Ignoring The Left Tail

Kaushalendra Kishore* Nishant Ravi†

December 23, 2024

Abstract

This article studies a principal-agent problem with flexible information acquisition. The agent has to choose between investing in an innovative asset about which costly information can be acquired or in a conventional asset about which there is enough historical data and no information is acquired. In the first-best problem, the principal acquires information about the entire distribution of cash flow from the innovative asset. Contrarily, in the second-best problem, less information is acquired about the left tail of the distribution compared to the first-best. The optimal contract of the agent when the innovative asset is chosen pays zero wage below a cutoff value of cash flow and an increasing wage above the cutoff. As a consequence, the information intensity is zero for these low cash flows where no wage is paid. This results in investment in even those assets that have a thick left tail and thus a high likelihood of failure ex post. Our paper explains why agents make investments without learning about the left tail of the distribution which in turn may lead to a financial crisis.

*Centre for Advanced Financial Research and Learning. Email- kaushalendra.kishore@cafral.org.in
†Indian School of Business. Email- Nishant_Ravi@isb.edu

1 Introduction

A financial crisis is often followed by a technological or a financial innovation (Kindleberger (1978), Goldfarb and Kirsch (2019)). Recent examples include the dot-com bubble and the global financial crisis of 2008 which involved innovative financial products such as collateralized debt obligations (CDOs). One of the explanations for why a financial crisis occurs is that economic agents are irrational and ignore the tail risks in their investment choices because of various biases such as the availability heuristic (Gennaioli et al. (2012), Barberis (2013)) or "blindness to outliers" (Payzan-LeNestour and Woodford (2022)).

However, the financial industry is known to hire talented individuals (Bond and Glode (2014), Glode and Lowery (2016)). Therefore it is surprising that these talented individuals make systematic mistakes, that too in an environment where there are strong incentives to make large profits if their portfolios perform well. In this paper, we try to answer this puzzle and provide an explanation for why rational agents may choose to learn less about the left tail than the right tail of the distribution of investments they make. As a consequence, they end up investing in even those assets that have a fat left tail and therefore a high likelihood of failure ex post.

We build a principal-agent model with flexible information acquisition. The principal ("she") wants to make an investment and has two choices, an asset A with unknown returns about which costly information can be acquired to learn about its distribution or another asset B about which there is nothing to learn. However, the principal does not have the skill to acquire the information about asset A, so she hires an agent ("he") to do so. There are several interpretations of assets A and B. Asset A can be considered an investment in a firm with new technology such as information technology or artificial intelligence, or in a financially innovative product such as CDOs. Asset B can be thought of as old-technology firms and conventional loans about which there is enough historical data and hence there is much less need to acquire costly information, or as safe securities such as government bonds with a very low probability of default.

Since an agent can choose both the intensity and nature of information acquisition, we use the approach of flexible information acquisition to model our information structure (see Yang and Zeng (2019), Yang (2020)).[1] Flexible information acquisition implies that the agent can choose any possible information structure which allows us to model state-contingent attention allocation. Since information acquisition is costly, the contract offered to the agent determines how much effort he exerts in learning about different states. Thus our approach allows us to endogenously model differential attention allocation to different

[1] In traditional models of information acquisition the information structure is exogenously given. See, for example, Lambert (1986), Gromb and Martimort (2007)Veldkamp (2023), among others.

states by the agent and the optimal contract that the principal offers. We model information cost as the reduction in Shannon's entropy of a random variable (Sims (2003)).

The optimal contract of the agent is contingent on the asset chosen by the agent and the cash flow generated by it. The contract and the information structure the agent chooses are endogenously determined in the model. The key friction in our model is that the nature and intensity of information acquired by the agent is unobservable and that the agent has limited liability.

In the benchmark (first-best) case, the principal herself exerts effort to acquire information about asset A. In the benchmark information structure, the principal acquires information about all values in the support of the distribution of cash flows from asset A. This is because she fully internalizes the profit (or loss) from each possible outcome of cash flow generated by asset A.

However, if the agent acquires the information (second-best), then less information is acquired about the left side of the distribution of asset A compared to the first-best. In the optimal contract, if the agent chooses asset A, then for low cash flows the agent is paid zero. If the cash flow is above a certain cutoff, then the agent is paid an increasing concave wage. The expected wage paid if the agent chooses asset B is positive.

