

Remark 1. Under Assumption 1.4.1 and 1.4.2, we have the following results.

(1) $\tilde{a}'(K)$ satisfies

$$\tilde{a}'(K) = \left\{ \frac{2\mu}{\sigma^2} \frac{1}{-\left(\frac{h''(1-\tilde{a})}{h'(1-\tilde{a})} + \frac{h''(\tilde{a})}{h'(\tilde{a})}\right)} \right\}(K), \text{ for } K \in I \tag{1.13}$$

(2). $\tilde{a}'(K)$ is symmetric around $K^* := \tilde{K}(\frac{1}{2})$, i.e. $\forall \delta$, $\tilde{a}'(K^* + \delta) = \tilde{a}'(K^* - \delta)$ and $\tilde{a}(K^* + \delta) + \tilde{a}(K^* - \delta) = 1$.

The proof is in the Appendix. There are two takeaways. First, $\tilde{a}'(K)$ is biased compared to the fundamentals. Notice that K^* is the fundamental value that equalizes posterior beliefs for the two truths. It is on the opposite side to prior or utility. While the fundamentals are centered at 0, $\tilde{a}'(K)$ is centered at K^*. Second, $\tilde{a}'(K)$ is positive yet small at both tails, and leads to less elaborate reporting and more pooling at both infinities. That contributes to the extreme bias.

Next examine the two adjustment factors $\lambda_h(K)$ and $\lambda_F(K)$. They respectively capture the effects of the curvatures of preference and fundamental distribution. Under the Gaussian-mixture assumption, $(\lambda_F(K))^{\frac{1}{6}}$ is the Gaussian density of $N(0, 3\sigma^2)$. Hence, it is only necessary to examine the shape of $\lambda_h(K)$.

Remark 2. Under Assumption 1.4.1 and 1.4.2, $\lambda_h(K)$ is symmetric around $K^* := \tilde{K}(\frac{1}{2})$.

The proof is in the Appendix. The question of interest is what affects the shape of the report curve and what biases can be generated. A classic reporting bias is herding on the prior or utility, and in the model it is represented by a report curve $\beta_\infty(K)$ tending to map fundamentals to reports leaning towards the prior or the utility side. Or equivalently, it is represented by $\beta'_\infty(K)$ tilted towards fundamentals on the opposite side of prior or utility (or by the higher newsworthiness of the opposite side or the higher level of attention allocated to fundamentals on the opposite side).

It is obvious that the interval I is centered at K^*, which is on the opposite side of prior or utility. That contributes to the bias, but it alone is not enough to fully picture it. In the following proposition, I give a sufficient condition for generating the bias.

Proposition 1.4.4. *(A sufficient condition for prior/utility bias)*
(1) the newsworthiness curve "leans to the prior or utility", or

$$\beta'_\infty(K) \geqslant \beta'_\infty(-K),\ \forall K \in I \ s.t.\ KK^* > 0 \tag{1.14}$$

if $(\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}$ *is weakly increasing for* $K < K^*$ *and decreasing for* $K > K^*$.
(2) $(\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}$ *is weakly increasing for* $K < K^*$ *and decreasing for* $K > K^*$ *if and only if*

$$3\frac{d}{da}\left(\ln(h'(a)h'(1-a))\right)+2\frac{d}{da}\left(\ln(-h''(a)h'(1-a)-h''(1-a)h'(a))\right) > 0 \text{ for } a < \frac{1}{2} \tag{1.15}$$

The proof is in the Appendix. Proposition 1.4.4 argues that, a sufficient condition for generating the prior- or utility-leaning bias is to ensure the bell shape of $(\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}$. It also gives an equivalent characterization of the condition. Notice that, by Theorem 1.4.3 and Remark 1, the monotonicity of $\beta'_\infty(K)$ is fully determined by the shape of the utility function. That is because the shape of $\beta'_\infty(K)$ is determined by $(\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}$, which can be fully represented by a function of $\tilde{a}(K)$ without standalone K. Since $\tilde{a}(K)$ is strictly monotone, the monotonicity of $\beta'_\infty(K)$ is hence only determined by the shape of h. Another takeaway is that the prior or utility leaning bias is not guaranteed. In fact, specific curvature assumptions are required to ensure the presence of the bias.

Another popular bias is the extreme bias. Because the model features a bounded support for optimal reports, the monotone increasing mapping from fundamentals to reports easily guarantees fundamentals to be mapped to reports that look more extreme.

Next, directly characterize the distributions of the reports. It is summarized in the next proposition.

Proposition 1.4.5. *(The report distribution) Denote the optimal report as α. Then*

$$\begin{cases} \Pr(\alpha = 0|\theta) = \Phi\left(\frac{\tilde{K}(0)-(2\theta-1)\mu}{\sigma}\right) & \text{if } \alpha = 0 \\ \alpha|\theta \sim \Phi\left(\frac{\tilde{K}(\alpha)-(2\theta-1)\mu}{\sigma}\right) & \text{if } \alpha \in (0,1) \\ \Pr(\alpha = 1|\theta) = 1 - \Phi\left(\frac{\tilde{K}(1)-(2\theta-1)\mu}{\sigma}\right) & \text{if } \alpha = 1 \end{cases} \tag{1.16}$$

where Φ is the cumulative distribution function for standard Gaussian distributions.

To intuitively understand the shapes of the newsworthiness and report curves, consider the following examples. They all feature the conditions proposed by Proposition 1.4.5. In fact, $\lambda_h(K)$ is constant across all these examples, which simplifies the analysis.

Example 3.1. Assume Assumption 1.4.1 and 1.4.2. Following Example 1, consider cosine difference utility $u(a;\theta) = u_\theta \cos(\frac{\bar{\pi}}{2}|a-\theta|)$, where with slight abuse of notation let $\bar{\pi}$ denote 3.14... rather than π. Then $h(a) = \sin(\frac{\bar{\pi}}{2}a)$ and $\lambda_h(K)$ is a constant. Hence

$$\tilde{a}(K) = \frac{2}{\bar{\pi}} \arctan\left(\frac{u_1\pi}{u_0(1-\pi)} \exp(\frac{2\mu K}{\sigma^2})\right) \tag{1.17}$$

and therefore

$$\beta'_\infty(K) \propto \exp(-\frac{K^2}{6\sigma^2})\sqrt{\tilde{a}'(K)} = \exp(-\frac{K^2}{6\sigma^2})\left(\frac{4\mu}{\bar{\pi}\sigma^2}\frac{\frac{u_1\pi}{u_0(1-\pi)}\exp(\frac{2\mu K}{\sigma^2})}{1+(\frac{u_1\pi}{u_0(1-\pi)}\exp(\frac{2\mu K}{\sigma^2}))^2}\right)^{\frac{1}{2}} \tag{1.18}$$

Example 3.2. Assume Assumption 1.4.1 and 1.4.2. Consider quadratic preference $u(a;\theta) = u_\theta(A(a-\theta)^2 + B)$ $(A < 0)$ and $h(a) = A(1-a)^2 + B$. Just like in Example 3.1., $\lambda_h(K)$ is constant. Hence $\tilde{a}(K)$ is a logistic function:

$$\tilde{a}(K) = \frac{\frac{u_1\pi}{u_0(1-\pi)}\exp(\frac{2\mu K}{\sigma^2})}{1 + \frac{u_1\pi}{u_0(1-\pi)}\exp(\frac{2\mu K}{\sigma^2})} \tag{1.19}$$

and

$$\beta'_\infty(K) \propto \exp(-\frac{K^2}{6\sigma^2})\sqrt{a'(K)} = \exp(-\frac{K^2}{6\sigma^2})\frac{(\frac{u_1\pi}{u_0(1-\pi)})^{\frac{1}{2}}\exp(\frac{\mu K}{\sigma^2})}{1 + \frac{u_1\pi}{u_0(1-\pi)}\exp(\frac{2\mu K}{\sigma^2})} \tag{1.20}$$

Example 3.3. Assume Assumption 1.4.1 and 1.4.2. Consider exponential preference $h(a) = C_0 - C_1\exp(-Aa)$, where $C_0 \in \mathbb{R}$ and $C_1, A > 0$. Just like in Example 3.1., $\lambda_h(K)$ is constant. And

$$\tilde{a}(K) = \begin{cases} 1 & \text{for } K \geqslant -\frac{\sigma^2}{2\mu}\ln(\frac{\pi u_1}{(1-\pi)u_0}) + \frac{A\sigma^2}{2\mu} \\ \frac{\mu}{A\sigma^2}K + \frac{1}{2} + \frac{1}{2A}\ln(\frac{\pi u_1}{(1-\pi)u_0}) & \text{otherwise} \\ 0 & \text{for } K < -\frac{\sigma^2}{2\mu}\ln(\frac{\pi u_1}{(1-\pi)u_0}) - \frac{A\sigma^2}{2\mu} \end{cases} \tag{1.21}$$

Hence, for the report curve, let $\beta_\infty(K) = 1$ for big K and $\beta_\infty(K) = 0$ for small K, and in the middle the newsworthiness curve $\beta'_\infty(K)$ is a truncated Gaussian density, i.e.

$$\beta'_\infty(K) \propto \begin{cases} \exp(-\frac{K^2}{6\sigma^2}) & \text{for } -\frac{\sigma^2}{2\mu}\ln(\frac{\pi u_1}{(1-\pi)u_0}) - \frac{A\sigma^2}{2\mu} < K < -\frac{\sigma^2}{2\mu}\ln(\frac{\pi u_1}{(1-\pi)u_0}) + \frac{A\sigma^2}{2\mu} \\ 0 & \text{otherwise} \end{cases} \tag{1.22}$$

1.5 Welfare Analysis

Consider again the baseline discrete model $[P1]$. Because it is assumed that the payoff discrepancies do not exist, the loss of welfare in the game comes solely from the fact that the number of different action recommendations N is bigger than the number of different reports n and there are not enough vehicles for lossless communication. It is only possible to study the welfare effects under the discrete setting, because in the continuous setting there is no welfare loss in the limit due to 1-to-1 correspondence between the fundamentals and reports. The proper definition of welfare is the value of $[P1]$ since both parties share this preference.

Intuitively, more information guarantees higher value, because with better information the media can make better recommendations. Also, more available codewords guarantee higher value, because that reduces pooling loss. The result is summarized in the next proposition. The proof is in the Appendix.

Proposition 1.5.1. *Under Assumption 1.3.1 and 1.3.2, $U^*(N, n)$ is strictly increasing in N and n.*

The discussion of welfare effects for media communications often focuses on how biases may reduce welfare. For instance, Gentzkow and Shapiro (2006) models media with additional reputational concerns and argues that eliminating bias may increase total welfare, because it decreases garbling and thus increases listeners' willingness to buy the reports. Mullainathan and Shleifer (2005) argues that profit-driven media cater to listeners' biases, and whether listeners may see the full picture depends on how heterogeneous their biases are. Also, in terms of media ethics, it seems to be widely accepted that media should try to produce unbiased projections of reality, rather than to game its reports or to deliberately create biases. Such reasoning often rests on the prerequisite that the biases are a consequence of agency problems, and in order to serve selfish motives, the media outlet will hurt the listeners' interests.

However, I argue that biases may be a natural consequence of information compression, and may be present even if the media is fully extending the audience's interests. The welfare implications are exactly the opposite to the traditional wisdom, because the bias itself results from welfare maximization. In particular, strategic reporting is strictly better than the behavioral strategy of randomly choosing n out of N and trying to project the unbiased reality:

Remark 3. Random reporting is strictly inferior to strategic reporting when $N > n$.

Proof. The value of the former is equal to $U^*(n, n)$, strictly smaller than $U^*(N, n)$ by Proposition 5. □

1.6 Applications

Applicable types of contents. This model is useful in analyzing contents comprising of presented signal pieces fit into a certain space. It can be applied to study a collection of stories, taking each story as a signal in the fundamentals. For example, a business newspaper as a collection of stories of the state of the economy can be the subject of study. Prime time news broadcast or online news presentation can fit in analogously. Beyond news media, examples may include a dossier of evidence files that try to make a point for a legal case with a page limit. In such reports, the order of the presentation in the reports usually does not matter. Selective inclusion of presented materials creates apparent biases, as discussed in this paper.

Within a text, the model can also be applicable. The relatively obvious applications are forming a briefing or writing an argument. In a briefing, the sender needs to find the appropriate set of events to serve the decision-maker. When making an argument, the writer needs to not only make her view clear but also find the right combination of supporting and hedging materials for argumentation.

Less obvious is the application to analyze wording. Unlike the examples above, texts as collections of words are ordered vectors. They are high-dimensional data with inter-dependencies within. To apply the model to analyze wording, one must assume the inter-dependencies do not matter, taking the text as a bag of words. That, in fact, is the take on texts by many content analysts and economists. Because computational methods taking into account high-dimensional features are often data-mining and hard to interpret economically, bag-of-words assumptions about texts and word frequency as content proxies are often adopted when an interpretable analysis is desired. For example, in the sentiment analysis by Tetlock (2007), words are categorized as having either optimistic or pessimistic sentiment, and their counts are used to represent contents for certain articles in the *Wall Street Journal*. If the high-dimensional features of contents is ignored, then the writer of an article can be viewed as putting together the correctly-proportioned collection of linguistic tokens in competing categories to deliver information about competing hypotheses.

Applicable settings in practice. The framework of this paper can be applied to study contents where the assumptions of the model apply. To be specific, the contents studied need to be vehicles for transmission of information to decision-makers who want to learn about a state variable. The content should feature presentation of fragmented pieces of information, such as pieces of evidence of a trial, stories in a news cycle, or aspects for or against an argument. For such contents, the framework can generate common apparent biases observed in everyday life. Using this model, one can identify the look of biases and its contextual determinants, describe the reporting and attention allocation strategies, and interpret the report distributions.

An important example is media bias. It is a widely held perception that media coverage is often biased, and this model helps to understand how biases may arise. The framework includes a faithful media outlet, a given physical capacity of information transmission, and two competing hypotheses that the listeners need to evaluate and base their decisions on. For instance, consider the coverage of the safety trials of a new drug for emergency use. The two hypotheses are "safe" and "unsafe", and the decision is whether to take the medicine or not. Depending on the public prior and payoff relevance, the media outlet may selectively cover positive or negative pieces of evidence to serve decision making. Using this framework, one can make sense of apparent selective coverage and biases found in news reports.

The application in the analysis of financial and business media deserves special mention. While previous theories on financial communication biases are usually supply-side ones with manipulation by media's informants, this model provides a sharp demand-side story that focus on the audience. In practice, news outlets such as CNBC, Bloomberg News or WSJ News are popular among investors and business executives, serving to facilitate decision-making. There is an obvious action recommendation flavor in such news. With the model, the two competing hypotheses may be the boom and bust of the economy, the rise and fall of prices, or high and low projected profitability of a listed firm. The action may be the amount of investment. The smooth preferences and different payoff relevance parameters may represent economic payoffs. Hence, this framework may offer a perspective to understand the coverage strategies for economic, financial or corporate news.

Although popular opinion often criticizes media biases, this paper urges one to think harder about the cause behind apparent selective coverage, tone twisting, and media biases before making judgments. While media biases may be undesirable under other demand-side mechanisms with agency problems or market segmentation or under supply-side mechanisms of informant manipulation, biases do not necessarily hurt welfare in the setup of this paper. Biases can be used to create value for decision-making. Hence, although it is justified to be concerned about many potentially negative effects of media biases, it is also sensible that sometimes one can relax and be less worried about the news biases misleading the audience and hurting welfare.

Beyond media applications, this model also provides a framework to understand a broad variety of contents. For example, the intelligence briefings to a policy maker also presents the features of the model setup: a fixed physical length, selection of the most newsworthy from a sea of raw candidate materials, and a sender faithfully facilitating decision making. Business consultancy and market research reports are similar examples. Other fitting situations in daily life include doctors explaining the physical examination results to a patient, or parents teaching their children information about the world by presenting evidence. Furthermore, even under circumstances where the sender is not faithful, the information compression effect is also present along with other manipulative effects.

Empirical implications. This framework has particular value to the empirical research of content analysis, pioneered by textual analysis. While textual analysis covers a broad range of topics (Gentzkow et al. 2019), this framework is particularly fitting to facilitating the strand of research that uses the counts of tokens, articles or covered events as the proxies for the content, and tries to dig information from these content about beliefs, sentiments, or fundamental information that often involve two alternative hypotheses. Usually, this line of works follow these typical steps: first, it constructs a count or frequency-based proxy for the content; second, it interprets the proxy itself or the residual of a regression model that involves the proxy; and third, it uses these variables in further statistical analysis, such as regressions. These proxies can be the explanatory variable or the explained variable, depending on the research question.

In this paper, under simplified assumptions that all constituent fragmented signals are identically distributed and take the same amount of physical space to report, the reports are treated as the total count of the fragmented signals that favor one hypotheses over the other. Or equivalently, they are treated as the frequency of the fragments favoring one side of the hypotheses if using scaling, as in the analysis of the limiting case. That draws a parallel comparison to the empirical variable constructed with counts and frequencies of a bag of linguistic tokens that appear in libraries of, for example, different sentimental categories, or other empirical constructions using counts and frequencies of sided articles or event coverage.

A standing question is how to contextualize these empirical content, which requires theoretical backing of these empirical constructions. This framework proposes an analytical answer. The strategic contexts are summarized in the parameters of preference curvatures, payoff relevance, the prior belief, and in the discrete case, the amount of attention. It is understood analytically how the mapping between fundamentals and reports is shaped by these contextual parameters. The asymptotic expression, in particular, proposes a convenient tool for empirical contextualization.

I suggest two uses of the framework. The first one is to interpret content frequency measures. In previous literature, researchers have some rough idea about the observed contents being some reflections of market beliefs or market fundamentals, and tend to use relatively vague terms such as tones, sentiment, etc. to describe the content measures. What information is embedded in those measures, however, is not perfectly clear. This model provides a sharp dissection of the content measure into different constituents that include the fundamentals, beliefs, and preferences. That helps to understand more thoroughly the content data that one is dealing with.

The second suggested use is its application in empirical analysis. Past literature directly use content measures in regressions. Without a micro-founded theory to dissect content measures, however, it is unclear what is the economic meaning of such content measure and what the statistical analysis really means. This paper suggests that, instead of directly use the content measure as a feature, researchers may consider generating content-implied information with the result of Theorem 1.4.3, and using that desired information to perform statistical analysis. Using proper structural setups, the information that can be backed out from the content measures include the implied fundamentals, beliefs, and preference parameters. This process is referred to as "contextualization". It helps to eliminate the biases due to information compression, and to recover the relevant information that is understood by the players in that specific strategic environment.

