Investor Inattention, Information, and Firm Investment

Jinge Liu*

November 20, 2021

Abstract

An investor with limited attention resources 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 with an attention cost. Because benefit to attention is convex, the investor will choose to optimally focus on a subset of firms and acquire full information while giving up learning the rest. Firms in the scrutinized subset are with low investigation costs and high Expected Value of Perfect Information (EVPI) and always receive efficient investments. The other firms are invested in the most inefficiently and present maximum asymmetry in information transparency. The result rationalizes the use of convertible debt as a socially optimal financing instrument for private firms. It can be applied to analyze a range of investment relationships, such as between a VC and start-ups or an LP and GPs.

*Fuqua School of Business, Duke University. jinge.liu@duke.edu. I thank my co-chairs Simon Gervais and S. "Vish" Viswanathan for their invaluable advice and support. I thank my committee members Felipe Varas and Arjada Bardhi for their invaluable support and suggestions. Comments are welcome. For the most recent version, please visit jingeliu.com. All remaining errors are mine.

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.

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, Lim and Teoh (2004) and the information-theoretic model of Bertomeu, Hu and Liu (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 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, Persico, and Todd (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 A Model of Disclosure and Attention Allocation

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 $\xi_i(K_i, A_i) \geq 0$, and the company will retain payoff $u_i(K_i, A_i) := Y_i(K_i, A_i) - \xi_i(K_i, A_i) \geq 0$. 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 $w \in (0, 1)$, $\xi_i\left(wK_i^{(1)} + (1-w)K_i^{(0)}, A_i\right) \leq w\xi_i\left(K_i^{(1)}, A_i\right) + (1-w)\xi_i\left(K_i^{(0)}, A_i\right)$.

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, $\int_i c_i \pi_i g(i) di \leq d < \int_i c_i g(i) di$. 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 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.

2.2.1 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 [Proposition 1.1]. (Proof is in [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

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

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 $\bar{A}$:*

$$u(K^*_{\bar{A}}, \bar{A}) = u(K^*_{nobs|\pi}, \bar{A}) \tag{2}$$

(ii). Equivalently, it is a solution for

$$K^*_{\bar{A}} = K^*_{nobs|\pi} \tag{3}$$

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

$$K^*_{nobs|\pi} = \min_{\bar{A}} K^*_{nobs|\pi,\bar{A}} \tag{4}$$

(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 $K^*_{\bar{A}} = K^*_{nobs|0}$.*

2.2.2 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:

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

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

$$
\begin{cases}
\Pr(S_{\Delta\pi} = A|A) = \frac{\pi_2-\pi_1}{1-\pi_1}, & \text{for } A > A^*_{\pi_1} \\
\Pr(S_{\Delta\pi} = A|A) = \pi_2, & \text{for } A^*_{\pi_2} < A \leq A^*_{\pi_1} \\
\Pr(S_{\Delta\pi} = A|A) = 0, & \text{for } A \leq A^*_{\pi_2}
\end{cases}
\tag{5}
$$

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

$$
\frac{dV(\pi)}{d\pi} = \frac{dV}{d\pi} + \frac{dV}{dA^*_\pi}\frac{dA^*_\pi}{d\pi} + \frac{dV}{dK^*_{nobs|\pi}}\frac{dK^*_{nobs|\pi}}{d\pi} \tag{6}
$$

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.

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

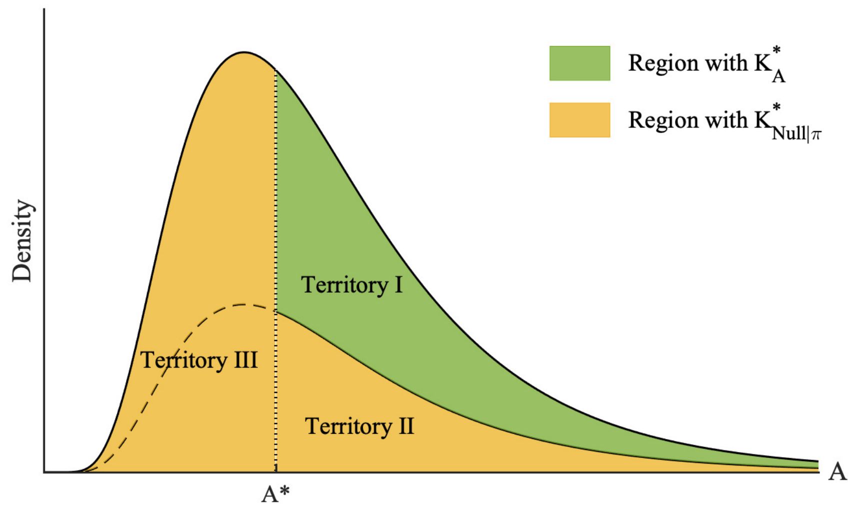


Figure 1: Regions of the prior belief of A

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.,

$$\frac{dV(\pi)}{d\pi} = \int_{A>A^*_\pi} \Delta_\pi(A) dF(A) \tag{7}$$

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

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

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

$$\int_{i \in J} E[Y(K_A^*, A) - R_0 K_A^*] g(i) di + \int_{i \notin J} E[Y(K_{nobs|0}^*, A) - R_0 K_{nobs|0}^*] g(i) di \tag{9}$$

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.