The intuition for the shape of the optimal contract offered by the principal and the lack of information acquisition by the agent about the left side of the distribution is as follows. To incentivize the agent to acquire information about asset A and then choose the right project conditional on that information, the agent should be punished adequately if he chooses asset A and the return is very low. This punishment should be large enough relative to the expected payoff from choosing asset B. However, because of limited liability, the best the principal can do is give the agent zero if low outcomes occur, as opposed to giving him a negative wage. Given the flat nature of the contract at low returns from the asset A, the agent has no incentive to exert effort to acquire information to distinguish between these low outcomes. Hence the information intensity is zero in the region where the wage is zero. The consequence of this is that the posterior distributions after observing the signals are very close to the prior on the left side of the distribution, but much apart from the prior on the right side of the distribution. In other words, the agent learns very little about the left side of the distribution.

Efficiency requires that asset A be chosen with a higher likelihood if its cash flow is higher. Therefore, in the optimal contract, after a certain cutoff value of cash flow from asset A, the wage increases as cash flow increases. This increasing part of the wage incentivizes the agent to choose an information structure such that asset A is chosen with a higher probability if the true cash flow is higher. Thus, the increasing wage ensures efficiency in asset allocation.

The principal will also desire efficient allocation only if her profit also increases with the cash flow from asset A. We find that this is also true in equilibrium. In the optimal contract, the increasing part of wage from asset A is concave with a slope less than one which implies that the profit of principal increases with cash flow. This increasing profit of the principal together with a higher likelihood of asset A being chosen when cash flow is higher is desirable to the principal.

The key implication of our result that less information is acquired about the left tail of the distribution is that investors fail to understand the thickness of the left tail, hence they may invest in an asset even if it has a thick left tail. These investments have a high likelihood of failing ex post. Several recent papers have highlighted the role of neglecting the crash risk before a financial crisis. Baron and Xiong (2017) show that shareholders in banks do not price the crash risk during credit expansions. Jordà et al. (2021) document that higher bank capital does not decrease the likelihood of a banking crisis, and interpret this result as evidence of optimism and neglect of crash risk before a crisis. Fahlenbrach et al. (2018) show that stocks of banks with high loan growth underperform the stock of banks with low loan growth. Cheng et al. (2014) provide evidence that mid-level managers in the securitization business were unaware of problems in the housing market. All these results can be explained by economic agents making decisions after receiving high signals drawn from an information structure that does not focus on learning about the thickness of the left tail of the distribution.

While our model fits the setting of the principal-agent problem within an investment firm, it can equally well be applied to a setting where the CEO is considered the agent and the shareholders are the principal. Suppose the CEO has to choose between making a new investment such as setting up a new factory or acquiring a new firm and retaining cash. This investment can be thought of as an investment in asset A, while retaining cash can be thought of as an investment in asset B. A large literature has established that CEOs can be overconfident (Malmendier and Tate (2005)).[2] However, this appearance of overconfidence could merely be a consequence of the CEOs choosing not to learn about the left side of the distribution of their investment.

1.1 Literature Review

Our paper contributes to several strands of literature. First, we contribute to the large literature on financial crisis. Several papers have highlighted that if the economic environment changes because of the arrival of innovative technologies or financial products, then there may be a crisis because of hubris, behavioural biases and neglect of crash risk (Kindleberger

[2]For a survey of literature on CEO overconfidence, see Malmendier and Tate (2015).

(1978), Shiller (2000), Barberis (2013), Gennaioli et al. (2012)) or "this time is different" syndrome (Reinhart (2009)). Thakor (2012), which provides an alternative theory, builds a model where banks have the incentive to innovate to earn positive profits in a competitive market. However, innovation creates a risk that the investors may disagree with the banks at an interim stage and withdraw funding, thereby leading to a crisis.[3] The novelty of our paper is that our model generates endogenous beliefs of rational agents that ex post appear to be irrational as has been highlighted by several papers on financial crises.

We also add to the relatively new literature on contracting with flexible information acquisition (Yang and Zeng (2019), Yang (2020)).[4] In Yang (2020), a seller needs to carve out a security from an existing asset for liquidity purposes and the buyer acquires information flexibly. The optimal contract is a debt contract as it is the most information-insensitive contract and reduces the adverse selection problem.