Remark I use a simple OLS model to illustrate the consequence of the model-free approach to contents without adjusting for contexts structurally. Suppose the correct data-generating process is

$$y_i = \eta_0 K_i + \boldsymbol{\eta}^T \mathbf{m_i} + \varepsilon_i \tag{1.23}$$

where $(x_i, K_i, \mathbf{m_i}, \epsilon)$ are i.i.d., K_i is the fundamental variable, and $\mathbf{m_i}$ (n-by-1) is a set of control variables including a constant. It is assumed that $\mathrm{E}[\epsilon_i | K_i, \mathbf{m_i}] = 0$. Now, without contextualization, a researcher takes the content $\beta_\infty(K)$ for the fundamental variable and uses it in place of K, regressing y_i on $\beta_\infty(K_i)$ and $\mathbf{m_i}$ and causing a specification error in the functional form. Using the OLS estimator, the pseudo-truth of convergence systematically deviates from the true value, hurting consistency and asymptotic unbiasedness. Specifically,

$$\hat{\eta}_0^{OLS} \rightarrow^p \eta_{pseudo} = \eta_0 \left(I_1 + I_2 \right) \tag{1.24}$$

where

$$I_1 = \frac{\mathrm{E}[\beta_\infty(K_i) K_i]}{\mathrm{E}[\beta_\infty(K_i)^2]}$$

$$I_2 = \frac{\mathrm{E}[\beta_\infty(K_i) K_i] \mathrm{E}[\beta_\infty(K_i) \mathbf{m_i}^T] D^{-1} \mathrm{E}[\beta_\infty(K_i) \mathbf{m_i}]}{\mathrm{E}[\beta_\infty(K_i)^2]^2} - \frac{\mathrm{E}[\beta_\infty(K_i) \mathbf{m_i}^T] D^{-1} \mathrm{E}[K_i \mathbf{m_i}]}{\mathrm{E}[\beta_\infty(K_i)^2]}$$

and

$$D_{n \times n} = \mathrm{E}[\mathbf{m_i} \mathbf{m_i}^T] - \mathrm{E}[\beta_\infty(K_i) \mathbf{m_i}] \mathrm{E}[\beta_\infty(K_i) \mathbf{m_i}^T] / \mathrm{E}[\beta_\infty(K_i)^2]$$

The asymptotic bias has two components. The first component measures the effect of the shape distortion. The second component involve the interactions between the misspecification and other control variables.

1.7 Model Extensions

In this section, I discuss two extensions of the benchmark model: (1) what if the preferences of the media and the audience are not aligned; and (2). an extension of the model to study ratings on a discrete scale, such as star ratings found on Amazon or IMDb.

1.7.1 Persuasion and Physical Capacity

In the benchmark model, I made two assumptions: the auxiliary function $h(x)$ is concave, and the preferences of all players are aligned. The assumption of preference alignment is used for isolating the compression effect, and is also reasonably descriptive of the reality in various settings. However, that assumption may not always hold, and this section studies the potential persuasion effects using a simplified setup.

Assume alternatively that the auxiliary function $h(x)$ is linear (or convex). Under preference alignment, the problem is trivial. Since the optimal action is either 0 or 1, only 2 report codewords are necessary for making action recommendationss, while there are $n+1$ available. Notice that each fundamental can be categorized as corresponding to an action of either 0 or 1. As long as under an information structure, the fundamentals of the same action category are pooled together, the information structure will transmit all information perfectly. The optimal information structure always exists and can be both honest and self-consistent. An obvious construct is for all fundamentals with optimal $a = 1$ map to their highest honest report possible ($C(K) = \max\{n, K - (N - n)\}$) and those with optimal $a = 0$ map to their lowest honest report possible ($C(K) = \min\{0, K\}$).

Under such binary action space settings, I discuss the persuasive effects when the payoff relevance parameters of the sender and the receiver differ. Specifically, let $u = \frac{u_1}{u_0}$ be a parameter for the audience member and $u_m = \frac{u_{1m}}{u_{0m}}$ for the media.

Then the media firm and the audience member have different ideas in mind which fundamentals shall map to an action of 1 or 0. Without loss of generality, assume that $u < u_m$, or that the media firm has more interests vested in the state $\theta = 1$. The listener wants to choose $a = 1$ for $K \geqslant \bar{K}$ with $\Pr(\theta|K) \geqslant \frac{1}{u+1}$, while the media firm wants the listener to choose $a = 1$ for $K \geqslant \underline{K}$ with $\Pr(\theta|K) \geqslant \frac{1}{u_m+1}$. Obviously $\bar{K} \geqslant \underline{K}$, and the tension exists if $\bar{K} \geqslant \underline{K}$. That tension lies in the fundamentals $\underline{K}, \underline{K}+1, ..., \bar{K}-1$.

I further make an assumption of commitment to the information structure. Essentially, this is a problem of Bayesian persuasion under limited physical capacity. The commitment assumption did not matter when discussing the benchmark model earlier, because with preference alignment the media firm does not have the incentive to shift its reporting decisions after realization. When there is preference discrepancy, however, the assumption for commitment to information structure starts to carry weight.

First, I examine the unconstrained optimal information structure. Those fundamentals $K < \underline{K}$ can pool together and are not relevant to the analysis. How can the tension fundamentals be treated properly? Essentially, the media would like to persuade the audience to choose $a = 1$ by increasing the inferred probability of $\theta = 1$ upon the audience observing a corresponding report. To achieve that, the media must pool the fundamentals in $[\underline{K}, \bar{K})$ with the fundamentals in $[\bar{K}, N]$ which lean to $\theta = 1$, so that when the audience member see the pools, he believes the probability of $\theta = 1$ is $\frac{1}{u+1} + \varepsilon$, just enough to support a bet of $a = 1$. The fundamentals $K \in [\bar{K}, N]$ with $\Pr(\theta = 1|K) > \frac{1}{1+u}$ are the "resources" of persuasive power that the media firm can utilize to confuse with $K \in [\underline{K}, \bar{K})$ and cause the $a = 1$ outcome.

The critical question is what is the optimal way to match the two categories of fundamentals together for pooling. Intuitively, the more that a fundamental $K \in [\bar{K}, N]$

leans towards $\theta = 1$, the stronger power it has in bringing up people's beliefs about $\theta = 1$ for a pool. Specifically, including one unit of probability of that fundamental in some pool will improve the pool's received perception of $\theta = 1$ by the most. On the other hand, the more that a fundamental $K \in [\underline{K}, \bar{K})$ already leans towards $\theta = 1$, the easier it is for that fundamental to be brought up in a pool that justifies $a = 1$. Hence, the most economic way of pooling is to match the strongest $\theta = 1$ leaning fundamental down from N with the easiest-to-sway fundamental down from $\bar{K} - 1$, until each pool is with $\Pr(\theta = 1|k) = \frac{1}{u+1}$ or either pool of probabilities is exhausted.

Specifically, if N and n are fairly big and assume that the discretization residual at the end of the algorithm can be ignored, then the procedure to find an equilibrium is as follows:

Step 0. Pool all fundamentals $[0, \underline{K})$ together (and map to $k = 0$, for instance); on another note let $K = \bar{K} - 1$. Let $K' = N$. Let $\delta = 1$. Initialize $s = 0$.

Step 1. Now find the pool k for K.

Denote $P_1 = u_1\pi \Pr(K|\theta = 1)$ and $P_0 = u_0(1 - \pi) \Pr(K|\theta = 0)$.

Step 2. Is $\delta = 1$?

⇒ Yes: let $v_1 = u_1\pi \Pr(K'|\theta = 1)$ and $v_0 = u_0(1 - \pi) \Pr(K'|\theta = 0)$.

⇒ No: let $v_1 = (1 - \delta)u_1\pi \Pr(K'|\theta = 1)$ and $v_0 = (1 - \delta)u_0(1 - \pi) \Pr(K'|\theta = 0)$.

Step 3. Is $P_1 + v_1 \geqslant P_0 + v_0$?

⇒ Yes: then solve for $\delta \in (0, 1 - s]$ such that $P_1 + \delta v_1 = P_0 + \delta v_0$ and update δ's value, and the pool k for K involves $\Pr(k|K') = \delta$. Let $s \leftarrow s + \delta$.

If $\delta = 1$, then $K' \leftarrow K' - 1$ and $s \leftarrow 0$. Otherwise no action for this line.

Let $K \leftarrow K - 1$. Repeat Step 1 if $K \geqslant \underline{K}$ or break to the end.

⇒ No: let $\delta \leftarrow 1$, $P_1 \leftarrow P_1 + v_1$ and $P_0 \leftarrow P_0 + v_0$. Let $K' \leftarrow K' - 1$. Repeat Step 2 if $K' \geqslant \bar{K}$ or break to the end.

There are three major differences between the information structures with or without persuasion effects. The first difference is the involvement of mixed strategies. It is possible that δ in the algorithm to be smaller than 1. It is in the interest of the media firm to confuse the audience when necessary. Second, honesty or self-consistency generally do not hold in equilibrium. They will become constraints, rather than equilibrium selection criteria. Third, the posterior distribution looks different. If assuming persuasion effects, then the posterior beliefs are equal and exactly enough to support the action preferred by the media firm, when the audience member observes a report for the fundamentals on the side preferred by the media.

1.7.2 Modeling Product Ratings

The benchmark model can be divided into two parts. The first part is to use dimension reduction and transform the problem to the strategic pooling of $N + 1$ fundamental values to $n + 1$ equidistant report codewords. The second part is to solve that pooling problem. Proposition 1.3.5(i) and Theorems 1.3.6 and 1.4.3(i) can deal with the general pooling problem, including but not limited to the transformed strategic signal selection problem.

Consider, for example, that reviews for a certain product on a shopping website are on a scale of 1 to 5 stars. Suppose the product can be a good or a bad type. A good product can generate a better customer experience, which is a random variable distributed with a higher mean; a bad product with a lower mean. The action involved is a purchasing decision; a higher a stands for higher probability of purchase, or a higher amount of purchase. The customers review the product faithfully to help later shoppers and all customers align in preferences and prior beliefs. The star-rating process can be understood as an "elaboration and recalibration" practice. Specifically, the reviewers of the product choose the information structure that maps the observed fundamental, or the customer experience, to equidistant report codewords, or star-ratings. The apparently observed star-rating distribution needs to be interpreted based on the strategic context, as the model indicates.

Assume that the Bayes factors are monotone. In other words, the customer satisfaction distributions conditioning on good and bad product qualities satisfy the monotone likelihood ratio property. The higher the customer satisfaction is, the more likely that the quality of the product is inferred to be high. According to Proposition 1.3.5(i), the optimal information structure features the proper ordered partitioning of the fundamental space (into 5+1=6 sections) and the pooling of fundamentals in the same partition set. Theorem 1.3.6 points out the common-sense equilibrium is to map the lowest partition set to 0 star, second lowest to 1 star, until the highest partition set to 5 stars. An asymptotic approximation of the mapping is featured in Theorem 1.4.3(i).

Therefore, this model helps us to understand the way people submit ratings and the observed empirical rating distributions. In practice, it is commonly observed that the distribution of the stars look tilted. For example, too many products on Amazon have many 5-star or 4-star ratings, and relatively few 1-3 star ratings. This model provides a potential explanation. Because the customers who search for the product have purchasing motives to begin with, they derive more utility from purchasing a good product over a bad one, than from not purchasing a bad product over missing a good one. Specifically, $u_1 > u_0$. Hence the customer may need more elaborate information on the opposite side of the payoff relevance, the fundamentals that affect values the most. With only a 5-star grid available, the faithful reviewers will choose to differentiate the opposite side of the payoff relevance with distinct star-rating assignments, leaving the relatively usable products of little value for differentiation with the remaining extremely good star-ratings. Effectively, this generates an apparent bias of slanting to the payoff relevance.

A similar argument can be sketched with prior beliefs. Generally speaking, there can be a perception for product quality of listings on a certain platform. Products listed on a credible platform, for instance, may be viewed as having a certified up-to-standard quality. Hence, the informative news is contradictory evidence, and the faithful reviewers elaborate on such evidence. Effectively, this apparently induces herding-on-the-prior phenomenon with abnormally many good reviews.

1.8 Conclusion

This paper identifies the channel how media's optimal information compression creates apparent reporting biases. Such reporting strategy actually improves welfare compared to the naive randomized strategy. Without the need to assume misaligned preferences or beliefs among the media outlet and people in the audience, the model generates two biases: herding on the prior or utility and slanting to the extremes. That is by using the fact that the media must allocate its limited capacity to reporting on the fundamentals where the audience's optimal action is the most sensitive. The advantage of the model is that it uses relatively few assumptions and generates clear-cut results. It can be applied to more general settings beyond news media where an information intermediary compresses information up to a physical constraint for communication.

Moreover, this paper analytically derives the mapping from the fundamentals to the reports in an asymptotic setting. I wish to highlight its empirical implication to the content analysis literature. While existent literature sometimes adopts token-frequency-based measures verbatim in their empirical analysis, I suggest it may be a good idea for one to decode the frequency-based measures with the derived mapping in this paper, recover the fundamentals that prompted the content generation, and use these fundamentals in the empirical analysis. This is because contents are strategically generated under particular contexts, which include the listener's beliefs, payoff relevance, and attention level, as well as signal strengths. One need to take proper account of these contexts in order to capture the content's message accurately and avoid unnecessary systematic measurement errors.

This model of information compression and transmission of course does not capture all aspects of the issue. For instance, it does not consider an arbitrary combination of joint distributions and physical sizes for fundamental signals. Consequently, reports in the model are assumed to only contain presentation of equally sizable and informative fragmented details, and do not contain other elements, such as their summary statistics. These extensions awaits future research. Nonetheless, this paper aims at pointing out a common channel that gives clear explanations to how apparent content biases arise, and proposing a means to take contexts into account when performing economic analysis of contents.

Chapter 2

Investor Inattention, Information, and Firm Investment

2.1 Introduction

The efficient flow of information from entrepreneurs to investors is essential for efficient capital allocation and price discovery in the financial market. Managers know more than investors do, and it is conventional wisdom that entrepreneurs may withhold bad news and supply good news for personal gains. This leads to asymmetry in uncertainty resolution and in turn to investment inefficiencies. In this context, investors can effectively demand more information from entrepreneurs by allocating more attention resources to their firms by investigating them more thoroughly. This is due to the fact that investor attention makes it more difficult for the entrepreneur to withhold information. However, investors have limited attention and cannot thoroughly investigate all the firms at all times; hence they need to determine the best attention allocation across the various investments in their portfolio. In equilibrium, the endogenous flow of information between the firms and investors results from the interaction between the demand and the supply, strategically determined by both sides of the market.

Such a game is especially descriptive for a situation where the portfolio companies are not strongly bound to information release and when the capital is provided by one or few investors. In particular, when the portfolio companies are private, such as start-ups or private equities, the flow of information can be described as endogenous voluntary disclosure up to investor attention. The investor's attention strategy, investment strategy, and firms' disclosure strategy interact in equilibrium.

Hence, the game can be applied to study, for instance, the information and investment strategies of venture capital investing in start-ups, or of pension funds, sovereign funds, or mutual funds investing in private equities. In all such proposed settings, the private portfolio companies have incentives to keep the beliefs of the investor high in order to maintain high levels of capital inputs. Start-ups tell brighter stories to attract capital. PE managers withhold bad news to maintain LP's interest. The investor has a monitoring budget and cannot possibly invest as much of it in every firm as she would like to were she unconstrained.

This paper proposes a model to study attention allocation and endogenous information flow, combining both demand and supply of strategic information. Heterogeneous entrepreneurs receive all of their capital from an investor and inject it into their production technologies. The output from an enterprise is split between the investor and the entrepreneur by a predetermined contract. The entrepreneur receives his own productivity, and following Dye (1985), can choose to either disclose that information honestly or withhold it. In contrast to Dye (1985), the entrepreneur always knows his own productivity. However, if he decides to transmit that type to the investor, the investor receives the communication only with a probability π that is determined by the attention she has dedicated to the firm. That is, the investor does not necessarily become aware of the firm's disclosures. Hence, when the investor does not observe any communication, she is unsure whether it was withheld or lost, thus preventing the full revelation of types.

The firm-specific probability π that the disclosures will reach the investor corresponds to the attention level that the investor pays to the firm. It be increased at a cost to capture the idea that paying more attention to a firm inevitably means that some other firms will receive less of it. In this way, attention serves to summarize the demand side of information acquisition and describes the desire of the investor to know more about a firm. It can be interpreted as the pressure put on the firm to communicate. With a high level of attention, the investor tends to believe the firm is of a low type when observing no disclosure. Because the optimal level of investment increases with the firm type, the investor will punish the firm with a low level of investment upon observing no disclosure. This, in turn, incentivizes the firm to disclose more. Essentially, it is related to resources put into investigating the firm or into monitoring its disclosure. The attention cost is higher when, for instance, the investor is unfamiliar with the industry or more frictions preclude effective investor engagement.

The firm receiving attention π will choose a cutoff strategy and disclose if and only if the type is higher than the cutoff. The cutoff decreases with π, and hence the disclosure set of types expands when the allocated attention is greater. Intuitively, an increase in investigation resources leads to a higher level of revelation and a more transparent information environment for a given firm. Like Dye (1985), for a given π, one may observe asymmetries in investment efficiency and uncertainty resolution between good and bad types. It will be clear later, however, that such asymmetries do not provide a complete picture of the cross-section because they fail to incorporate the endogenous, and potentially heterogeneous, amount of attention that each firm receives.

The investor's attention is limited, and so she needs to optimally allocate different amounts of it to different firms. With a continuum of firms, I first show that the investor will choose a bang-bang solution: the only two possible equilibrium values of attention are $\pi = 100\%$ for some firms and $\pi = 0$ for others. In other words, the investor chooses a subset of companies to investigate fully and completely gives up on monitoring the rest. The investigated subset includes firms yielding the highest ex-ante returns to attention. That return is determined positively by the expected utility improvement of learning perfectly about the firm or, as is usually defined in the decision theory, the Expected Value of Perfect Information (EVPI), and negatively by the attention costs. The bang-bang solution is an important result, as it says that the investor will prefer to concentrate her attention rather than spread it. She should determine the concentrated subset by EVPI and ease to investigate, rather than other often readily available metrics such as levels of firm quality, size, or value.

The optimal allocation is concentrated because the benefit of attention paid to any given firm is convex in the attention input. In the Blackwell sense, the benefit of greater attention is to allow the investor to make more suitable investment decisions based on circumstances. Specifically, that benefit is threefold: first, the investor is less likely to miss above-cutoff news announcements; second, the cutoff is lower and more types are revealed; investments for those types can be more efficient; third, the investment without observing disclosures is more efficient. The second effect is zero due to the way cutoffs are set: around the strategically picked cutoff, the levels of investment under perfect or imperfect information are the same, and hence revealing more types marginally does not affect value. The third effect is zero because of the envelope theorem. The first effect remains and is stronger for higher levels of π. When she is more attentive, the investor believes the type to be worse upon observing no disclosure and so will invest less. Hence, missing good disclosures is a graver mistake, resulting in the aforementioned convexity and bang-bang solution.