2.3 Robustness and Comparative Statics

2.3.1 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.

2.3.2 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

$$EVPI = \int_A \big(v(K_A^*, A) - v(K_{nobs|0}^*, A)\big) f(A)dA \tag{10}$$

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). $\xi_{AA} \leq 0$, (ii). $\xi_{AA}\xi_{KK} \leq \xi_{AK}^2$.*

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

$$EVPI = \int_A \left(v(K_A^*, A) - v(K_{nobs|0}^*, A)\right) \frac{1}{\sigma} f_X(\frac{A-\mu}{\sigma}) dA \tag{11}$$

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

$$EVPI = \left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right) \left((R_0/w)^{\frac{\alpha}{\alpha-1}}\right) \left(E[A^{\frac{1}{1-\alpha}}] - (E[A])^{\frac{1}{1-\alpha}}\right) \tag{12}$$

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

$$EVPI \propto \frac{1}{2e} \frac{1-\alpha}{2-\alpha} \left[(\mu + e)^{\frac{2-\alpha}{1-\alpha}} - (\mu - e)^{\frac{2-\alpha}{1-\alpha}}\right] - \mu^{\frac{1}{1-\alpha}} \tag{13}$$

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

$$\frac{dEVPI}{d\mu} \propto E[(A^{\frac{1}{1-\alpha}})'] - (\mu^{\frac{1}{1-\alpha}})' \tag{14}$$

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.

2.4 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.

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 $Y(K, A) - R_0 K$ and $\mathrm{E}[Y(K, A)] -$

$R_0 K$, 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 $U_h(1) \geq U_h(0)$. There are another $(1-\beta)$ firms not receiving attention (group l, for low EVPI). They have a lower EVPI (denote as δ_l; $\delta_l \leq \delta_h$) and optimally choose to hide away from attention with $U_l(1) \leq U_l(0)$.

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 $EVPI \geq \delta_h$, the $U_h(1)$ is the highest $U(1)$, hence shutting down further motives to compete for attention. Similarly, among all legit contracts with $EVPI \leq \delta_l$, 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*

$$\xi^{(h)}(K, A) = \begin{cases} R_0 K & \textit{for } K \leq \hat{K}_A^* < \hat{K}_{nobs|0}^* \\ Y(K, A) - (Y(\hat{K}_A^*, A) - R_0 \hat{K}_A^*) & \textit{for } \hat{K}_A^* \leq K < \hat{K}_{nobs|0}^* \\ Y_K(K_{nobs|0}^*, A) K & \textit{for } K < \hat{K}_{nobs|0}^* \leq \hat{K}_A^* \\ Y(K, A) - \left(Y(\hat{K}_{nobs|0}^*, A) - Y_K(K_{nobs|0}^*, A) \hat{K}_{nobs|0}^* \right) & \textit{for } K > \hat{K}_{nobs|0}^* \end{cases} \tag{15}$$

for the high EVPI group of measure β *and*

$$\xi^{(l)}(K,A)=\begin{cases} R_0K & \text{for } K\le \hat{K}^*_A<\hat{K}^*_{nobs|0}\\ Y(K,A)-(Y(\hat{K}^*_A,A)-R_0\hat{K}^*_A) & \text{for } \hat{K}^*_A\le K<\hat{K}^*_{nobs|0}\\ Y_K(\hat{K}^*_{nobs|0},A)K & \text{for } K<\hat{K}^*_{nobs|0}\le\hat{K}^*_A\\ Y_K(\hat{K}^*_{nobs|0},A)\hat{K}^*_{nobs|0}+Y_K(\hat{K}^*_A,A)(\hat{K}^*_A-\hat{K}^*_{nobs|0}) & \text{for } \hat{K}^*_{nobs|0}<K\le\hat{K}^*_A \end{cases} \tag{16}$$

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, $\xi_K(\hat{K}^*_{nobs|0},A)\le Y_K(\hat{K}^*_{nobs|0},A)$. 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.