Our paper is closest to Yang and Zeng (2019) which studies a security design problem in a production economy where an entrepreneur needs financing for a project from an investor who acquires information. If the ex ante prospects of the project is good, then information acquisition is not very useful and the optimal security is a debt contract, a security that disincentivizes information acquisition. However, if the ex ante project profitability is not high, then information acquisition is valuable and the optimal security needs to incentivize efficient financing decisions. The optimal contract in this case is a combination of debt and equity. The shape of the equity component in their paper is similar to that in ours and plays the role of incentivizing information acquisition to increase efficiency. The key difference between the investor's contract in their model and the agent's contract in our model is that when the investment results in a low cash flow, then the entire output goes to the investor because of the debt component of the contract; however, the exact opposite happens in the agent's contract who receives nothing if the cash flow from the investment in asset A is low. The reason for this difference is as follows. In our model, the agent is paid an endogenous positive wage if asset B is chosen. Hence he must be given very low wages at low values of cash flow from asset A to incentivize him to acquire information. Contrarily, the outside option of the investor is zero in Yang and Zeng (2019), hence there need not be any punishment to incentivize information acquisition.

Another class of papers that has studied contracting within financial institutions is also connected. Some of these papers have highlighted that competition to hire talented employ-

[3]For another model on innovation and crises, see Biais et al. (2009).

[4]Thus our paper also adds to the literature on contracting with an agent for information acquisition. See, Demski and Sappington (1987), Lambert (1986),Gromb and Martimort (2007), Häfner and Taylor (2022), for example. In these papers, the information structure is exogenously given, while in our paper it is endogenously derived.

ees can lead to excessive risk-taking (Bénabou and Tirole (2016), Acharya et al. (2016)). Heider and Inderst (2012) build a model in which effort is required to prospect for loans. If competition increases in the loan market, then soft information is ignored while sanctioning the loans.[5] In another interesting paper Bouvard and Lee (2020), risk management is modeled as information acquisition, which requires time, about a trading opportunity. However, since banks preemptively compete for the same trading opportunity, there is a race to the bottom and a socially inefficient amount of time is spent on information acquisition. They do not solve for optimal contracts though. In our model, competition for talent or investment opportunities has no role, and yet socially inefficient information acquisition and project allocation takes place because of moral hazard and limited liability.

The rest of the paper is organized as follows. In section 2, we build the principal-agent model. Section 3 solves for the optimal first best contract. In section 4, we solve the second-best contract and highlight the inefficiencies. Section 5 shows some comparative statics and section 6 concludes.

2 Model

In our model, there are two risk-neutral players, a principal ("she") and an agent ("he"), and two dates, 0 and 1. The principal has capital I that can be invested at $t = 0$ in one of two assets, asset A and asset B. If the investment is made in asset A, then at $t = 1$, a nonnegative verifiable random cashflow θ is produced with a distribution F that admits a probability distribution function f over its support $\Theta = [\underline{\theta}, \bar{\theta}]$. If the investment is made in the asset B, it produces a cash flow given by the number $\theta_B \in \mathbb{R}_+$. We make the following assumption to ensure that the asset allocation decision is non-trivial.

Assumption 1. $\underline{\theta} \geq 0, \theta_B \in (\underline{\theta}, \bar{\theta})$

At $t = 0$, the principal offers a contract to the agent that specifies the agent's payout based on the agent's asset choice and the cash flow realized. That is, the principal commits to a payment rule given by the pair $(w, \bar{w})$ where $w : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}_+$ is wage when the agent chooses asset A and $\bar{w}$ is the wage when he chooses asset B. We assume limited liability for both the principal and the agent, i.e., the wages satisfy the conditions $0 \leq \bar{w} \leq \theta_B$ and $0 \leq w(\theta) \leq \theta$.

If the agent accepts the contract, the agent first acquires information about asset A. We assume that the agent needs to exert costly effort to acquire information, which is measured by entropy reduction (Sims (2003)). The cost per unit of reduction in entropy is denoted

[5] Agarwal and Ben-David (2018) and Cole et al. (2015) provide evidence that if volume-based incentives are given to loans officers then there is less reliance on soft information and loan performance decreases.

by $\mu > 0$. The information structure is flexible (Yang (2020), Yang and Zeng (2019)) and is discussed in detail later. After acquiring information and updating his beliefs, the agent makes the binary decision ($a \in \{0, 1\}$) of investing in the asset A ($a = 1$) or asset B ($a = 0$). After the agent chooses the asset, the asset return is realized, and the contracted amount is paid to the agent. The underlying assumption here is that the principal does not have expertise in acquiring information about asset A, hence she hires the agent to do so.