In the model's equilibrium, the investor pays more attention to firms that are cheaper to investigate. Although it is generally difficult to study comparative statics of EVPI, it can be shown under some assumptions how EVPI changes with the location and scale of the productivity distribution. Under certain assumptions, greater uncertainty in productivity leads to higher EVPI, meaning the investor prioritizes the scrutiny of risky firms. Location of the productivity also affects EVPI, and the direction depends on parameters. The investor is not guaranteed to be interested in firms with better productivity distributions; she can be interested in worse ones if her payoff is more sensitive in the region of low types.

A large theoretical literature has contrasted the information flow that endogenously emanate from firms with different prospects or shocks, contrasting them in terms of uncertainty resolution, announcement clustering, and investment efficiency. However, most of these models of information disclosure (e.g., Dye (1985), Archarya, DeMarzo, and Kremer 2011) take the information supplier as the only determinant of new information reaching financial markets. Taking the demand-side fr information disclosures as exogeneous, good firms perform better than bad firms in the sense that their information shocks are absorbed in a more timely manner, thereby accelerating the resolution of risks and improving economic efficiency.

Such predictions break down when the demand for information disclosures is jointly considered. The firms that are more closely monitored will be more transparent and more efficiently priced in financial markets; one cannot observe any asymmetry in transparency or efficiency between better and worse firms. On the other hand, firms that are under less scrutiny by investors will vary significantly in terms of their transparency and investment efficiency. Importantly, the set of firms that receive the most attention does not depend on firm quality. Hence, better information transparency and efficiency cannot imply higher firm quality. An opaque firm could be on the relatively lower end of an ex-ante higher but more certain productivity distribution, still much better than a transparent firm with a shallow type for a risky or low productivity distribution. One cannot predict a monotone increasing relationship between firm quality and a good information environment.

I further analyze the optimal contract design that split the surplus between the investor and entrepreneurs. Because the model does not use many assumptions, I can solve for the equilibrium contract design with a general set of feasible contracts. Conventional contract design interprets the payoff as a function of the productivity or outputs, with a fixed amount of capital. I study the contract of splitting of the output as a function of both capital inputs and the productivity, hence designing a predetermined profile of contracts indexed by capital inputs. In other words, I focus on the contracts adopted in the norms and are predetermined before investments are made or information is exchanged. That is because the forms of contracts in a particular setting, e.g., for VC financing entrepreneurs, or an LP pays GPs, are relatively stable and expected. The predetermination of the contracts allows for the interaction between attention, disclosure, and investment strategies.

In particular, I focus on the socially efficient contracts, or the contracts that the social planner will choose to induce the maximum total surplus. All firms are assumed to be ex-ante homogeneous. In equilibrium, I show that the firms competing for attention and get it use potentially different contracts from the firms competing for inattention and get it. The analysis also rationalizes the use of convertible debt in the financing of start-ups by venture capitalists.

This model can serve as a workhorse model with demand-supply interactions that analyze the information and investment strategies in various settings. First of all, it can be applied to study the world of venture capital and the financing of start-ups. Suppose a venture capitalist "sprays and prays" by investing some money in a wide range of start-ups, gaining access to knowing the priors, production functions, and attention costs. Then the investor must decide how much to invest in future rounds, and the decision depends on the learned type of the start-ups. In this case, the monitoring process resembles what the model describes. The model also applies to setups where an LP seeks to learn about the quality of GPs, or when the headquarter of an organization tries to monitor its subsidiaries.

The rest of the paper is organized as follows. Section 2 presents the benchmark model. I show the equilibrium attention allocation, disclosure, and investment strategies, along with robustness checks and comparative statics. That includes the analysis of the cross-section of information shocks in 2.4. Section 3 discusses the optimal socially efficient contract. Section 4 discusses applications of the benchmark model in the different settings of private investments. Section 5 concludes.

2.1.1 Related Literature

This paper combines the demand and supply of strategic communication models between investors and portfolio companies. There are two perspectives in explaining the injection of endogenous information into the market. Demand-side models argue that new information shock into the economy is due to the investors' acquisition decisions. On the other hand, the supply-side models explain information flows with the strategic information release of firms. A typical demand-side perspective is investor inattention. This paper relates to the rational inattention (Sims (2003)) or inattentiveness (Reis 2006) models in portfolio management and asset pricing in terms of using attention to capture demand.

On the supply side, the models explain information shocks with the corporate release of new information. In accounting, economics, and finance, this topic is sometimes discussed in the literature of voluntary information disclosure. Dye (1985) outlines how an informed agent discloses or withholds private information when faced with a receiver that assumes the worst upon observing no disclosure. In the setup of Dye (1985), there is some probability that the sender does not have private information, and that prevents complete revelation of the sender's type in equilibrium. My model assumes that the senders are always privately informed, shutting down the Dye (1985) channel. Meanwhile, it is the endogenous probability of the inattentive receiver missing disclosure that prevents full revelation. This setup also contrasts other attempts to combine the voluntary disclosure literature with inattention, including the behavioral model of Hirshleifer et al. (2004) and the information-theoretic model of Bertomeu et al. (2020). Notably, this paper also explicitly models the production of the firms, thus contributing to the study of the real effects of disclosures.

More broadly, this paper fits into the discussion of disclosure and persuasion with limited attention. Using a commitment approach, Gentzkow and Kamenica (2014) and Bloedel and Segal (2020) investigate Bayesian persuasion with inattention. Using Dye information, my model is a model of verifiable information release with inattention. My specification of attention is different from Sims (2003), which has been popular in the literature. Other models without verifiability include Di Pei (2015) that discusses cheap talk with limited attention.

This model also contributes to the optimal enforcement literature. There is an idea that, when the monitoring resource is scarce, it could be socially optimal to concentrate monitoring on a publicized subgroup of individuals rather than spread it out evenly and thinly. Lazear (2006) and Eeckhout et al. (2010) both assume an exogenous distribution for non-compliance payoffs and relate the concentration strategy to the curvature of the distribution. In this respect, my model uses a principal-agent setup and provides an alternative perspective why the concentration of monitoring to a subgroup is always optimal for a voluntary disclosure game.

2.2 A Model of Disclosure and Attention Allocation

2.2.1 Setup

Agents. One investor chooses from many target firms to form a portfolio. The types of firms are indexed by i. The production function of the firm is $Y_i(K_i, A_i)$, the productivity is $A_i(\omega)$ And its prior belief is F_i. There are firms of measure $g(i)$ with the same production function and the prior belief. The total measure of all companies is ψ, or $\int_i g(i)di = \psi$. To give an intuitive interpretation, think of the investor as a venture capitalist and firms as start-ups. Or take the investor as a large LP and firms as many PEs in its portfolio.

The investor is endowed with initial capital W_0 and has access to an alternative borrowing and lending technology with a risk-free gross return of R_0. The production of firm i, $Y_i(K_i, A_i)$, satisfies $\frac{\partial Y_i}{\partial K_i} > 0$ and $\frac{\partial Y_i}{\partial A_i} > 0$. The investor provides all investment K_i to penniless entrepreneurs. In practice, funding partners provide the most capital, and co-investment is usually comparatively very small in many investment settings of PE/VC. Y_i can represent a general class of firm investments, including real investments where Y_i is the output or financial investments where Y_i is exit cash flows. I assume that the marginal product of investment is smaller than R_0 when the amount of investment is big. In other words, there exists a $\bar{K}_i$ for company i, such that for all $K_i > \bar{K}_i$, $\frac{\partial Y_i}{\partial K_i} < R_0$.

The investor and company i will share the output by an exogenous contract. The investor will get $\xi_i(K_i, A_i) \geqslant 0$, and the company will retain payoff $u_i(K_i, A_i) := Y_i(K_i, A_i) - \xi_i(K_i, A_i) \geqslant 0$. Both payoffs ξ_i and u_i are assumed to be weakly increasing in K and A. An implicit assumption is that both A and K are contractible. That is reasonable in the model because both K and Y are observable, and one can simply back out A. In the benchmark model, such contracts are taken as given. Further, I assume that $\xi_i(K_i, A_i)$ is concave in K_i, for any given A_i. In other words, for any weight $w \in (0,1)$, $\xi_i\left(wK_i^{(1)} + (1-w)K_i^{(0)}, A_i\right) \leqslant w\xi_i\left(K_i^{(1)}, A_i\right) + (1 - w)\xi_i\left(K_i^{(0)}, A_i\right)$.

Another key assumption is the complementarity between investment and productivity. I assume that for all i, $\xi_i(K_i, A_i)$ has increasing differences. That ensures the investor would like to invest more in a better project (Milgrom and Shannon 1994), as is common in practice. In addition, since the entrepreneurs do not contribute their own money, such investment inputs are the size of the firms. A_i also implicitly describes the deserving size of a firm.

The investor forms a portfolio of all the firms and the alternative lending, and wants to maximize total payoff

$$\int_i \mathrm{E}^{(f_i)}[\xi_i(K_i, A_i)]g(i)di + R_0\left(W_0 - \int_i \mathrm{E}^{(f_i)}[\xi_i(K_i, A_i)]g(i)di\right)$$

Let $v_i(K_i, A_i) := \xi_i(K_i, A_i) - R_0 K_i$ denote the excess return for investing in a company, and then the objective of the investor can be written as $\int_i \mathrm{E}^{(f_i)}[v_i(K_i, A_i)]g(i)di$. It requires the investor to put an appropriate amount of money into firm i, according to its productivity A_i. The preference of the firms, on the other hand, is to maximize $u_i(K_i, A_i) := Y_i(K_i, A_i) - \xi(K_i, A_i) + \gamma_i(K_i)$, where $\gamma_i(K_i)$ (with $\gamma_i'(K_i) > 0$) is an extra term for empire building. With such preference, firm i will try to get the most investment. Hence, the tension is that the companies will try to induce the investor to over-invest. They can potentially exploit the fact that the investor does not know all information and that it is costly to acquire information.

Information. Firms' production functions and prior beliefs are common knowledge in the economy. Productivity realization A_i is private information only known to firm i with probability 1, not the investor or other firms. For all firms, A_i are assumed to be mutually independent, hence shutting down the cross inference of the investor. Firm i, however, can voluntarily disclose A honestly and fully to the investor if it wants to; otherwise, it can remain silent. Like Dye (1985), it is assumed that the entrepreneur cannot lie about its type and disclose A perfectly if it communicates any information.

The disclose-or-withhold assumption is reasonable in the discussion of many investor-firm relationships. Legal procedures are in place to ensure that the firms receiving investment release information honestly. For instance, the legal due-diligence process verifying information authenticity is required before a VC deal with an entrepreneur is finally struck. Apart from procedural scrutiny, there are strong economic incentives for surviving fact checks as well. Dishonesty may bring severe consequences in litigation and reputation harm. Hence I adopt the setup where the action space of a firm includes either an honest full disclosure or withholding.

The key ingredient of the game is the investor's limited attention. It is assumed that even if firm i communicates some information, the investor may not hear it because he has not paid enough attention to that company. What captures the investor's attention level is π_i, the probability that the investor observes disclosure conditioning on company i has provided a disclosure. A higher π_i indicates a higher level of interest the investor has for that company. The investor can pay to boost up to π_i with monitor cost c_i per unit of probability, which captures how difficult it is for the investor to investigate firm i. The total attention resource d is limited for the investor. That is to say, $\int_i c_i \pi_i g(i) di \leqslant d < \int_i c_i g(i) di$. Unable to monitor every company with full attention, the investor must make an attention allocation decision.

This information setting is related to but different from the evidence of Dye (1985). In Dye (1985), there is some probability that the privately informed sender does not receive information from nature. Hence, when the receiver does not observe anything, he is unsure whether it is because the sender's type is too bad or because the sender is uninformed. That hesitation prevents the receiver from confirming the sender as the worst type and prevents full revelation of types. In my setting, the sender is always privately informed. Meanwhile, it is the inattentiveness of the investor that prevents full revelation. When the investor observes no information, he is unsure whether it is because bad news is hidden or the communication fails to deliver due to inattention. The investor will pick the level of attention endogenously.

Timing. There are three stages: the information stage (Stage 1), the investment stage (Stage 2), and the realization stage (Stage 3). In Stage 1, the investor allocates attention, and companies observe the assigned levels and then make disclosure decisions. I will also discuss the case where companies make disclosure decisions without observing allocated attention levels. In Stage 2, after the investor receives disclosure or nothing, she invests with his updated beliefs. In Stage 3, everything is realized, and everyone is paid.

2.2.2 The Benchmark Model

In the benchmark model, I consider the case where the borrowing limit does not bind. The firm has access to ample funds and will continue to invest until the marginal return to investment fails to exceed the opportunity cost R_0. For massive investors such as pension funds, sovereign funds, or sizable private equities, it is common that they do not exhaust all funds available onto active strategy portfolios. Passively managed components may correspond to a positive weight in alternative exit lending in the model. In such scenarios, the funds are ample, and investors are trading off between portfolio firms with the alternative opportunity R_0 rather than among portfolio firms.

The game is sequential and solved by backward induction. It is necessary to first focus on the investment and disclosure decisions for any given level of attention. After that, the focus moves to find the optimal level of attention.

Investment and disclosure

To endogenize π_i for the investor, it must be studied how subsequent disclosure and investment decisions are made under any given π_i. The following analysis of investment and disclosure will exclusively focus on firm i. Hence I omit the subscripts i.

Given π, the firm will form a disclosure strategy by choosing a set $D_\pi \subset \mathbf{R}_+$. If $A \in D_\pi$, then company i will disclose its type. If otherwise $A \in ND_\pi = \neg D$, it will withhold. That is because at almost everywhere on $supp(A)$, the firm will not mix disclosure with non-disclosure out of indifference due to the monotonicity of $u(K, A)$.

The investor will either observe a disclosure of A or not see anything. In other words, if referring to the signal the investor eventually receives from the firm as

S_π, then S_π takes the value of either A when he observes communication, or *nobs* when he observes nothing. The distribution of S_π depends on the received attention $\pi = \Pr(S_\pi = A | A, A \in D)$.

Either way, the investor will update her belief on A given S_π and invest. That investment will determine the payoff of the firm, which prefers higher input levels. Firm i understands how the investor interprets its choice of D_π. Hence it determines D_π such that a disclosure is made if and only if the expected payoff of disclosure exceeds concealment.

Investment. Given any posterior belief about company i's type, the investor will invest $K^*_{S|\pi}$ that maximizes his expected payoffs. When there is no borrowing constraint, the investor trades off between the marginal return for investing in a given company and the reservation return R_0. When the investor observes a disclosure of company i's A, the investor can simply make a perfect-information investment $K^*_{S|\pi} = K^*_A$ that maximizes $v(K, A)$. Otherwise, she needs to update his beliefs with not observing disclosure and make an imperfect-information investment $K^*_{S|\pi} = K^*_{nobs|\pi}$ that maximizes $\mathrm{E}[v(K, A)|nobs, \pi]$, which depends on p, D_π and π that together pin down the posterior belief of the investor.

Disclosure. The disclosure set D_π is characterized by a cutoff A^*_π. Company i will disclose its type iff. $A > A^*_\pi$. That is because a better A will be matched with a higher level of investment due to complementarity, and hence when a company has a good type, it wants to disclose it to attract more capital. All types of A in the disclosure set are almost surely above those in the non-disclosure set. That is because the company's expected payoff should be the highest for disclosing the types in D_π, the second-highest for not making a disclosure, and the lowest for disclosing the types in ND_π, to avoid contradictions. Hence the disclosure strategy is necessarily characterized by a cutoff, similar to Dye (1985). The result is summarized in the following proposition. (Proof is in the Appendix.)

Proposition 1.1 *(cutoff strategy) For any given π, the disclosure strategy of company i features a cutoff A^*_π with $D_\pi = \{A \in supp(f) | A > A^*_\pi\}$. In other words, there exists A^*_π such that company i discloses almost surely iff. $A > A^*_\pi$.*

Meanwhile, only inferior companies hide their types, and therefore the investor will make bad assumptions about firms from which he does not observe information flows. However, because the investor is not fully attentive when $\pi < 1$, he is not confident that the firm is below the cutoff type. Specifically, upon seeing no information flows, the investor's posterior belief of A is

$$f_{A|nobs,\pi}(t) = \frac{1}{1 - \pi + \pi F(t)} \left((1 - \pi) f(t) + \pi f(t) \mathbf{1}_{t \leqslant A^*_\pi} \right) \tag{2.1}$$

Such belief prevents the company's type from full revelation and allows for an interior cutoff, around which the firm is indifferent between disclosure and no disclosure. The determination of the equilibrium cutoff is summarized in Proposition 1.2.

Proposition 1.2 *(equilibrium cutoff) (i). The cutoff A^*_π is a solution of $\bar{A}$ for an equation of $\bar{A}$:*

$$u(K^*_{\bar{A}}, \bar{A}) = u(K^*_{nobs|\pi}, \bar{A}) \tag{2.2}$$

(ii). Equivalently, it is a solution for

$$K^*_{\bar{A}} = K^*_{nobs|\pi} \tag{2.3}$$

(iii). The investor "assumes the worst" upon seeing no disclosure, i.e.

$$K^*_{nobs|\pi} = \min_{\bar{A}} K^*_{nobs|\pi,\bar{A}} \tag{2.4}$$

(iv). The cutoff A_π^* exists and is unique.

The proof is in the [Appendix]. Result (i) states that the firm is indifferent at the cutoff between disclosure or non-disclosure. The cutoff divides the support of A into the good news region where disclosure wins and the bad news region where concealment wins. Result (ii) shows that due to the monotonicity of $u(K, A)$, the indifference condition translates to the equality of perfect information and "no news" imperfect information investment levels at the cutoff. Result (iii) argues that, under given π, the investment strategy will lead to the cutoff picked, among all candidate cutoff values, to induce a posterior such that the no-news investment is the lowest. Intuitively, the investor "assumes the worst" when he observes nothing. The "worst" does not sink to the lowest type due to the probability that the investor inattentively misses the communication. Still, the investor uses the lowest reasonable level of investment for maximum punishment. This result derived with production complements the similar result of Dye (1985) in an analogous setting.

Comparative statics of π. Intuitively, attention increases revelation. A classic result of Dye (1985) is that the disclosure cutoff decreases with the chance that the company does not receive its type. Similar results hold for the benchmark model, presented in Proposition 1.3 (Proof see [Appendix]):

Proposition 1.3 (revelation increases with attention) (i). A_π^* is strictly decreasing in π. (ii). A_1^* is the lower bound of supp(A); $\lim_{\pi \to 0^+} A_\pi^*$ solves $\bar{A}$ from the equation $K_{\bar{A}}^* = K_{nobs|0}^*$.

Optimal attention allocation

The question moves to the search for the most efficient allocation of attention that maximizes the ex-ante expected value for the investor. The benefit to the investor for paying attention π to company i is to make a more informed and customized investment decision. By standard Blackwell arguments, that improves the ex-ante value of the expected utility maximization.