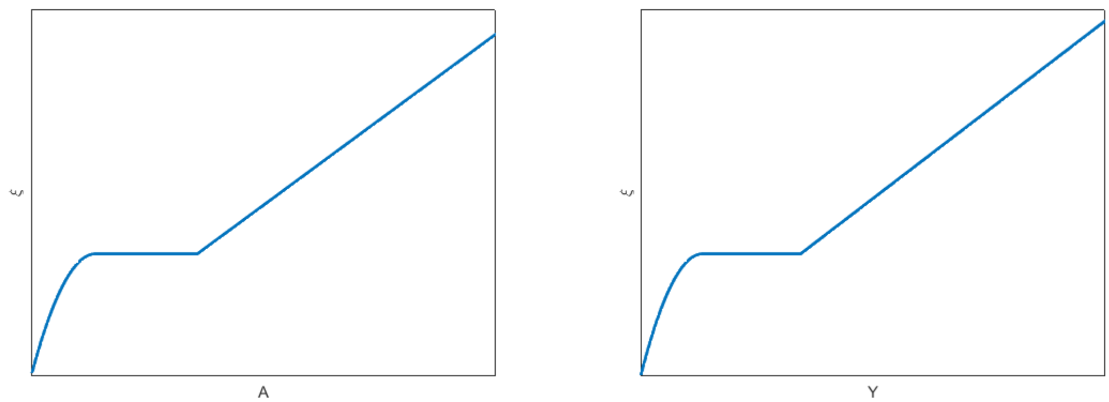
From the conventional perspective of security design, I interpret the contracts by their payoff diagrams. Fix K, then $\xi(K, A)$ as a function of A for small $K < \hat{K}^*_{nobs|0}$, regardless of attention, is

$$\xi = \begin{cases} Y(K,A) - \Big(Y(\hat{K}^*_A, A) - R_0 \hat{K}^*_A\Big) & \text{for } A \ s.t. \ \hat{K}^*_A < K \\ R_0 K & \text{for } A \ s.t. \ K \leq \hat{K}^*_A < \hat{K}^*_{nobs|0} \\ Y_K(\hat{K}^*_{nobs|0}, A) K & \text{for } A \ s.t. \ \hat{K}^*_A \geq \hat{K}^*_{nobs|0} \end{cases} \tag{17}$$

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.

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 $R_0 K$ with a constant reservation



(a) Investor payoff ξ as a function of A (b) Investor payoff ξ as a function of Y

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

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.

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.

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.

References

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.

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.

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

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.

Robinson, D. T. and Sensoy, B. A. (2013). Do private equity fund managers earn their fees? compensation, ownership, and cash flow performance. *The Review of Financial Studies*, 26(11):2760–2797.

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

Appendix I: Benchmark Model

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

$$u(K^*_{\bar{A}}, \bar{A}) = u(K^*_{nobs|\pi}, \bar{A}) \tag{18}$$

(ii). Equivalently, it is a solution for

$$K^*_{\bar{A}} = K^*_{nobs|\pi} \tag{19}$$

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

$$K^*_{nobs|\pi} = \min_{\bar{A}} K^*_{nobs|\pi,\bar{A}} \tag{20}$$

(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

$$\int_A \frac{(1-\pi)f(A) + \pi f(A)\mathbf{1}_{A\leq\bar{A}}}{1-\pi+\pi F(\bar{A})} v_K(K, A)dA = R_0 \tag{21}$$

(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 $\frac{dK}{d\bar{A}} = 0$, and hence $\int_A \frac{dG}{d\bar{A}}v_K(K, A)dA = 0$. Simplify the expression, getting

$$v_K(K, \bar{A}) = \int_A \frac{(1-\pi)f(A) + \pi f(A)\mathbf{1}_{A\leq\bar{A}}}{1-\pi+\pi F(\bar{A})} v_K(K, A)dA \tag{22}$$

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 $K^*_A = K^*_{prior}$.

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

$$\frac{dV}{dA^*_\pi} = \pi f(A^*_\pi)\left(v(K^*_{A^*_\pi}, A^*_\pi) - v(K^*_{nobs|\pi}, A^*_\pi)\right) \tag{23}$$

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 $D_\pi = \{A | A \geq A^*_\pi\}$. 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,

$$\frac{dV(\pi)}{d\pi} = \frac{dV}{d\pi} + \frac{dV}{dK^*_{nobs;A^*}} \frac{dK^*_{nobs;A^*}}{d\pi} \tag{24}$$

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:

$$\frac{d^2V(\pi)}{d\pi^2} = -\left(\int_{A>A^*} \frac{dv(K^*_{nobs;A^*}, A)}{dK^*_{nobs;A^*}} dF(A)\right)\left(\frac{dK^*_{nobs;A^*}}{d\pi}\right) \tag{25}$$

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.