Asset A can be interpreted as a new technology or innovative financial product about which very little is known and hence costly information needs to be acquired about it. On the other hand, asset B can be thought of as a conventional asset about which there is enough historical data and little to learn about, or it can be thought of as a safe asset such as government bonds. The key point about asset B is not that it has a degenerate distribution but that no information needs to be acquired about it to learn more about its distribution. We can assume both the cash flow from asset B and the wage paid when it is chosen by the agent to be random variables and our result will still hold.[6]

Our model fits several economic situations. For example:

(i.) A CEO (principal) is trying to incentivize a trading desk to choose to invest between innovative products such as CDOs (asset A) or conventional loans (asset B).

(ii.) The head of a hedge fund or a mutual fund asks an investment manager to choose to invest in new technology firms such as IT firms or AI firms versus old technology firms.

(iii.) A CEO asks a trader to choose between hedging an asset on its portfolio or not. The unhedged asset can be thought of as asset A while the hedged asset can be thought of as asset B.

(iv.) The owners of a firm (principal) ask the CEO (agent) to choose between investing in a new factory (or acquiring a new firm) versus retaining cash.

2.1 Flexible Information Structure

We model flexible information acquisition following the work of Woodford (2008), Yang (2020) and Yang and Zeng (2019). A flexible information structure allows the agent to choose any information structure and the associated cost is μ times the expected entropy reduction.

Since the information acquisition is flexible, the agent can acquire information about asset A by arranging to receive a signal, which could be correlated arbitrarily to the underlying return θ. Specifically, under the information structure π, which is chosen by the

[6]See section 5 for details.

agent, the signal x is drawn from some set X, according to a conditional probability $\pi(x|\theta)$ chosen by the agent. We need not impose any restrictions on X. Note that $\pi : \Theta \to \Delta X$. ΔX denotes the set of all probability distributions over X. However, since the action, a, of the agent is binary, where $a = 1$ implies investing in asset A and $a = 0$ implies investing in asset B, the agent will always choose a binary-signal information structure. We denote this signal by $s \in \{0, 1\}$. Any finer information structure will result in an unnecessary expense on information acquisition without adding any value to decision-making. This result is now standard in literature (Yang (2020), Yang and Zeng (2019)), hence we do not derive it. Therefore, the information structure can be characterized by a measurable function of θ, $m(\cdot) : \Theta \to [0, 1]$, which is the probability of seeing the signal 1 if the cashflow realized from the risky project is going to be θ. The agent takes the action $a = 1$ if the signal takes the value 1, else she takes the action $a = 0$. Therefore, we restrict m to belong to the set of all measurable functions from Θ to $[0, 1]$ given by $\mathcal{M}$. We refer to $m(\cdot)$ as the signal structure or the information structure.

The key question we are trying to answer in our paper is why economic agents are unaware of the left side of the distribution. To answer this question we need to model our information structure in such a way that the agent can choose how much it wants to learn about each possible state, i.e., the information structure needs to be completely flexible. The main advantage of modeling information structure the way we have done it is that it allows us to simultaneously solve for a flexible information structure and the corresponding contract in a principal-agent setting. In this setting, the absolute value of the first-order derivative of the function $m(\cdot)$, i.e., $|\frac{dm(\theta)}{d\theta}|$, measures the intensity of information acquisition around θ (Yang (2020)). Thus this term allows us to capture the intensity of endogenous information acquisition. A larger $|\frac{dm(\theta)}{d\theta}|$ implies that the agent acquires more information around θ, and hence, the more her decision changes with a change in θ. The observation of the signal reduces the uncertainty in the random variable θ, where the uncertainty of a random variable is measured by Shannon's entropy $H(\cdot)$. The cost associated with a signal structure, $m(\cdot)$, is proportional to the expected reduction in uncertainty after observing the signal realization. The difference between the expected posterior entropy and the prior entropy is given by

$$I(m(\cdot)) = \mathbb{E}[g(m(\theta))] - g(\mathbb{E}[m(\theta)]),$$

where $g(x) = x\ln(x) + (1 - x)\ln(1 - x)$, and the expectation is taken under the probability measure F. Given the per unit cost of information acquisition, μ, the total cost associated with a signal structure $m(\cdot)$ is