Let $V(\pi) = E[v(K^*_{S|\pi}, A)|\pi] = E[E[v(K^*_{S|\pi}, A)|S, \pi]\pi]$. The "Expected Value of Information", or the improvement in expected utility after learning the signal S_π regarding firm i is $EVI(\pi) = V(\pi) - V(0)$. In equilibrium, the attention allocation $\{\pi_i\}_{i \in I}$ should maximize $V = \int_{i \in I} EVI_i(\pi_i) g(i) di$, or equivalently $\int_{i \in I} V_i(\pi_i) g(i) di$ for the optimization purpose, up to the attention resource constraint.

For studying the optimal attention allocation, it is necessary to investigate what is $V(\pi)$ for a given firm i. For a more straightforward interpretation, rewrite it as:

$$
\begin{aligned}
& V(\pi) \\
&= E[v(K^*_{S|\pi}, A)|\pi] \\
&= E[E[v(K^*_{S|\pi}, A)|S, \pi]\pi] \\
&= E[v(K^*_{nobs|\pi}, A)|nobs, \pi] \Pr(nobs|\pi) + E[v(K^*_A, A)|\neg nobs, \pi] \Pr(\neg nobs|\pi) \\
&= \pi \int_{A > A*} v(K^*_A, A) dF(A) + \pi \int_{A \leqslant A*} v(K^*_{nobs|\pi}, A) dF(A) + ... \\
&\qquad (1 - \pi) \int_A v(K^*_{nobs|\pi}, A) dF(A) \\
&= \pi \int_{A > A*} v(K^*_A, A) dF(A) + (1 - \pi) \int_{A > A*} v(K^*_{nobs|\pi}, A) dF(A) + ... \\
&\qquad \int_{A \leqslant A*} v(K^*_{nobs|\pi}, A) dF(A) \\
&= E[v(K^*_A, A)|\pi] - E[\Delta_\pi(A)|nobs, \pi] \Pr(nobs|\pi)
\end{aligned}
$$

where $\Delta_\pi(A) = v(K^*_A, A) - v(K^*_{nobs|\pi}, A)$ is the increased payoff increment for knowing A. (In fact, $EVI(\pi) = E[\Delta_\pi(A)|\pi]$.)

Line [5] and Line [6] are two ways of rewriting the value function. Line [5] indicates that the value function is the expectation of investment payoffs under the prior distribution. Specifically, the prior distribution can be dissected into three territories. That dissection is shown in Figure 1. In territory (i), the investor makes a perfectly informed investment. In territory (ii), the investor fails to observe the communication and makes an imperfectly informed investment. In territory (iii), the investor never sees a disclosure because none is provided and can only make an imperfectly informed investment. On another note, Line [6] uses a different way of grouping terms and shows that the value function is equal to the ideally efficient expected utility (the first term) minus the welfare loss (the second term). Unsurprisingly, the loss of welfare occurs if and only if the investor fails to observe the firm's type, either because it is hidden or inattentively missed. Such failure of observation induces making an imperfectly informed investment.

The optimal allocation of attention is determined by the fact that the shape of $V(\pi)$ is increasing and convex. First, it should be natural that $V(\pi)$ is increasing in π. Intuitively, a higher level of attention will induce an increased amount of information being revealed, and by Blackwell's arguments, more information will contribute to more informed decision-making and increase the expected value. Specifically, for $\pi_1 < \pi_2$, learning S_{π_2} will yield the same posterior profile as learning both S_{π_1} and $S_{\Delta\pi}$, with $S_{\pi_1} \perp S_{\Delta\pi}|A$ and $S_{\Delta\pi}|A$ taking values of either A and *nobs* and

$$\begin{cases} \Pr(S_{\Delta\pi} = A|A) = \frac{\pi_2 - \pi_1}{1-\pi_1}, & \text{for } A > A^*_{\pi_1} \\ \Pr(S_{\Delta\pi} = A|A) = \pi_2, & \text{for } A^*_{\pi_2} < A \leqslant A^*_{\pi_1} \\ \Pr(S_{\Delta\pi} = A|A) = 0, & \text{for } A \leqslant A^*_{\pi_2} \end{cases} \tag{2.5}$$

Since $S_{\Delta\pi}$ is a non-trivial signal, it generates positive value and hence $V(\pi_1) < V(\pi_2)$.

Apart from directly citing the Blackwell-style result, one can alternatively take the derivative of $V(\pi)$ with respect to π directly and observe its sign. That practice will help us to understand details of how attention contributes to $V(\pi)$ and is also critical in investigating the curvature of $V(\pi)$. The derivative is

$$\frac{dV(\pi)}{d\pi} = \frac{dV}{d\pi} + \frac{dV}{dA^*_\pi}\frac{dA^*_\pi}{d\pi} + \frac{dV}{dK^*_{nobs|\pi}}\frac{dK^*_{nobs|\pi}}{d\pi} \tag{2.6}$$

The decomposition shows the three channels how attention brings benefits. The second and third components are both zero, making the first component the sole surviving term. The detailed derivation is in the [Appendix] (Proof of Theorem 1).

The first component describes the direct benefit of increasing π. In that component, more attention leads to a lower chance of missing good news. When $A > A^*_\pi$, the firm will communicate, and therefore, increasing attention raises the opportunity for the investor to receive the communication and invest efficiently. The term is positive. In [Figure 1], this is represented by an expansion of territory (i) into territory (ii). On a side note, when $A < A^*_\pi$, there is no chance of observing disclosure, and hence territory (iii) is irrelevant.

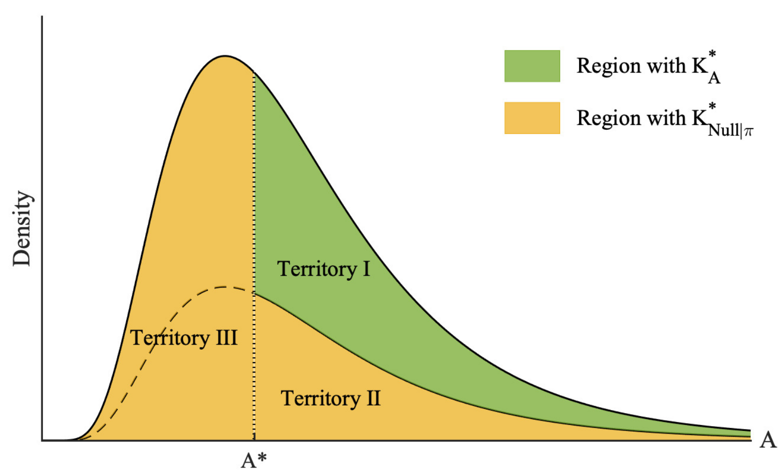

Figure 2.1: **Regions of the Prior Belief of** A

The second and the third components are indirect effects of increasing π. In the second component, the company responds to increased attention by lowering its disclosure cutoff. That expands the territories of (i) and (ii) while shrinking the territory of (iii). Specifically, an increase in π will flip some marginal types around the cutoff from surely inducing imperfectly informed investment $K^*_{nobs|\pi}$ to, with probability π, inducing perfectly informed investment $K^*_{A^*_\pi}$. That, however, will not affect the utility of the investor because by Proposition 1.2 (ii), $K^*_{A*} = K^*_{nobs|\pi}$ at the cutoff. That is an interesting result. Eventually, despite the revelation of more types, such revelation does not improve the investor's expected utility.

In the third effect, as the attention level increases, the investor's posterior belief conditioning on no observation will change, thus influencing $K^*_{nobs|\pi}$ and hence the value function. That effect, however, is also zero. Because $K^*_{nobs|\pi}$ is the optimizer of utility, it follows directly from Envelope Theorem that $\frac{dV}{dK^*_{nobs|\pi}} = 0$. Hence both indirect effects are zero. Just like the second effect, despite increased efficiency, the expected utility of the investor does not improve.

To summarize, the investor wants to increase attention π_i only because doing so decreases his chance of missing good news from that firm. Other effects do not play a role at all. That is a crucial result that explains the only reason why attention has some worth. Specifically, increased attention only helps to catch more good disclosures, i.e.,

$$\frac{dV(\pi)}{d\pi} = \int_{A>A^*_\pi} \Delta_\pi(A)dF(A) \tag{2.7}$$

Having obtained an analytical expression for the first derivative, it is straightforward to obtain the second derivative:

$$\frac{d^2V(\pi)}{d\pi^2} = -\left(\int_{A>A^*_\pi} \frac{dv(K^*_{nobs|\pi}, A)}{dK^*_{nobs|\pi}} dF(A)\right)\left(\frac{dK^*_{nobs|\pi}}{d\pi}\right) \tag{2.8}$$

The sign for the second bracket is negative by Proposition 1.3. The question is on the sign of the first bracket. It can be shown that the first term must be positive.

The proof is in the Appendix (Proof of Theorem 1).

Intuitively, the convexity of $V(\pi)$ is due to the increased stakes of making mistakes with higher levels of π. It has been established that the marginal value of π is its ability to avoid missing good news. The consequence of missing good news is graver as the attention level increases. If with high attention levels, the investor who observes nothing tends to believe it is highly likely due to firm i's concealment. Hence, the investor will invest very little in the firm. However, that leads to him making a big mistake when no observation results from an inattentive miss. Therefore, the higher the investor has boosted attention, the more incentivized the investor feels to boost it up even more.

With convexity of $EVI(\pi)$, the investor naturally tend to concentrate her attention. With the linear cost assumption, the effect is clean-cut and the optimal attention allocation strategy is a mix of bang-bang solutions. The only reasonable π values are 100% and 0. The payoff of attention 100% attention is $EVI(1)$. In decision theory, this value is also called Expected Value of Perfect Information ($EVPI$; $EVPI := EVI(1)$). It takes up c amount of resource. Hence the returns to attention is $\rho = \frac{EVPI}{c}$. The investor should sort all ρ_i and pay attention to those that feature the highest ρ_i until the attention resource is exhausted. Formally, these results are summarized in [Theorem 1] (additional proof see Appendix):

Theorem 1 (i). (monotonicity) $V(\pi)$ *and* $EVI(\pi)$ *increase in* π.
(ii). (convexity) $V(\pi)$ *and* $EVI(\pi)$ *are strictly convex in* π.
(iii). (attention allocation) In the equilibrium attention allocation $\{\pi_i\}_{i\in I}$, π_i *is equal to either 1 or 0. The investor will calculate* $\rho_i = \frac{EVPI_i}{c_i}$ *for each* i*, sort from high to low, and pay attention resource to companies with the highest* ρ_i *until the attention resource is exhausted.*

That is a striking result. The nature of the strategic voluntary disclosure problem creates a tendency for the investor to focus his attention instead of spreading it out evenly. The investor is effectively a concentration strategy. Denote $J = \{i \in I | \pi_i = 1\}$. She looks for those companies that monitoring pays off the most and then devotes all her resources to investigating those firms. Meanwhile, she ignores other companies. Those companies that receive attention will fully reveal their types. If for marginal i, the investor's attention resource is not enough to include all $g(i)$ companies in his monitor set, then the investor will announce clearly which ones in the $g(i)$ firms will be monitored. In a later section, I will describe the characteristics of firms selected in J.

Welfare. The model's proper measure of social welfare is the output surplus, i.e., the aggregate

$$\int_{i \in J} E[Y(K_A^*, A) - R_0 K_A^*] g(i) di + \int_{i \notin J} E[Y(K_{nobs|0}^*, A) - R_0 K_{nobs|0}^*] g(i) di \quad (2.9)$$

There are two sources of inefficiency. The first is that the investor is sometimes making imperfectly informed decisions. The second is that, even if the investor is perfectly informed, she maximizes her own payoff, not the social surplus. Let $\hat{K}_A^*$ denote the solution to $\max_K Y(K, A) - R_0 K$, the socially efficient investment level. The mentioned second friction is that $\hat{K}_A^*$ and K_A^* are not necessarily equal. When information friction must be present, the only way for maximizing efficiency is to have the investor totally own all cash flows, i.e. $\xi = Y$. That happens when the investor is induced to produce the efficient amount of outputs for any beliefs of A on the equilibrium path. Section 3.2 discusses this issue in greater detail.

2.2.3 Robustness and Comparative Statics

Robustness to firms not observing π_i

It may be argued that, sometimes, it is a strong assumption that the companies can observe the attention allocated to them. Start-up entrepreneurs do not necessarily know whether venture capitalists have focused on them in an early stage. A PE general partnership does not necessarily realize a sovereign fund is evaluating it. A subsidiary manager of an organization may not necessarily know the CEO's upcoming supervision plans. If the firms do not observe the investor's attention when making disclosure decisions, then the attention and disclosure strategies are part of a simultaneous game. The solution will be characterized by a Bayesian Nash Equilibrium (BNE) rather than the sequential equilibrium discussed above in the previous setting. In the simultaneous setup, there are generally multiple equilibria. In all equilibria, however, the investor is adopting a concentration strategy, i.e., she chooses and publicly announces a subset of firms to which she will pay 100% attention. The strategy profile in the previously discussed sequential game forms one PBE among all possible PBEs, and out of all equilibria, it yields the highest expected payoff for the investor. From this perspective, the PBE solution enjoys robustness to the entrepreneurs' information sets. The results are presented in the following [Proposition 2.1]:

Proposition 2.1 (disclosure without observing attention) (i). There exists BNE for simultaneous attention allocation and disclosure. In any BNE, the attention allocation $\{\pi_i\}_{i\in I}$ *features a fully monitored subset and the neglected rest.*
(ii). the equilibrium attention allocation in the PBE for sequential attention allocation and disclosure supports the BNE that maximizes payoff for the investor.

Interestingly, the concentration structure still holds even if the disclosure decisions are made without observing attention assignments. Essentially, that is because the benefit to attention, given any cutoff disclosure strategy of firms, is still convex. Out of the three components in the derivative of expected utility improvement with respect to attention, the first direct component is still the sole remaining term, and despite slight differences, increases with the attention level. The second term of increased revelation no longer exists due to the fixed strategy of the firm. The third effect is zero again because of the Envelope Theorem. Hence, the dominant strategy for the investor involves choosing a subset of firms to fully investigate. Meanwhile, as long as there is some positive probability for a firm to be chosen into the monitored subset, it will fully reveal its private information. Such correspondences lead to the bipartite solution.

How are returns to attention ranked?

A natural question is what are the characteristics of the subset of companies selected by the investor. An obvious observation is that an increased level of c_i decreases ρ_i. Essentially, the parameter describes how costly the investor is to monitor the firm. For example, if the investor is specialized in the firm's business, then c_i will be low, and the returns to attention will be high. Another example is that if the firm has bad presentation skills or is ill-managed, it may be challenging to dig information, so c_i may be high. A low c_i signifies a better match between the fund and the firm.

Less obvious is the relationship between $EVPI$ and firm attributes, namely the prior F and the production function, as well as the relationship between $EVPI$ and the reservation return R_0.

As Lawrence (2012) points out, it is generally difficult to examine $EVPI$'s relationship with prior, since a change in the prior will affect both $V(0)$ and $V(1)$, thus making the change in their differences unclear. The $EVPI$ for the investor in learning a given firm is

$$EVPI = \int_A \left(v(K_A^*, A) - v(K_{nobs|0}^*, A)\right) f(A)dA \tag{2.10}$$

To begin with, I examine the prior belief's influence on $EVPI$. The two common attributes associated with a probability distribution are scale and location. The prior scale, or risk, is critical in determining gains of investigation. A sufficient condition for higher risk (in terms of Mean Preserving Spread) leading to higher $EVPI$ is for the value increment function $\Delta_\pi(A)$ to be convex. Past research (Gould (1974), Hess (1982)) show that it holds under some regularity conditions:

Proposition 2.2 (cf. Hess (1982)) $EVPI$ increases with the the uncertainty of the prior (in the sense of MPS), if (i). $\xi_{AA} \leqslant 0$, (ii). $\xi_{AA}\xi_{KK} \leqslant \xi_{AK}^2$.

Sufficiently, the production function needs to be linear or concave in productivity but not concave in both productivity and capital as a bivariate function. A special case is when productivity and capital are separable in the form of $\xi(K, A) = \xi_1(A)\xi_2(K)$, or when the payoff is linear in the random state $\xi_1(A)$. Under such linearity, the prior decision and the expected utility $V(0)$ will be the same for distributions of $\xi_1(A)$ that differ by a mean-preserving spread. $V(1)$ is bigger under higher uncertainty due to $\xi(K, A)$'s concavity in K. An example is a constant share of the Cobb-Douglas production $\xi(K, A) = wAK^\alpha$. Intuitively, under some common conditions, the investor wants to select firms with more uncertainty into its subset of investigation because she gains more from real optionality.

The location of the prior also affects informational value of monitoring. To put on more structures, I assume that the distributions F are in a location-scale family parameterized by (μ, σ). Let X be a random variable with cdf F_X and a smooth pdf f_X. Let the location-scale family follows the distribution $F^{(\mu,\sigma)} \sim^d \mu + \sigma X$, or

$F^{(\mu,\sigma)}(A) = F_X(\frac{A-\mu}{\sigma})$ (such distributions are independent). Then

$$EVPI = \int_A \left(v(K_A^*, A) - v(K_{nobs|0}^*, A)\right) \frac{1}{\sigma} f_X(\frac{A-\mu}{\sigma}) dA \tag{2.11}$$

and this can help to examine how parameters μ and σ affect $EVPI$:

$$\begin{aligned}
\frac{dEVPI}{d\mu} &= -\frac{1}{\sigma^2} \int_A \left(v(K_A^*, A) - v(K_{nobs|0}^*, A)\right) f'_X(\frac{A-\mu}{\sigma}) dA \\
&= \frac{1}{\sigma} \int_A \left(v_A(K_A^*, A) - v_A(K_{nobs|0}^*, A)\right) f_X(\frac{A-\mu}{\sigma}) dA \\
&= \frac{1}{\sigma} E\left[v_A(K_A^*, A) - v_A(K_{nobs|0}^*, A)\right]
\end{aligned}$$

The second line follows from integration by parts. The sign is undetermined, and depends on the relative size of $E[v_A(K_{nobs|0}^*, A)]$ and $E[v_A(K_A^*, A)]$. That is to say, the investor is not necessarily interested in companies with higher productivity or bigger in implicit size, for that matter. It depends on the average sensitivity of values with respect to productivity:

Proposition 2.2 (EVPI and location) For productivity distribution $F_X(\frac{A-\mu}{\sigma})$, $EVPI$ *increases with* μ *iff.* $E[v_A(K_{nobs|0}^*, A)] < E[v_A(K_A^*, A)]$.

Example. It is helpful to examine an example with Cobb-Douglas production and uniform distributions for illustration. Set $Y(K, A) = AK^\alpha$ and $v(K, A) = wY(K, A) - R_0 K$. It is straightforward that

$$K_A^* = \alpha^{\frac{1}{1-\alpha}} (R_0/w)^{\frac{1}{\alpha-1}} A^{\frac{1}{1-\alpha}}$$

and

$$K_{nobs|0}^* = \alpha^{\frac{1}{1-\alpha}} (R_0/w)^{\frac{1}{\alpha-1}} E[A]^{\frac{1}{1-\alpha}}$$

Hence

$$EVPI = \left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right)\left((R_0/w)^{\frac{\alpha}{\alpha-1}}\right)\left(E[A^{\frac{1}{1-\alpha}}] - (E[A])^{\frac{1}{1-\alpha}}\right) \tag{2.12}$$

Let A be drawn from a uniform distribution $U(\mu - e, \mu + e)$. Then

$$EVPI \propto \frac{1}{2e}\frac{1-\alpha}{2-\alpha}\left[(\mu+e)^{\frac{2-\alpha}{1-\alpha}} - (\mu-e)^{\frac{2-\alpha}{1-\alpha}}\right] - \mu^{\frac{1}{1-\alpha}} \tag{2.13}$$

It is obvious how location μ and scale e affect $EVPI$. Take derivatives and get

$$\frac{dEVPI}{d\mu} \propto E[(A^{\frac{1}{1-\alpha}})'] - (\mu^{\frac{1}{1-\alpha}})' \tag{2.14}$$

By Jensen's inequality, whether the investor prefers to investigate higher or lower productivity depends on the shape of the production α. If $\alpha > 0.5$, then the investor investigates high μ firms. If $\alpha < 0.5$, then she investigates low μ firms. If $\alpha = 0.5$, then there is no location preference. The preference for scale follows Hess (1982) result.

The opportunity cost R_0 also affects $EVPI$. It is obvious that a rising required return R_0 decreases K_A^* and $K_{nobs|0}^*$, due to concavity of payoff functions. A higher R_0 also decreases $V(0)$ and $V(1)$. By Envelope Theorem, $\frac{dEVPI}{dR_0} = K_{nobs|0}^* - \mathrm{E}[K_A^*]$. The direction of $EVPI$ change is unclear; it depends on the shape of the contract. Same analysis applies to the equivalent case of $\xi(K, A)$ changing proportionately to $c \times \xi(K, A)$.

In summary, the rank of returns to attention is nontrivial. In particular, the selection of the firms receiving attention is not simply following a naive metric of firm quality, such as firm size or value. Scale may boost $EVPI$ under certain circumstances, and other attributes' effects are less clear and may depend on parameters.

2.2.4 Application: Firm Transparency and Efficiency

Since Dye (1985), there have been discussions about the asymmetry between good and bad news in the timing of uncertainty resolution. Because entrepreneurs use a cutoff strategy for voluntary disclosures, it is conventional wisdom that the uncertainty to the investors will be resolved at the interim stage by disclosures for the good types above the cutoff and at the realization stage for the bad types below the cutoff. According to this view, the cross-section of remaining uncertainty in investors' beliefs is inversely related to the quality of the firms, and one can infer a firm is of high quality from its increased information transparency.

That traditional view, however, does not consider the endogenous information acquisition by the investor. In other words, it implicitly assumes that the strategic information injected into markets is driven by the supply side, while taking the demand side as exogenous. This paper, however, differs from the traditional perspective in adding demand side into the picture, thus leading to the prediction that earlier uncertainty resolution and better information transparency does not necessarily imply more desirable firm quality. In this paper, all firms are divided into two groups based on attention received. Those firms in the subset receiving attention will fully reveal at the interim stage, hence resolving all uncertainty earlier and presenting no asymmetry. Those outside the subset will disclose if their types are better than a highest possible cutoff level, at which the informed investment level is equal to the completely uninformed investment level based on prior. In those firms, one may observe the maximum level of asymmetry of uncertainty resolution.

Hence, the traditional cross-sectional prediction breaks down because the determination of the investigated subset does not necessarily coincide with the levels of firm types or other measures of desirability. As discussed, the investor will focus her attention on the firms with the highest returns to attention, determined by costs to investigate and their EVPI. The rankings of returns to attention do not necessarily correlate with more desirable firm quality. Under some conditions, the investor would like to investigate those firms with higher risks due to their higher EVPI. In certain ranges of parameters, the investor intends to pay attention to firms with a lower location of productivity distribution because of higher EVPI. Such firms may be considered as having "lower quality" but will be pressed to be transparent at the interim stage. In addition, the easiness to learn about a firm also matters in the ranking of returns to attention, which is not necessarily associated with firm quality. Hence, one cannot infer a firm to be "good" in comparison simply because it resolves its uncertainty; it could have drawn investors' attention and received more pressure in the early stage to release information.

2.3 Attention Attitudes and Security Design

This paper's attention and information release results do not rely on strong assumptions on the securities splitting the surplus. Hence, I study the features of optimal contracts used in financing that satisfy the assumptions in this model. Because the optimal contract eventually depends on the bargaining powers of the two sides, I find bounds to optimal contracts that respectively give all bargaining power to either the investor or the entrepreneur. Assume all market power goes to the investor and the reservation payoff to an entrepreneur is zero, then with limited liability, the contract will give all payoffs to the investor, i.e., $\xi = Y$. The investor will maximize total output, achieving maximum economic efficiency, while the firm will receive a payoff that increases with a tiny slope ε, or effectively zero. This boundary contract is trivial. The other bound with the firm enjoying bargaining power is worthy of discussion.

Before discussing the optimal contract, I clarify the meaning of contracts in the setup for a better understanding. First of all, contractible variables are both A and K. Productivity A is contractible because one can back out A from observables K and Y. Secondly, in convention, contract design focuses on the relationship between ξ and Y (or A), taking K as given. Hence, the characterization of the contract as "equity", "debt", or other types is to interpret the shape of $\xi(K, A)$ as a function of A (or equivalently Y) for a given K. However, it is also important to examine

$\xi(K, A)$ as a function of K for a given belief in A because that perspective determines investor inputs. The discussed object, $\xi(K, A)$, is essentially a profile of contracts indexed by K. The basic idea of the contract analysis is first to fix a belief in A and examine investment decisions based on ξ as a function of K, essentially pinning down the contract profiles, and then to switch perspectives and examine what ξ looks like as a function of A for each fixed K.

Finally, the contract $\xi(K, A)$ is agreed upon before attention allocation and before the entrepreneur observes A. That is undoubtedly before the investment stage. Only then can the interaction between information and investment strategies be obtained. Such contracts can be interpreted as the routine that players will abide by in convention. In practice, the forms of financing contracts and terms used, for instance, between VCs and start-ups or between LPs and GPs, are relatively stable and expected.

Suppose all firms are homogeneous with the same prior beliefs and production functions. Assume the entrepreneurs propose leave-it-or-take-it contracts. The entrepreneurs can retain surplus from the investor; however, they still face competition for limited attention resources from the investor. I require all regularity conditions must hold. Namely, the contracts must satisfy exogenous conditions including $\xi(K, A)$ increasing in A, K and concave in K, complementarity between A and K, and $u(K, A)$ increasing in A, K.

Another requirement is that I focus on studying socially efficient contracts. Such contracts maximize the expected total output, conditioning on the information set. The financing instruments adopted evolve and form over the long term. It makes sense to focus the study on socially efficient contracts since, over the long run, such contracts maximize the total welfare and are eventually beneficial. Specifically, I design $\xi(K, A)$ such that $K_A^* = \hat{K}_A^*$ and $K_{nobs|0}^* = \hat{K}_{nobs|0}^*$; the latter investment levels maximize $Y(K,A)-R_0K$ and $\mathrm{E}[Y(K,A)]-R_0K$, respectively. An implication is that for any given possible belief of A, the expected output is fixed. Hence the contract design is a fixed-sum game between the investor and firms.

Equilibrium characterization. Suppose the measure of firms is 1, attention cost is 1 and amount of attention resources is $\beta < 1$. Let $U(1) = \mathrm{E}[u(K_A^*, A)]$ and $U(0) = \mathrm{E}[u(K_{nobs|0}^*, A)]$ be the expected payoffs receiving or not receiving attention, respectively. The definition is analogous to $V(1)$ and $V(0)$.

In equilibrium, there will be at most two groups of firms. There are β firms receiving attention (group h, for high EVPI). They have a higher EVPI (denote as δ_h). They optimally choose to receive attention with $U_h(1) \geqslant U_h(0)$. There are another $(1-\beta)$ firms not receiving attention (group l, for low EVPI). They have a lower EVPI (denote as δ_l; $\delta_l \leqslant \delta_h$) and optimally choose to hide away from attention with $U_l(1) \leqslant U_l(0)$.

In equilibrium, no player can profitably deviate by choosing a different legit contract. An obvious implication is $U_h(1) = U_l(0)$. Also, among all legit contracts with $EVPI \geqslant \delta_h$, the $U_h(1)$ is the highest $U(1)$, hence shutting down further motives to compete for attention. Similarly, among all legit contracts with $EVPI \leqslant \delta_l$, the $U_l(0)$ is the highest $U(0)$, shutting down motives to compete for avoiding attention.

The following proposition describes the best socially optimal contract profiles:

Proposition 3.1 (optimal socially efficient contract) If the firms have the bargaining power, then the socially efficient contract profile in equilibrium is

$$\xi^{(h)} = \begin{cases} R_0 K & K \leqslant \hat{K}_A^* < \hat{K}_{nobs|0}^* \\ Y(K,A) - (Y(\hat{K}_A^*, A) - R_0 \hat{K}_A^*) & \hat{K}_A^* \leqslant K < \hat{K}_{nobs|0}^* \\ Y_K(K_{nobs|0}^*, A) K & K < \hat{K}_{nobs|0}^* \leqslant \hat{K}_A^* \\ Y(K,A) - \Big(Y(\hat{K}_{nobs|0}^*, A) - Y_K(K_{nobs|0}^*, A) \hat{K}_{nobs|0}^* \Big) & K > \hat{K}_{nobs|0}^* \end{cases}$$

$$(2.15)$$

for the high EVPI group of measure β *and*

$$\xi^{(l)} = \begin{cases} R_0 K & \text{for } K \leqslant \hat{K}^*_A < \hat{K}^*_{nobs|0} \\ Y(K,A) - (Y(\hat{K}^*_A, A) - R_0 \hat{K}^*_A) & \hat{K}^*_A \leqslant K < \hat{K}^*_{nobs|0} \\ Y_K(\hat{K}^*_{nobs|0}, A)K & K < \hat{K}^*_{nobs|0} \leqslant \hat{K}^*_A \\ Y_K(\hat{K}^*_{nobs|0}, A)\hat{K}^*_{nobs|0} + Y_K(\hat{K}^*_A, A)(\hat{K}^*_A - \hat{K}^*_{nobs|0}) & \hat{K}^*_{nobs|0} < K \leqslant \hat{K}^*_A \end{cases} \tag{2.16}$$

for the low EVPI group of measure $1-\beta$*.*

The verification of the equilibrium is as follows. First, both payoff profiles are legit: they satisfy social efficiency ($K^*_A = \hat{K}^*_A$ and $K^*_{nobs|0} = \hat{K}^*_{nobs|0}$ for both contracts) and in either case, the conditions of monotonicity, concavity and complementarity hold. Second, $\delta_h > \delta_l$. That is because $V(0)$ are equal for both contracts, while $V(1)$ on the section of $K > \hat{K}^*_{nobs|0}$ is bigger for the high-EVPI contract. Third, there is no profitable deviation. $U_h(1) = U_l(0)$ (and also $= U_h(0)$). In fact, the contracts are constructed so that $V(0)$ is the smallest possible for all legit contracts (explained in the next paragraph). Suppose a firm from the low-EVPI group wants attention. Then it must shoot up the contract-implied $V(1)$ to higher than that of the high-EVPI group in order to get attention. Since socially optimal contract design is a fixed-sum game, $V(1) > V_h(1)$ implies $U(1) < U_h(1)$ (which is equal to $U_l(0)$). Hence no low-EVPI group firm will deviate. On the other hand, no high-EVPI group firm wants to evade attention. That is because $V(0)$ for both contracts are already the lowest possible as mentioned above, meaning $U(0)$ are the highest possible, given the fixed sum game. Evading attention by moving to any other legit contract cannot yield a higher $U(0)$.

The most important construction concerns minimizing $V(0)$. For social efficiency upon no observation, for any A, there must be $\xi_K(\hat{K}^*_{nobs|0}, A) = Y_K(\hat{K}^*_{nobs|0}, A)$. That is because for $K^*_{nobs|0} = \hat{K}^*_{nobs|0}$ to hold, it must be true that $\mathrm{E}[\xi_K(\hat{K}^*_{nobs|0}, A)] = \mathrm{E}[Y_K(\hat{K}^*_{nobs|0}, A)] = R_0$, while due to the monotonicity of $u(K, A)$ in K, $\xi_K(\hat{K}^*_{nobs|0}, A) \leqslant Y_K(\hat{K}^*_{nobs|0}, A)$. Hence the only possible scenario is $\xi_K(\hat{K}^*_{nobs|0}, A) = Y_K(\hat{K}^*_{nobs|0}, A)$. Under such slope restrictions, the minimal $V(0)$ involves looking for the lowest possible level of $\xi(\hat{K}^*_{nobs|0}, A)$ for any given A. That effort is presented in the optimal contracts for $K < \hat{K}^*_{nobs|0}$.

The contract solutions offer an insightful result. The two groups of firms that will receive different levels of attention choose different contracts. They are assigned different levels of attention, on the other hand, precisely because they have chosen different contracts, to begin with. The equilibrium generates endogenous heterogeneity out of ex-ante identical firms. The firms that will receive attention and are efficient are associated with high EVPI, and they prefer to be noticed (in fact, at the margin, they are indifferent about attention). The other firms do not want attention and do not receive it due to low EVPI. Attention competition takes place on $K > \hat{K}^*_{nobs|0}$, among the high-EVPI group. The firms can only retain a constant amount of output because of the pressure from competition to keep EVPI high.

From the conventional perspective of security design, I interpret the contracts by their payoff diagrams. Fix K, then $\xi(K, A)$ as a function of A for small $K < \hat{K}^*_{nobs|0}$, regardless of attention, is

$$\xi = \begin{cases} Y(K, A) - \Big(Y(\hat{K}^*_A, A) - R_0\hat{K}^*_A\Big) & \text{for } A \ s.t.\ \hat{K}^*_A < K \\ R_0 K & \text{for } A \ s.t.\ K \leqslant \hat{K}^*_A < \hat{K}^*_{nobs|0} \\ Y_K(\hat{K}^*_{nobs|0}, A)K & \text{for } A \ s.t.\ \hat{K}^*_A \geqslant \hat{K}^*_{nobs|0} \end{cases} \tag{2.17}$$

The payoff diagram has three parts. For low levels of A, the supposed efficient investment level is lower than K, and the payoff to the investor is the output less the amount paid to the maximum surplus the entrepreneur can extract under the efficient level of $\hat{K}^*_A$. For middle levels, the payoff is constant at an interest rate of R_0. For high levels of A, the payoff shoots up. The return $Y_K(\hat{K}^*_{nobs|0}, A)$ is the marginal product at $\hat{K}^*_{nobs|0}$ and increases with A. Effectively, the payoff to the investor looks like convertible debt. Figure 2 illustrates the payoff diagrams using a Cobb-Douglas production function.[1]

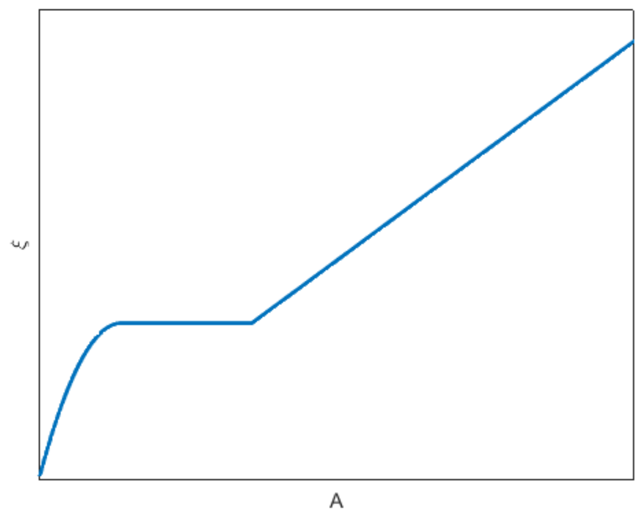

Figure 2.2: **Investor Payoff** ξ **as a Function of** A

From the analysis, it can be told that the essence of security design is to use on-equilibrium investment decisions to back out off-equilibrium payoffs. Although the payoff diagrams of the security show the payoffs ξ for all A, in fact only one point is on the equilibrium path, and that is the joint of the first and second sections of the payoff diagram. K is the perfect-information investment level for that particular A. All other points on the payoff diagram are not equilibrium pairs of values; however, they are critical in supporting the desired equilibrium.

The second kink that joins the second and third sections is the defining feature for a convertible bond. Below that level, the investor makes R_0K with a constant reser-

[1] Optimal contracts for $K < \hat{K}^*_{nobs|0}$. Linearity for high productivity states due to Cobb-Douglas production used in the illustration ($Y = AK^\alpha$, $R_0 = 2$, $K = 3$, $K_0 = 20$, $\alpha = 0.5$).

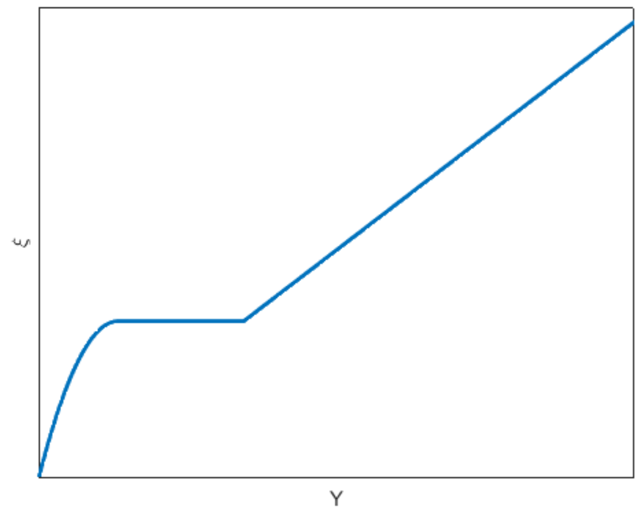

Figure 2.3: **Investor Payoff ξ as a Function of Y**

vation return, while above that level she makes $Y_K(\hat{K}^*_{nobs|0}, A)K$ with the effective return increasing in productivity. That kink is due to different payoff patterns between low and high productivity states that respectively induce a perfect-information investment level below or above the no-information investment level. In low productivity states, all surplus is extracted by the entrepreneur, and the investor only retains a net zero payoff. In high productivity states, the entrepreneur cannot extract all surplus, because he needs to sustain the no-information efficient investment level.

That payoff structure can explain the security design in many settings. For example, a venture capitalist may use convertible debt to finance a start-up. Fund managers from PE's general partnerships are paid with a waterfall structure comprising a scale component and a performance component upon very good states.

For large values of K, the securities are on the fourth lines of equations in Proposition 3.1.

2.4 Applications

The benchmark model has a wide range of applications. The key prediction generated by the model is that the firms will be categorized into two subsets with significantly different received attention levels. That demand-side factor with supply-side information release mechanisms generates predictions for many setups where information generation increases with more pressing investigation efforts. It seems common that the investor focuses cognitively on a subset of firms in the portfolio in the world of private finance. I use three examples to illustrate how the model can apply to the scenarios.

Example 1 (VC and start-ups) The world of venture capital and start-up finance is a fitting example of the model. Start-up entrepreneurs tend to speak highly of their own projects to attract financing. Although legal and economic mechanisms are in place to ensure they do not fake information, there is no binding obligation that they reveal all information unless pressed by the investor. Venture capitalists do not have unlimited human resources to spare and cannot possibly know every entrepreneur's details. VCs need to decide how to allocate scarce resources for maximum gains.

In practice, venture capitals in practice often adopt a multi-round financing strategy. Often PE/VCs invest in many start-ups in the first round, or "spray and pray", hoping to gain access to the start-ups and learn about the companies with the most potential. The practice can be viewed as the investor purchasing tickets to learning entrepreneurs' inside operations and acquiring basic information such as the prior beliefs, production function, and investigation costs. As a major stakeholder, the VC firm will closely follow the management and receive the private information flow. The investor will then focus on following up a subset of firms subsequently in later rounds of financing, as is described in the model.

For example, the setup can be applied to study the scope for the first-round investments. Assume that the investor can spend a certain amount of money on an entrepreneur to build relations with the entrepreneur and effectively purchase access to further insider details, such as the distribution of productivity and the production functions. The more the investor spends, the greater the second-round investment opportunities she can access and the higher payoffs she may harvest. The scope of the first-round spray-and-pray investments can be endogenously determined by the gains in payoffs and costs for first-round down payments.

There has been a discussion of the specialization of PE/VCs. Some investment firms are specialized in the investment of certain industries, while others are generalists without an obvious focus. In the model, specialization can be captured by the parameter c_i, the investigation cost specific to a pair of investor-entrepreneur. A VC is specialized in the investment of an industry if its c_i for firms in that industry is generally low. Hence, it often prioritizes learning about those firms and can invest efficiently.

Example 2 (LP and GP dynamics) While PE/VCs are the investor for start-ups, they are the investment targets for large funds, such as pension funds, sovereign funds, asset management, or other types of funds. Facing those Limited Partnerships, the General Partnership becomes the one with insider knowledge about the firms they manage and the portfolio outcomes. In practice, the LPs can be huge compared to their human resources. As a consequence, the attention limit will be binding, and the investor must allocate its monitoring resources wisely.

The information flow from GP to LPs can be studied with this model. LPs want to find competent asset managers, and the performance of the candidate GPs signifies their competence. Consider a scenario where a GP manages a portfolio, and before it closes, the GP aims to raise new funds for the same projects or new portfolios. The

LP will infer the quality of this investment opportunity from GP's ongoing projects, and hence the GP will strategically choose to release information to the LPs. A big LP will listen to many pitches from GPs, selectively investigate some of them, and then make the investment decision.

Example 3 (beyond private finance) The model can apply to a broader set of scenarios beyond financing private firms. One example is the relationship between the headquarter and subsidiaries within an organization. The CEO will be responsible for both supervising the branches and allocating resources to subsidiaries. Local managers will strategically inform the CEO about the operations to sway the CEO and attract more resources. The CEO only has constrained monitoring resources, and the model predicts that she tends to focus such resources to get to the bottom of some subsidiaries thoroughly. Similar patterns may also occur in the interaction between the central government, e.g., the fiscal authority and the local officials.

2.5 Conclusion

This paper introduces a combined model of attention allocation and investment by an investor and strategic information provision by her portfolio companies. The central mechanism is the concentration of attention by the investor in equilibrium, driven by the fact that expected utility improvement $EVI(\pi)$ is naturally convex in π. The portfolio companies are divided into two groups based on whether received attention is complete or nothing, based on the ranking of returns to attention ρ. A higher investigation cost lowers ρ. Firm attributes, such as prior location or scale, along with production and contracts, affect ρ. Under some assumptions, the investor may prefer to prioritize acquiring information of firms with higher uncertainty. I also present an example of investors prioritizing the investigation of firms with lower mean in productivity.

The model can be applied in a wide range of settings, in particular involving investments in private firms. Examples include VC investment in start-ups and PE's LP investment in GPs. The model generates predictions of information transparency and investment efficiency different from previous models that only consider the supply-side mechanisms. The link between firm quality and more information release is broken. The model can also study security design under the restrictions of assumptions made for the contracts. Ex-ante identical firms endogenously seek or avoid attention using different contracts. Using the optimal contract structure, I provide a rationalization for using convertible debt in private firm financing. The model can provide analyses of the interaction between investment and information strategies found in the relationship between, for instance, VC and start-ups, between LP and GPs, or between headquarters and subsidiaries.

Appendix A

Appendix to Chapter 1

A.1 Proofs of Propositions 1.3.3, 1.3.4, 1.3.5

For simplicity, in this section I refer to these propositions as Props 1, 2, and 3 respectively. For notational brevity, denote binomial masses $p_K := C_N^K p^K (1-p)^{N-K}$ and posterior beliefs $\pi'_k := \Pr(\theta = 1|k) = \frac{\pi \Pr(k|\theta=1)}{\Pr(k)}$. Utility function is

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

where $a^*(k) = \arg\max_a \pi'_k u_1 h(a) + (1-\pi'_k) u_0 h(1-a)$.

Proof of Proposition 1, 2. Prove Propositions 1 and 2 together. First, consider a weaker version of Proposition 1:

Proposition 1'. There exists an equilibrium with pure strategies.

Proceed the proof following the roadmap: $\to [1'] \to [2] \to [1]$.

Proof. (Proposition 1') For σ_{Kk},

$$\frac{\partial U}{\partial \sigma_{Kk}} = \frac{\partial U_k}{\partial \sigma_{Kk}} \tag{A.1}$$

$$= \frac{\partial U_k(\sigma_{Kk}, a^*(k))}{\partial \sigma_{Kk}} + \frac{\partial U_k(\sigma_{Kk}, a^*(k))}{\partial a^*(k)} \frac{\partial a^*(k)}{\partial \sigma_{Kk}} \tag{A.2}$$

$$= \frac{\partial U_k(\sigma_{Kk}, a^*(k))}{\partial \sigma_{Kk}} \tag{A.3}$$

$$= \pi u_1 p_K h(a^*(k)) + (1-\pi) u_0 p_{N-K} h(1-a^*(k)) \tag{A.4}$$

The third line follows the Envelope Theorem.

Now focus on such k that $\sum_K \sigma_{Kk} > 0$. Notice that the optimality of $a^*(k)$ is given by F.O.C.: $0 = \pi'_k u_1 h'(a) - (1-\pi'_k) u_0 h'(1-a)$, and by the Implicit Function Theorem,

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

Hence,

$$\frac{\partial^2 U}{\partial \sigma_{Kk}^2}$$
$$= (\pi u_1 p_K h'(a^*(k)) - (1-\pi) p_{N-K} u_0 h'(1-a^*(k))) \frac{\partial a^*(k)}{\partial \sigma_{Kk}}$$
$$= -\frac{(\pi u_1 p_K h'(a^*(k)) - (1-\pi) u_0 p_{N-K} h'(1-a^*(k)))^2}{\pi u_1 (\sum_{K'=1}^{N} p_{K'} \sigma_{K'k}) h''(a^*(k)) + (1-\pi) u_0 (\sum_{K'=1}^{N} p_{N-K'} \sigma_{K'k}) h''(1-a^*(k))}$$
$$\geqslant 0$$

The sign of the last line is nonnegative and in alignment with the numerator, because $h''(\cdot) < 0$. Hence Jensen's Inequality can be applied. And also notice that

$$\frac{\partial^2 U}{\partial \sigma_{Kk_1} \partial \sigma_{Kk_2}} = 0 \tag{A.5}$$

Thus if taking all $\{\sigma_{(\neg K')k}\}$ as given, then the problem of choosing optimal $\sigma_{K'k}$ is to optimally allocate 1 unit of probability resource of K' to convex and independent technologies $k = 0, 1, ..., n$.

The utility function is continuous and defined on a compact feasible set, so its optimization must have solutions. Now I show by contradiction that there must exist a pure strategy solution.

Suppose that optimally, $\sigma^*_{Kk_1} \in (0,1)$. Then there must exist $\sigma^*_{Kk_2} \in (0,1)$. Then 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 (Jensen's Inequality), 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$(\sigma^*_{Kk_1}, \sigma^*_{Kk_1} + \delta)$, then $\sigma_{Kk_1} = \sigma^*_{Kk_1} + \varepsilon$ and $\sigma_{Kk_2} = \sigma^*_{Kk_2} - \varepsilon(\varepsilon < \delta)$ will improve U (Jensen's Inequality), 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 $(\sigma^*_{Kk_2} - \delta, \sigma^*_{Kk_2} + \delta)$, then shift mass from σ_{Kk_1} (or σ_{Kk_2}) to σ_{Kk_2} (or σ_{Kk_1}). One of the two following cases will occur:

(iii-a). If $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} > 0$ or $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} > 0$ on $(0, \sigma_{Kk_1^*} + \sigma_{Kk_2^*})$, and w.l.o.g. assume $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} > 0$ when $\sigma_{Kk_1} = \sigma^0 \in (\sigma^*_{Kk_1}, \sigma_{Kk_1^*} + \sigma_{Kk_2^*})$, $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} = 0$ on $(\sigma^*_{Kk_1}, \sigma^0)$, and $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} = 0$ on $(\sigma^*_{Kk_2} - (\sigma^0 - \sigma^*_{Kk_1}), \sigma^*_{Kk_2})$, then let $\sigma_{Kk_1} = \sigma^0$ and $\sigma_{Kk_2} = \sigma^*_{Kk_2} - (\sigma^0 - \sigma^*_{Kk_1})$. It is an information structure with the same utility and falls into the category of (ii). Resort to the reasoning of (ii), and there is a contradiction to optimality;

(iii-b). If both $\frac{\partial^2 U}{\partial \sigma^2_{Kk_1}} = 0$ and $\frac{\partial^2 U}{\partial \sigma^2_{Kk_2}} = 0$ on $(0, \sigma_{Kk_1^*} + \sigma_{Kk_2^*})$, then shift all masses from $\sigma_{Kk_1} to \sigma_{Kk_2}$. If $\sigma_{Kk_1^*} + \sigma_{Kk_2^*} = 1$, then found an equivalent optimal solution but in pure strategies. If $\sigma_{Kk_1^*} + \sigma_{Kk_2^*} < 1$ then iterate the process of (i)(ii)(iii). $\square$

Note 1. In the proof of Proposition 1', I only focus on k with $\sum_K \sigma_{Kk} > 0$, i.e. the actually used codewords. Also when constructing contradiction, the only scenario where the alternative strategy is only weakly better is in Case (iii-b) when $\sigma_{Kk_1^*} + \sigma_{Kk_2^*} = 1$. In Case (iii-b), $\frac{\partial U}{\partial \sigma_{Kk_1}}$ is constant, and by the formula of $\frac{\partial U}{\partial \sigma_{Kk}}$, this only happens when $a^*(k_1)$ is constant with respect to σ_{Kk_1}. That implies K is the only fundamental with any mass that maps to codeword k_1, because the change of σ_{Kk_1} does not change the composition of probabilities that map to k_1. The same argument goes for k_2. Hence, by the operations of (iii-b) when $\sigma_{Kk_1^*} + \sigma_{Kk_2^*} = 1$, the resulting equivalent optimal solution must involve either k_1 or k_2 with $\sum_K \sigma_{Kk} = 0$, i.e. a codeword ending up unused.

Therefore, there are only two candidate scenarios for an optimal solution. It is either a pure strategy equilibrium with all codeword used, or a mixed strategy equilibrium that is equivalent to a pure strategy equilibrium with some codewords unused. To show Proposition 1, I just need to rule out the latter scenario. The help of Proposition 2 is needed. Next prove Proposition 2:

Proof. (Proposition 2) By Proposition 1' and Note 1, to show that no codewords should be wasted, I only need to consider pure strategy equilibria.

Proof by contradiction. Suppose in a pure strategy equilibrium Σ, $\sum_K \sigma_{Kk_1} = 0$, i.e. codeword k_1 is not used. Then N fundamentals are grouped into at most $n-1 \leqslant N-1$ bunches, and there exist one bunch with $m \geqslant 2$ fundamentals. Let this set of fundamentals be $\{K^{(1)}, ..., K^{(m)}\}$ and the corresponding codeword is k_2. Denote its contribution to the value function as U_{k_2} and optimal action $a^*(k_2)$. Consider another plan of codeword assignment Σ', with $C(K^{(1)}) = k_1$ and $C(K^{(2)}), ..., C(K^{(m)}) = k_2$, and other assignments unchanged. The contribution to the new value function by k_1 and k_2 are U'_{k_1} and U'_{k_2}, and the optimal actions are $a^{*'}(k_1)$ and $a^{*'}(k_2)$. By the standard argument for the positiveness of the value of information,

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

and the inequality holds strictly because $a^*(k_2)$, $a^{*'}(k_1)$ and $a^{*'}(k_2)$ cannot be equal. □

Note 2. For Proposition 2 to hold, only the inequality of $a^*(k_2)$, $a^{*'}(k_1)$ and $a^{*'}(k_2)$ is needed. It is sufficient that the distributions of the fundamentals are $K|\theta \sim Bi(N, p\theta + (1-p)(1-\theta))$. However, any distributions for K such that $a^*(k_2)$, $a^{*'}(k_1)$ and $a^{*'}(k_2)$ can ensure Proposition 2 to hold.

Finally, return to Proposition 1:

Proof. (Proposition 1). By Proposition 2, solutions with codeword unused is not optimal. Hence rule out the weakly better cases in the proof of Proposition 1' (see Note 1) and the optimal solution must only involve pure strategies. □

Proof of Proposition 3. Only need to prove Proposition 3(i), and then 3(ii) will naturally follow by substituting in the binomial distribution masses.

Proof. Prove by induction. Let $P(N, n)$ be the problem of choosing n fragments from N signals, and the goal is to show $P(N, n)$ features ordered partitioning solutions. I proceed by firstly showing the necessity of ordered partitioning solutions for $P(N, 1)$ by induction starting from $P(2, 1)$ and $P(3, 1)$, and secondly showing it for $P(N, n)$. For fundamental K, denote $x_K = \pi u_1 \Pr(K|\theta = 1)$ and $y_K = (1 - \pi)u_0 \Pr(K|\theta = 0)$. Each fundamental can be fully characterized by (x_K, y_K). Let $\eta_K := y_K/x_K = \frac{\pi}{1-\pi}\frac{u_1}{u_0}\Lambda(K)$. Without loss of generality, let $\Lambda(0) < \Lambda(1) < ... < \Lambda(K)$.

Step 1. Prove that $P(N, 1)$ features ordered partitioning solutions.

Step 1.1. Prove that $P(2, 1)$ features ordered partitioning solutions.

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

$$U = ((1 - s)x_0 + (1 - t)x_2)h(a) + ((1 - s)y_0 + (1 - t)y_2)h(1 - a) + ... \\ + (sx_0 + x_1 + tx_2)h(b) + (sy_0 + y_1 + ty_2)h(1 - b) \quad (A.6)$$

Then the strategy at $s = t = 0$ becomes what needs to be proven suboptimal. Notice that by the Envelope Theorem,

$$\frac{\partial U}{\partial s} = x_0(h(b) - h(a)) + y_0(h(1 - b) - h(1 - a))$$

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

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, hence completing the proof.

Therefore I proceed by showing at least one of the two partial derivatives is positive. Show by contradiction.

Suppose both partial derivatives are nonpositive, then

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

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

Assume w.l.o.g. $a > b$. Then on one hand, $\eta_1 = \frac{h'(b)}{h'(1-b)}$ by optimality of b. On the other hand, however, $\eta_1 < \eta_2 < \frac{(h(a)-h(b))/(a-b)}{(h(1-b)-h(1-a))/((1-b)-(1-a))}$, and 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.

(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 actions are the same as $s = t = 0$. The utility is $U|_{s=t=1} = U|_{s=t=0}$. However, the strategy at $s = t = 1$ does not make use of both codewords, and thus is suboptimal by Proposition 2 and Note 2. Hence the strategy at $s = t = 0$ is also suboptimal.

Thus completing the proof of Step 1.1.

Step 1.2. Prove that $P(3,1)$ features ordered partitioning solutions.

Prove by contradiction. Notice that the non-ordered-partition grouping plans are $\{B_k\}_{k=0}^1 = \{\{1\},\{0,2,3\}\}$, $\{\{2\},\{0,1,3\}\}$, $\{\{0,3\},\{1,2\}\}$ and $\{\{0,2\},\{1,3\}\}$. Discuss why these cases cannot be optimal.

Case (i). For $\{\{1\},\{0,2,3\}\}$ and $\{\{2\}\ ,\ \{0,1,3\}\}$: assume the optimal partition is $\{\{1\}\ ,\ \{0,2,3\}\}$. Consider another problem of partitioning fundamentals $\{0,1,v\}$ where v is characterized by $x_v = x_2 + x_3$ and $y_v = y_2 + y_3$. Thus the optimal partition must be $\{\{1\},\{0,v\}\}$ to avoid contradiction to the optimality of $\{\{1\},\{0,2,3\}\}$. However, the problem is $P(2,1)$, and since $\eta_0 < \eta_1 < \eta_v$, by the result of Step 1.1., $\{\{1\},\{0,v\}\}$ cannot be optimal because it is not a ordered partition; a contradiction. Analogous arguments apply for $\{\{2\},\{0,1,3\}\}$.

Case (ii). For $\{\{0,3\},\{1,2\}\}$: assume it is optimal. Consider another problem of partitioning fundamentals $\{0,v,3\}$ where v is characterized by $x_v = x_1 + x_2$ and $y_v = y_1 + y_2$. Thus the optimal partition must be $\{\{v\},\{0,3\}\}$ to avoid contradiction to the optimality of $\{\{0,3\},\{1,2\}\}$. However, the problem is a $P(2,1)$, and since $\eta_0 < \eta_v < \eta_3$, by the result from Step 1.1., $\{\{v\},\{0,3\}\}$ cannot be optimal because it is not an ordered partition. Contradiction.

Case (iii). For $\{\{0,2\},\{1,3\}\}$: assume it is optimal. Consider the following strategy: (i). 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 (ii). 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

$$U = (x_0 + sx_1 + (1-t)x_2)h(a) + (y_0 + sy_1 + (1-t)y_2)h(1-a) + ...$$
$$+ ((1-s)x_1 + tx_2 + x_3)h(b) + ((1-s)y_1 + ty_2 + y_3)h(1-b) \quad \text{(A.7)}$$

Then the strategy at $s = t = 0$ becomes what needs to be proven suboptimal. Notice that by the Envelope Theorem,

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

Consider two scenarios:

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

(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 $U|_{s=t=0} = (x_0 + x_1 + x_2 + x_3)h(a) + (y_0 + y_1 + y_2 + y_3)h(1-a)$. 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 for the alternative bundling $\{\{0,1,2,3\},\varnothing\}$, the optimal actions are the same as $s = t = 0$ and thus the utility levels are the same. However, the alternative bundling does not make use of both codewords, and thus is suboptimal by Proposition 2 and Note 2. Hence the strategy at $s = t = 0$ is also suboptimal.

Step 1.3. Given that $P(N-1,1)$ have ordered partition solutions, prove that $P(N,1)$ features ordered partition solutions ($N \geqslant 4$).

Suppose the optimal partition $\{B_0, B_1\}$ for $P(N,1)$ is not an ordered partition, then consider two cases:

(i). If there exists 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 $\{B_0', B_1'\}$, with $B_0' = B_0 \cup \{v\}\backslash\{i,i+1\}$ and $B_1' = B_1$. However, this is not a ordered partition, and therefore contradictory to the induction hypothesis.

(ii). It there does 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 $\{B_0', B_1'\}$, 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, and therefore contradictory to the induction hypothesis.

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

Step 2. Prove that $P(N, n)$ features ordered-partitioning solutions.

The problem is to allocate $\{0, 1, ..., N\}$ into $(n + 1)$ groups and let the solution be $\{B_0, ..., B_n\}$. For any $0 \leqslant i, j \leqslant n$ $(i \neq j)$, 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 $P(m_i + m_j, 1)$ of $\{K_1^i, K_2^i, ..., K_{m_i}^i,$ $K_1^j, K_2^j, ..., K_{m_j}^j\}$. By the assumed optimality of $\{B_0, ..., B_n\}$, the solution has to be $\{B_i, B_j\}$. By the result of Step 1, it is an ordered partition. Therefore, any two groups of $B_0, ..., B_n$ form an ordered partition. The only way possible is that $\{B_0, ..., B_n\}$ is an ordered partition. $\square$

A.2 Proof of Theorem 1.4.3

I refer to 1.4.3 as Theorem 2. Notice that solving for cutoffs in the fundamental space $\mathbb{R}$ is equivalent to finding cutoffs in $(0, 1)$, if there exists a strictly monotone mapping from the fundamentals to $(0, 1)$. For such distributions $F_{K|\theta}$ that $\tilde{a}(K)$ is continuous and strictly monotone (e.g. Gaussian mixtures), $\tilde{a}(K)$ can serve as that mapping. Hence, to understand K_i^*, I can equivalently study $\tilde{a}(K_i^*)$ on $(0, 1)$. Hence, define

$$\delta_n(a) := \frac{1}{n} \int_{\tilde{a}(K') \leqslant a} d\mathbb{1}_{\{K' \in \kappa^*(n)\}} \tag{A.8}$$

and let $\hat{\delta}(a) := \lim_{n \to \infty} \delta_n(a)$. Then it is obvious that

$$\beta'_\infty(K) = \hat{\delta}'(\tilde{a}(K))\tilde{a}'(K) \tag{A.9}$$

To show Theorem 2, I need to first show that cutoffs are dense on $\mathbb{R}$ as $n \to +\infty$:

Lemma A.2.1. *(cutoffs are dense in the limit) Under [A1'][A2'],* $\forall (K_1, K_2) \subset \mathbb{R}$*,* $\exists n^*$*, such that* $\forall n > n^*$*,* $\exists K^{(n)} \in \kappa(n)$ *and* $K^{(n)} \in (K_1, K_2)$*.*

Proof. Equivalently need to show $\forall (a_1, a_2) \subset \mathbb{R}$, $\exists n^*$, such that $\forall n > n^*$, $\exists K^{(n)} \in \kappa(n)$ and $\tilde{a}(K^{(n)}) \in (a_1, a_2)$. Prove by contradiction. Assume that $\exists (a_1, a_2)$ such that

$\exists n_j$ (a subsequence of n) and $\kappa(n_j) \cap (a_1, a_2) = \varnothing$, $\forall j$. Because there are n_j distinct elements in $\kappa(n_j)$, as $n \to +\infty$, $\max |\tilde{a}(K_i) - \tilde{a}(K_{i-1})| \leqslant \frac{1}{n_j} \to 0$ $(K_i, K_{i-1} \in \kappa(n_j))$. Hence for big enough j, one can move a cutoff from out of (a_1, a_2) and almost not hurting any value, then place it at $\frac{a_1+a_2}{2}$, hence strictly increasing value. □

Then move on to the proof of Theorem 2:

Proof. (Theorem 2) First, prepare by analyzing, for any arbitrary four-times continuously differentiable functions $g_1(x)$ and $g_2(x)$ $(x \in (0,1))$, the properties of equation

$$\frac{g_1(x_2) - g_1(x_1)}{g_2(x_2) - g_2(x_1)} = \frac{g_1'(x)}{g_2'(x)} \tag{A.10}$$

where $x_1, x_2, x \in (0,1)$ (let $x_1 < x_2$). Assume that the shapes of g_1 and g_2 ensure that $x \in (x_1, x_2)$. Let $x_m := \frac{x_1+x_2}{2}$, and respectively Taylor expand the left-hand-side numerator and denominator around x_m, and evaluate at x:

$$\begin{aligned}\frac{g_1(x_2) - g_1(x_1)}{g_2(x_2) - g_2(x_1)} &= \frac{g_1'(x_m)(x_2 - x_1) + \frac{1}{24}g_1'''(x_m)(x_2 - x_1)^3 + o(|x_2 - x_1|^4)}{g_2'(x_m)(x_2 - x_1) + \frac{1}{24}g_2'''(x_m)(x_2 - x_1)^3 + o(|x_2 - x_1|^4)} \\ &= \frac{g_1'(x_m) + \frac{1}{24}g_1'''(x_m)(x_2 - x_1)^2 + o(|x_2 - x_1|^3)}{g_2'(x_m) + \frac{1}{24}g_2'''(x_m)(x_2 - x_1)^2 + o(|x_2 - x_1|^3)}\end{aligned} \tag{A.11}$$

where the first line follows from the fact that the 0th, 2nd and 4th order terms are cancelled out. Then substitute it into the equation, transform it, and obtain

$$\frac{g_1'(x)}{g_2'(x)} = \frac{g_1'(x_m)}{g_2'(x_m)} + \frac{1}{24}\frac{g_1'''(x_m)g_2'(x_m) - g_2'''(x_m)g_1'(x_m)}{g_2'(x)g_2'(x_m)}(x_2 - x_1)^2 + o(|x_2 - x_1|^3) \tag{A.12}$$

Then Taylor expand $\frac{g_1'(x)}{g_2'(x)}$ to the 1st order around x_m and evaluate at x, thus getting

$$x - x_m = \frac{1}{24}\frac{(g_2'(x_m))^2}{g_1''(x_m)g_2'(x_m) - g_1'(x_m)g_2''(x_m)} \times ... \tag{A.13}$$

$$\frac{g_1'''(x_m)g_2'(x) - g_2'''(x_m)g_1'(x)}{g_2'(x)g_2'(x_m)}(x_2 - x_1)^2$$

$$+ o(|x_2 - x_1|^3)$$

Let the multiplier to quadratic term be called $\Gamma(x_1, x_2, x; g_1, g_2)$. Then as $\frac{x_2}{x_1} \to 1$,

$$\Gamma(x_1, x_2, x; g_1, g_2) \to \bar{\Gamma}(g_1, g_2) := \frac{1}{24}\frac{g_1'''(x)g_2'(x) - g_2'''(x)g_1'(x)}{g_1''(x)g_2'(x) - g_1'(x)g_2''(x)}$$

$$= \frac{1}{24}\frac{d}{dx}(\ln|g_1''(x)g_2'(x) - g_1'(x)g_2''(x)|) \tag{A.14}$$

Also notice that $x - x_m = -\frac{1}{2}((x_2 - x) - (x - x_1))$. Hence

$$(x_2 - x) - (x - x_1) = -2\Gamma(x_1, x_2, x; g_1, g_2) \cdot (x_2 - x_1)^2 + o(|x_2 - x_1|^3) \tag{A.15}$$

That concludes the analysis for the properties of the equation.

Now, begin solving [P1']. Consider for any positive $\varepsilon > 0$ and the closed interval $[\varepsilon, 1-\varepsilon]$. Under assumptions [A1'][A2'], the later-used Taylor expansions have their multipliers on the remainder terms continuous functions bounded away from ∞, and on the closed interval $[\varepsilon, 1-\varepsilon]$, Lipschitz continuous with a uniform upper bound. The remainders across $[\varepsilon, 1-\varepsilon]$ are comparably small.

First, solve for the F.O.C. (with respect to K_i^*) and obtain the following

$$\frac{h(1-a_{i-1}^*) - h(1-a_i^*)}{h(a_i^*) - h(a_{i-1}^*)} = \frac{u_1\pi}{u_0(1-\pi)}\frac{F_1'(K_i^*)}{F_0'(K_i^*)} \tag{A.16}$$

and see by the definition of $\tilde{a}(K)$ that the right-hand-side term is $\frac{h'(1-\tilde{a}(K_i^*))}{h'(\tilde{a}(K_i^*))}$. Hence it can be rewritten as

$$\frac{h(1-a_{i-1}^*) - h(1-a_i^*)}{h(a_i^*) - h(a_{i-1}^*)} = \frac{h'(1-\tilde{a}(K_i^*))}{h'(\tilde{a}(K_i^*))} \tag{F.O.C.}$$

Now recall the result (A.14) obtained earlier in the proof. Let $g_1(x) = h(1-x)$, $g_2(x) = h(x)$, $x_1 = a^*_{i-1}$, $x_2 = a^*_i$, $x = \tilde{a}(K^*_i)$. Hence

$$(a^*_i - \tilde{a}(K^*_i)) - (\tilde{a}(K^*_i) - a^*_{i-1}) = -2\Gamma_{i-1} \cdot (a^*_i - a^*_{i-1})^2 + o(|a^*_i - a^*_{i-1}|^3) \quad \text{(A.17)}$$

where $\Gamma_{i-1} := \Gamma(a^*_{i-1}, a^*_i, \tilde{a}(K^*_i); h(1-\cdot), h(\cdot)) \to \bar{\Gamma}(a) = \frac{1}{24}\frac{d}{da}(\ln|h'(a)h''(1-a) + h'(1-a)h''(a)|)$, as $n \to +\infty$.

Next, solve for the obedience constraint of [P1']. By the definition of $\tilde{a}(K)$ (and $\tilde{K}(a)$), $\frac{h'(1-a^*_i)}{h(a^*_i)} = \frac{u_1\pi}{u_0(1-\pi)}\frac{F'_1(\tilde{K}(a^*_i))}{F'_0(\tilde{K}(a^*_i))}$, and hence the constraint can be rewritten as

$$\frac{F_1(K^*_{i+1}) - F_1(K^*_i)}{F_0(K^*_{i+1}) - F_0(K^*_i)} = \frac{F'_1(\tilde{K}(a^*_i))}{F'_0(\tilde{K}(a^*_i))} \quad \text{(Obedience)}$$

or alternatively

$$\frac{F_1(\tilde{K}(\tilde{a}(K^*_{i+1}))) - F_1(\tilde{K}(\tilde{a}(K^*_i)))}{F_0(\tilde{K}(\tilde{a}(K^*_{i+1}))) - F_0(\tilde{K}(\tilde{a}(K^*_i)))} = \frac{F'_1(\tilde{K}(a^*_i))}{F'_0(\tilde{K}(a^*_i))} \quad \text{(A.18)}$$

Again, recall the result (A.14) shown earlier in the proof. Let $g_1(x) = F_1(\tilde{K}(x))$, $g_2(x) = F_0(\tilde{K}(x))$, $x_1 = \tilde{a}(K^*_i)$, $x_2 = \tilde{a}(K^*_{i+1})$, $x = a^*_i$. Hence

$$(\tilde{a}(K^*_{i+1}) - a^*_i) - (a^*_i - \tilde{a}(K^*_i)) = -2\Xi_i \cdot (\tilde{a}(K^*_{i+1}) - \tilde{a}(K^*_i))^2 + o(|\tilde{a}(K^*_{i+1}) - \tilde{a}(K^*_i)|^3) \quad \text{(A.19)}$$

where $\Xi_i := \Gamma(\tilde{a}(K^*_i), \tilde{a}(K^*_{i+1}), a^*_i; F_1(\tilde{K}(\cdot)), F_0(\tilde{K}(\cdot))) \to$
$\bar{\Xi}(a) := \frac{1}{24}\frac{d}{da}(\ln\{|F'_1(\tilde{K}(a))F''_0(\tilde{K}(a)) - F''_1(\tilde{K}(a))F'_0(\tilde{K}(a))|(\tilde{K}'(a))^3\}$ as $n \to +\infty$.

By (A.17) and (A.19),

$$\begin{aligned}
&(\tilde{a}(K^*_{i+2}) - \tilde{a}(K^*_{i+1})) - (\tilde{a}(K^*_{i+1}) - \tilde{a}(K^*_i)) \\
&= -2\Big((\tilde{a}(K^*_{i+1}) - \tilde{a}(K^*_i))^2\Xi_i + (\tilde{a}(K^*_{i+2}) - \tilde{a}(K^*_{i+1}))^2\Xi_{i+1} \\
&\qquad + 2(a^*_i - a^*_{i-1})^2\Gamma_i + o(|\cdot|^3)\Big) \qquad \text{(A.20)}
\end{aligned}$$

and hence if define $i_z = \min\{i : \tilde{a}(K_i^*) \geqslant z\}$ ($z \in (0,1)$), then for $0 < s < t < 1$,

$$
\begin{aligned}
&(\tilde{a}(K_{i_t+1}^*) - \tilde{a}(K_{i_t}^*)) - (\tilde{a}(K_{i_s+1}^*) - \tilde{a}(K_{i_s}^*)) \\
&= -2\Big(\sum_{i=i_s}^{i_t-1}(\tilde{a}(K_{i+1}^*) - \tilde{a}(K_i^*))^2\Xi_i + \sum_{i=i_s}^{i_t-1}(\tilde{a}(K_{i+2}^*) - \tilde{a}(K_{i+1}^*))^2\Xi_{i+1} \\
&\quad + 2\sum_{i=i_s}^{i_t-1}(a_i^* - a_{i-1}^*)^2\Gamma_i + o(|\cdot|^2)\Big)
\end{aligned}
\tag{A.21}
$$

Again, by (A.17) and (A.19),

$$
(a_{i+1}^* - a_i^*) - (\tilde{a}(K_{i+1}^*) - \tilde{a}(K_i^*)) = -2\Big((\tilde{a}(K_{i+1}^*) - \tilde{a}(K_i^*))^2\Xi_i + (a_i^* - a_{i-1}^*)^2\Gamma_i + o(|\cdot|^3)\Big)
\tag{A.22}
$$

and hence

$$
(a_{i_t+1}^* - a_{i_t}^*)/(\tilde{a}(K_{i_t+1}^*) - \tilde{a}(K_{i_t}^*)) \to 1 \tag{A.23}
$$

Now, let $0 < x < y < 1$, and examine

$$\omega(x,y) := \frac{\frac{d}{da}\hat{\delta}(x)}{\frac{d}{da}\hat{\delta}(y)} \tag{A.24}$$

$$= \frac{\lim_{n\to\infty}(1/n)/(\tilde{a}(K^*_{i_x}) - \tilde{a}(K^*_{i_x-1}))}{\lim_{n\to\infty}(1/n)/(\tilde{a}(K^*_{i_y}) - \tilde{a}(K^*_{i_y-1}))}$$

$$= \lim_{n\to\infty}\frac{(\tilde{a}(K^*_{i_y}) - \tilde{a}(K^*_{i_y-1}))}{(\tilde{a}(K^*_{i_x}) - \tilde{a}(K^*_{i_x-1}))}$$

$$= 1 - 2\Big(\lim_{n\to\infty}\sum_{i=i_x}^{i_y-1}(\tilde{a}(K^*_{i+1}) - \tilde{a}(K^*_i))(\frac{\tilde{a}(K^*_{i+1}) - \tilde{a}(K^*_i)}{\tilde{a}(K^*_{i_x}) - \tilde{a}(K^*_{i_x-1})})\Xi_i + ...$$

$$... + \lim_{n\to\infty}\sum_{i=i_x}^{i_y-1}(\tilde{a}(K^*_{i+2}) - \tilde{a}(K^*_{i+1}))(\frac{\tilde{a}(K^*_{i+2}) - \tilde{a}(K^*_{i+1})}{\tilde{a}(K^*_{i_x}) - \tilde{a}(K^*_{i_x-1})})\Xi_{i+1} + ...$$

$$... + \lim_{n\to\infty} 2\sum_{i=i_x}^{i_y-1}(a^*_i - a^*_{i-1})(\frac{a^*_i - a^*_{i-1}}{\tilde{a}(K^*_{i_x}) - \tilde{a}(K^*_{i_x-1})})\Gamma_i + \lim_{n\to\infty} o(|\cdot|)\Big) \tag{A.25}$$

$$= 1 - \frac{1}{6}\Big(\int_x^y \frac{d}{da}\ln(-h'(a)h''(1-a) - h'(1-a)h''(a))\omega(x,a)da + ...$$

$$... + \frac{d}{da}(\ln\{|F_1'(\tilde{K}(a))F_0''(\tilde{K}(a)) - F_1''(\tilde{K}(a))F_0'(\tilde{K}(a))|(\tilde{K}'(a))^3\}\Big) \tag{A.26}$$

Differentiate ω w.r.t. y and get an ODE:

$$\frac{d\omega}{dy} = -\frac{1}{6}\frac{d}{dy}\ln(-h'(y)h''(1-y) - h'(1-y)h''(y))\omega$$

$$-\frac{1}{6}\frac{d}{dy}\ln(|F_1'(\tilde{K}(y))F_0''(\tilde{K}(y)) - F_1''(\tilde{K}(y))F_0'(\tilde{K}(y))|(\tilde{K}'(y))^3)\omega \tag{A.27}$$

The solution is

$$\omega(x,y) = C_0\Big((-h'(y)h''(1-y) - h'(1-y)h''(y))\cdot$$

$$(|F_1'(\tilde{K}(y))F_0''(\tilde{K}(y)) - F_1''(\tilde{K}(y))F_0'(\tilde{K}(y))|(\tilde{K}'(y))^3))\Big)^{-\frac{1}{6}} \tag{A.28}$$

where C_0 is a constant. Pin down

$$C_0 = \Big((-h'(x)h''(1-x) - h'(1-x)h''(x)) \cdot$$

$$(|F_1'(\tilde{K}(x))F_0''(\tilde{K}(x)) - F_1''(\tilde{K}(x))F_0'(\tilde{K}(x))|(\tilde{K}'(x))^3)) \Big)^{\frac{1}{6}} \tag{A.29}$$

by $\omega(x, x) = 1$. Hence

$$\hat{\delta}(a) \propto \Big((-h'(a)h''(1-a) - h'(1-a)h''(a)) \cdot$$

$$(|F_1'(\tilde{K}(a))F_0''(\tilde{K}(a)) - F_1''(\tilde{K}(a))F_0'(\tilde{K}(a))|(\tilde{K}'(a))^3)) \Big)^{\frac{1}{6}} \tag{A.30}$$

And notice that $(\tilde{K}'(a))^{-1} = \tilde{a}'(\tilde{K}(a))$. Hence

$$\beta_\infty(K) \propto \Big((-h'(\tilde{a}(K))h''(1-\tilde{a}(K)) - h'(1-\tilde{a}(K))h''(\tilde{a}(K))) \cdot$$

$$(|F_1'(K)F_0''(K) - F_1''(K)F_0'(K)|) \Big)^{\frac{1}{6}} (\tilde{a}'(K))^{\frac{1}{2}} \tag{A.31}$$

Notice that this is derived on $[\varepsilon, 1-\varepsilon]$ for any $\varepsilon > 0$. Hence it holds for $a \in (0, 1)$.

If $K|\theta \sim N(\mu(2\theta - 1), \sigma^2)$, then $|F_1'(K)F_0''(K) - F_1''(K)F_0'(K)|^{1/6} \propto \exp(-\frac{K^2}{6\sigma^2})$, a $N(0, 3\sigma^2)$ Gaussian density. □

A.3 Proofs of Remarks 1, 2 and Proposition 1.4.5

First show Remarks 1 and 2.

Proof. (Remark 1) (1). By definition, $\frac{h'(1-a)}{h'(a)} = \frac{u_1\pi}{u_0(1-\pi)} \exp(\frac{2\mu K}{\sigma^2})$, hence

$$\tilde{K}(a) = \frac{\sigma^2}{2\mu} \left(\ln h'(1-a) - \ln h'(a) - \ln \frac{u_1\pi}{u_0(1-\pi)} \right) \tag{A.32}$$

Thus $\tilde{K}'(a) = \frac{\sigma^2}{2\mu} \left(-\frac{h''(1-a)}{h'(1-a)} - \frac{h''(a)}{h'(a)} \right)$; substitute it into $\tilde{a}(K) = (\tilde{K}(\tilde{a}(K)))^{-1}$. Q.E.D.

(2). it is a trivial corollary of (1). □

Proof. (Remark 2). By Remark 1, $\tilde{a}(K^* + \delta) + \tilde{a}(K^* - \delta) = 1$. Evaluate $\lambda_h(K)$ at $K^* + \delta$ and $K^* - \delta$ and using that result, getting $\lambda_h(K^* + \delta) = \lambda_h(K^* - \delta)$. □

Next show Proposition 1.4.5.

Proof. (Proposition 1.4.5). (1). Without loss of generality, only need to show the case when $\frac{\pi u_1}{(1-\pi)u_0} < 1$, i.e. $K^* > 0$. Notice that $\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}$ is symmetric around K^*, and by assumption, its value is higher if K is closer to K^*. Then $\beta'_\infty(K) = \exp(-\frac{K^2}{6\sigma^2})\left((\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}\right) \geqslant \exp(-\frac{K^2}{6\sigma^2})\left((\lambda_h(-K))^{\frac{1}{6}}(\tilde{a}'(-K))^{\frac{1}{2}}\right) = \beta'_\infty(-K)$. The inequality is because the distance between K^* and K is closer than between K^* and $-K$.

(2). $(\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}$ is increasing iff. $\ln\left((\lambda_h(K))^{\frac{1}{6}}(\tilde{a}'(K))^{\frac{1}{2}}\right)$ is increasing. Take its derivative with respect to K on interval I and recognizing $\tilde{a}'(K) > 0$, thus obtaining the result. □

Appendix B

Appendix to Chapter 2

Lemma 1. (monotone comparative statics) (i). K^*_A is continuous and strictly increasing in A; (ii). K^*_{nobs} is increasing in the distribution of A (in the FOSD order).

Proof Both are standard monotone comparative statics results. (i). K^*_A is the solution to $v_K(K, A) = R_0$ with $v_K(K, A)$ decreasing in K. Due to complementarity, increasing A will point-wise raise $v_K(K, A)$ as a function of K, hence increasing the solution to $v_K(K, A) = R_0$. (ii). K^*_{nobs} is the solution to $E[v_K(K, A)] = R_0$ with $v_K(K, A)$ being an increasing function in A due to complementarity. Hence for F.O.S.D. higher distribution of A, $E[v_K(K, A)]$ is point-wise higher as a function of K. Since $E[v_K(K, A)]$ is a decreasing function in K, a point-wise increase in the function will raise the solution to $E[v_K(K, A)] = R_0$.

Proposition 1.1 (cutoff strategy) For any given π_i, the disclosure strategy of company i features a cutoff A^* with $D = \{A \in supp(f_p) | A > A^*\}$. In other words, there exists A^* such that company i discloses almost surely iff. $A > A^*$.

Proof of Proposition 1.1 Suppose for the sake of contradiction that there exists sets $S_1, S_2 \subset supp(f_p)$ with $\Pr_p(S_1), \Pr_p(S_2) > 0$, such that (i). $\forall A_1 \in S_1$ and $A_2 \in S_2$, $A_1 < A_2$, and (ii). $S_1 \subset D$ and $S_2 \subset ND$. Because $A_1 \in D$, the company's payoff is higher if disclosing A_1 than if remaining silent.

On the other hand, notice that when making a disclosure, the expected payoff of the company is $(1 - \pi)u(K^*_{nobs}, A) + \pi u(K^*_A, A)$. Because A_2 is bigger than A_1 and hence induces a higher level of first-best investment from the investor, the company's expected payoff is at least as high when disclosing A_2 as when disclosing A_1. Hence the company should prefer to disclose A_2 rather than to stay silent. That contradicts the assumption that $A_2 \in ND$.

Proposition 1.2 (equilibrium cutoff) (i). The cutoff A_π^* is a solution of $\bar{A}$ for

$$u(K_{\bar{A}}^*, \bar{A}) = u(K_{nobs|\pi}^*, \bar{A}) \tag{B.1}$$

(ii). Equivalently, it is a solution for

$$K_{\bar{A}}^* = K_{nobs|\pi}^* \tag{B.2}$$

(iii). The investor "assumes the worst" upon seeing no disclosure, i.e.

$$K_{nobs|\pi}^* = \min_{\bar{A}} K_{nobs|\pi,\bar{A}}^* \tag{B.3}$$

(iv). The cutoff A_π^* exists and is unique.

Proof (i) holds naturally for the marginal type. Now show (ii). Same K under same A will lead to same values of $u(K, A)$. Also because $u(K, A)$ is strictly monotone in K, same values of $u(K, A)$ with same A indicates equal K.

Now show (iii). Examine $K_{nobs|\pi,\bar{A}}^*$ as a function of $\bar{A}$. $K_{nobs|\pi,\bar{A}}^*$ is the solution to

$$\int_A \frac{(1-\pi)f(A) + \pi f(A)\mathbf{1}_{A \leqslant \bar{A}}}{1 - \pi + \pi F(\bar{A})} v_K(K, A) dA = R_0 \tag{B.4}$$

(as an equation of K). Denote the left-hand side as $L = \int_A G(\bar{A}, A) v_K(K, A) dA$. Take derivative with respect to $\bar{A}$ on both sides, getting

$$\int_A \left(\frac{dG}{d\bar{A}} v_K(K, A) + G(\bar{A}, A) v_{KK}(K, A) \frac{dK}{d\bar{A}} \right) dA = 0$$

At the minimum point $\bar{A}$ of $K_{nobs|\pi,\bar{A}}^*$, there is $\frac{dK}{dA} = 0$, and hence $\int_A \frac{dG}{dA} v_K(K, A) dA = 0$. Simplify the expression, getting

$$v_K(K, \bar{A}) = \int_A \frac{(1-\pi)f(A) + \pi f(A)\mathbf{1}_{A \leqslant \bar{A}}}{1 - \pi + \pi F(\bar{A})} v_K(K, A) dA \tag{B.5}$$

while RHS is equal to R_0. Hence the minimum point of $K^*_{nobs|\pi,\bar{A}}$ satisfies $v_K(K,\bar{A}) = R_0$, the F.O.C. for $K^*_{\bar{A}}$. In other words, as functions of $\bar{A}$, $K^*_{nobs|\pi,\bar{A}}$ crosses $K^*_{\bar{A}}$ at its own minimum point and hence that minimum point is the equilibrium cutoff.

Now show (iv). Denote the second-best investment amount upon under prior belief is K^*_{prior}. Without loss of generality assume the lower bound of $supp(A)$ is 0 and the upper bound is ∞. It is obvious that $K^*_0 < K^*_{prior} < K^*_\infty$. For $\bar{A} \to 0$ and $\bar{A} \to \infty$, $K^*_{nobs|\pi,\bar{A}} \to K^*_{prior}$. $K^*_{nobs|\pi,\bar{A}}$ is a continuous function of $\bar{A}$ due to the absolute continuity of the prior. $K^*_{\bar{A}}$ is strictly increasing and continuous. Hence they must intersect.

Now show they must only intersect once. Following the proof of (iii), at the intersection, there must be $v_K(K,\bar{A}) = R_0$ as a point on curve $K^*_{\bar{A}}$, and hence $\int_A \frac{dG}{d\bar{A}} v_K(K,A)dA = 0$. Also at the intersection there must also be

$$\int_A \left(\frac{dG}{d\bar{A}} v_K(K,A) + G(\bar{A},A) v_{KK}(K,A) \frac{dK}{d\bar{A}} \right) dA = 0$$

as a point on curve $K^*_{nobs|\pi,\bar{A}}$ Consequently, there must be

$$\left(\int_A G(\bar{A},A) v_{KK}(K^*_{nobs|\pi,\bar{A}}, A) d\bar{A} \right) \frac{dK^*_{nobs|\pi,\bar{A}}}{d\bar{A}} = 0$$

Because $v_{KK} < 0$, the first term is negative, and for the equation to hold there must be $\frac{dK^*_{nobs|\pi,\bar{A}}}{d\bar{A}} = 0$. That must hold for all intersections. For it to happen, it is only possible that the curves only intersect once. Hence the equilibrium cutoff exists and is unique.

Proposition 1.3 (revelation increases with attention) (i). A^*_π is strictly decreasing in π. (ii). A^*_1 is the lower bound of $supp(A)$; $\lim_{\pi \to 0^+} A^*_\pi$ solves $K^*_A = K^*_{prior}$.

Proof (i). For any given $\bar{A}$, the posterior belief in A decreases in the FOSD order as π increases. Hence by Lemma 1, $K^*_{nobs|\pi,\bar{A}}$ as a function of $\bar{A}$ is point-wise lower.

Then following the proof of Proposition 1.2, the intersection of $K^*_{\bar{A}}$ and $K^*_{nobs|\pi,\bar{A}}$ is strictly lower, thus finishing the proof.

(ii). When $\pi = 1$, withholding information means the company is for sure below the cutoff type, and the investment will be lower than or equal to the first-best investment at the cutoff; for the equality to hold it is only possible that the cutoff is the lowest type. When $\pi \to 0^+$, $K^*_{nobs|\pi,\bar{A}} \to K^*_{prior}$ point-wise as a function of $\bar{A}$, hence the intersection point converges to the solution of $K^*_{\bar{A}} = K^*_{prior}$.

Theorem 1 (i). (monotonicity) $V(\pi)$ and $EVI(\pi)$ increase in π.
(ii). (convexity) $V(\pi)$ and $EVI(\pi)$ are strictly convex in π.
(iii). (attention allocation) In the equilibrium attention allocation $\{\pi_i\}_{i\in I}$, π_i is equal to either 1 or 0. The investor will calculate $\rho_i = \frac{EVPI_i}{c_i}$, sort from high to low, and pay attention resource to companies with the highest ρ_i until the attention resource is exhausted.

Proof. (i). There are two proofs. First approach: see the main text for the Blackwell-based approach of $S_{\Delta\pi}$. Second approach: see the first-order approach in the main text. The derivative is $\frac{dV(\pi)}{d\pi} = \frac{dV}{d\pi} + \frac{dV}{dA^*_\pi}\frac{dA^*_\pi}{d\pi} + \frac{dV}{dK^*_{nobs|\pi}}\frac{dK^*_{nobs|\pi}}{d\pi}$. The first term is positive. The third term is 0 by Envelope Theorem. The second term is zero, because

$$\frac{dV}{dA^*_\pi} = \pi f(A^*_\pi)\left(v(K^*_{A^*_\pi}, A^*_\pi) - v(K^*_{nobs|\pi}, A^*_\pi)\right) \tag{B.6}$$

and by Proposition 1.2 $K^*_{A^*_\pi} = K^*_{nobs|\pi}$.

(ii). Due to optimality of $K^*_{nobs|\pi}$, I can obtain and regroup its F.O.C. as

$$\int_{A>A*} \frac{dv(K^*_{nobs|\pi}, A)}{dK^*_{nobs|\pi}} dF(A) = \frac{1}{\pi}\left(\int_A \frac{dv(K^*_{nobs|\pi}, A)}{dK^*_{nobs|\pi}} dF(A)\right)$$

. When $\pi = 0$, the right-hand-side term is zero, because it is precisely the F.O.C. of $K^*_{nobs|0}$. When π is positive, the right-hand side is positive. That is because $K^*_{nobs|\pi} < K^*_{nobs|0}$ (because the posterior of (A^*_π, π) is worse than $(A^*_\pi, 0)$, or the prior F), and hence for any A, $\frac{dv(K^*_{nobs|\pi}, A)}{dK^*_{nobs|\pi}} > \frac{dv(K^*_{nobs|0}, A)}{dK^*_{nobs|0}}$ (because of convexity). Therefore the sign of the first bracket is positive. To summarize, $\frac{d^2V(\pi)}{d\pi^2} > 0$ for $\pi > 0$ and $V(\pi)$ is strictly convex.

(iii). A natural corollary of (i)(ii).

Proposition 2.1 (i). There exists BNE for simultaneous attention allocation and disclosure. In any BNE, the attention allocation $\{\pi_i\}_{i \in I}$ is concentration. (ii). the equilibrium attention allocation in the PBE for sequential attention allocation and disclosure supports the BNE that maximizes payoff for the investor.

Proof. (i). For existence see proof of (ii). Now show concentration by examining best responses in BNE. Denote the mixed strategy of attention paid to a given firm as $\lambda(\pi)$, a probability cdf for $\pi \in [0, 1]$.

First, show that for any strategy $\lambda(\pi)$ of the investor, the best response of the firms is to adopt a cutoff strategy with $D_\pi = \{A | A \geqslant A^*_\pi\}$. Suppose, for the sake of contradiction, that $\exists$ positive probability sets $S_1 \subset D_\pi$, $S_2 \subset ND_\pi$ such that $A_1 < A_2$ for all $A_1 \in S_1, A_2 \in S_2$. Then for any given $\lambda(\pi)$, on one hand, company's expected payoff is higher to disclose A_1 than to withhold, and is lower to disclose A_2 than to withhold. However, the expected payoff is higher for A_2 than A_1, regardless of $\lambda(\pi)$, a contradiction. Hence the strategy is a cutoff one.

Second, for any cutoff strategy of the firms, the dominant strategy of the investor is a concentration strategy. That is because the $EVI(\pi)$ or $V(\pi)$ is still strictly convex. Denote the second-best investment with posterior of A^* as $K^*_{nobs;A^*}$. Since

A^* is taken as given and will not change with π,

$$\frac{dV(\pi)}{d\pi} = \frac{dV}{d\pi} + \frac{dV}{dK^*_{nobs;A*}}\frac{dK^*_{nobs;A*}}{d\pi} \tag{B.7}$$

The first term is equal to $\mathrm{E}[\Delta(A)]$ where $\Delta(A) = v(K^*_A, A) - v(K^*_{nobs;A*}, A)$. The second term is 0, because of the Envelope condition. The first-order expression is positive, meaning expected utility will increase with π due to fewer misses for good types, even if no more types are revealed. Further taking the second derivative:

$$\frac{d^2V(\pi)}{d\pi^2} = -\left(\int_{A>A*} \frac{dv(K^*_{nobs;A*}, A)}{dK^*_{nobs;A*}} dF(A)\right)\left(\frac{dK^*_{nobs;A*}}{d\pi}\right) \tag{B.8}$$

The reason for its concavity is the same as EVI convexity in the sequential setting in the benchmark model.

Finally, for any concentration strategy of the investor, the best response of the firms is to either fully disclose when there is positive probability that it will be selected in the subset, or to hide when there is zero probability. For a firm, if the investor pays $\pi = 1$ with chance λ and $\pi = 0$ with $1 - \lambda$, then the expected payoff of full disclosure is $\lambda u(K^*_A, A) + (1-\lambda)u(K^*_{nobs|0}, A)$. The expected payoff for any other cutoff strategy is $\lambda\Big(\Pr(D_\pi)\mathrm{E}[u(K^*_A, A)|D] + \Pr(ND_\pi)\mathrm{E}[u(K^*_{nobs|1}, A)|ND]\Big) + (1 - \lambda)u(K^*_{nobs|0}, A)$. The former is always bigger than the latter. Hence the BNE always involve a concentration strategy.

(ii). The concentration strategy may not not unique. Now show that the strategy as the equilibrium of PBE in the sequential setting forms a BNE. One just needs to show that the investor cannot be better-off deviating from his original attention allocation plan after disclosure decisions are made. By the proof in (i), a concentration strategy is the best response for any given cutoff disclosure strategies. Hence, suppose there is any profitable derivation, then there must exist another different concentration strategy that is a profitable deviation. That is impossible due to the fact that PBE already selected those with the highest sets of $\frac{EVPI}{c}$. Deviation will sacrifice the highest $\frac{EVPI}{c}$ and harvest some partial-revelation $\frac{EVI}{c}$ for companies with lower $\frac{EVPI}{c}$, decreasing expected utility. Hence there is no profitable deviation and PBE strategy also forms a BNE.

PBE yields value for the investor no lower than any BNE because of the first-mover advantage of the investor.

Bibliography

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.

Bénabou, 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.

Bertomeu, J., Hu, K. P., and Liu, Y. (2020). Disclosure and investor inattention. *Available at SSRN 3673225*.

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.

Di Pei, H. (2015). Communication with endogenous information acquisition. *Journal of Economic Theory*, 160:132–149.

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

Eeckhout, J., Persico, N., and Todd, P. E. (2010). A theory of optimal random crackdowns. *American Economic Review*, 100(3):1104–35.

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*.

Gould, J. P. (1974). Risk, stochastic preference, and the value of information. *Journal of Economic Theory*, 8(1):64–84.

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

Hess, J. (1982). Stochastic preference and the value of information. *Journal of Economic Theory*, 27(1):231–238.

Hirshleifer, D. A., Lim, S. S., and Teoh, S. H. (2004). Disclosure to an audience with limited attention. *Available at SSRN 604142*.

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

Lawrence, D. B. (2012). *The Economic Value of Information*. Springer Science & Business Media.

Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, 121(3):1029–1061.

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*.

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

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

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

Tetlock, P. C. (2007). Giving content to tnvestor 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.