$$c(m(\cdot)) = \mu I(m(\cdot)).$$

3 First Best

In this section, we analyze the first-best signal structure in which the principal himself acquires the information and makes the asset allocation decision. We denote the utility of the principal if he takes action a and the asset A returns θ by $v(a, \theta)$. The ex post gain in payoff from choosing asset A over B is given by $\Delta v(\theta) = v(1, \theta) - v(0, \theta) = \theta - \theta_B$. In the first best, the principal's problem is to choose the optimal information structure $m \in \mathcal{M}$ that maximizes the principal's payoff. It is given as follows:

$$\max_{m \in \mathcal{M}} V(m) = \mathbb{E}[m(\theta)\Delta v(\theta)] + \theta_B - c(m)$$

The characterization of the optimal m is given by Proposition 1 in Yang and Zeng (2019), which we restate here.

Proposition 1. *Let $m \in \mathcal{M}$ be an optimal strategy and $p_1 = \mathbb{E}[m(\theta)]$ be the corresponding unconditional probability of taking action $a = 1$. Then,*

(i) the optimal strategy is unique;

(ii) there are three possibilities for the optimal information structure:

(a) $p_1 = 0$ (i.e., $m(\theta) = 0$ almost surely) if and only if

$$\mathbb{E}[e^{\mu^{-1}.\Delta v(\theta)}] \leq 1; \tag{1}$$

(b) $p_1 = 1$ (i.e., $m(\theta) = 1$ almost surely) if and only if

$$\mathbb{E}[e^{-\mu^{-1}.\Delta v(\theta)}] \leq 1; \tag{2}$$

(c) $p_1 \in (0, 1)$ if and only if $\mathbb{E}[e^{\mu^{-1}.\Delta v(\theta)}] > 1$ and $\mathbb{E}[\mathbb{E}[e^{-\mu^{-1}.\Delta v(\theta)}] \leq 1] > 1$, in which case the optimal information structure $m(\theta)$ is given by equation

$$\Delta v(\theta) = \mu \cdot [g'(m(\theta)) - g'(p_1)]. \tag{3}$$

The proposition first says that optimal information structure is unique. This is because the functional $V(m(\cdot))$ is concave in $m(\cdot)$. Second, the proposition gives three cases for information acquisition and investment choice. In case (a), the principal does not acquire information and chooses asset B. This will happen if θ_B is relatively large or if the cost of information acquisition, μ, is large. In case (b), the principal does not acquire information and chooses asset A. This will happen if θ_B is relatively small or if μ is large. The interesting case is (c), when the principal acquires information. This will happen if μ is low and the ex ante profitability of any asset is not extreme relative to the other. The optimal information structure $m(\theta)$ is given by equation 3, which is obtained by equating pointwise marginal benefit to pointwise marginal cost.

4 Second Best

We now discuss the principal-agent problem which is a sequential game. The principal first offers the contract $(w, \bar{w})$ and then the agent chooses the signal structure and makes the asset allocation decision that maximizes his utility. The agent's outside option is assumed to be 0. We first characterize the best response of the agent for a given contract $(w, \bar{w})$, which we denote by $m_{w,\bar{w}}(\cdot)$. Let $\Delta u(\theta)$ denote the ex post gain in payoff from choosing asset A over B for the agent, i.e.,

$$\Delta u(\theta) = u(1, \theta) - u(0, \theta) = w(\theta) - \bar{w}.$$

The problem of the agent can be formally stated as

$$\max_{m \in \mathcal{M}} U(m) = \mathbb{E}[m(\theta) \cdot \Delta u(\theta)] + \bar{w} - c(m).$$

There are two points to note here. First, the agent's problem is a direct adaptation of the first best problem of the principal. Therefore the best response of the agent can be characterized by using proposition 1. We obtain $m_{w,\bar{w}}(\cdot)$ by simply replacing $\Delta v(\theta)$ by $\Delta u(\theta)$ in proposition 1. The second point is that the principal's and the agent's incentives are perfectly aligned, i.e., they choose the same information structure if

$$\theta - \theta_B = w(\theta) - \bar{w} \quad \forall \theta \in \Theta.$$

Now the equilibrium is formally defined as follows:

Definition 1. *A sequential equilibrium in our model is a combination of the principal's optimal wage offer $(w^*, \bar{w}^*)$ and the agent's optimal information structure $m_{w,\bar{w}(.)}$ for any wage contract $(w, \bar{w})$, such that*

(i.) the agent's optimal information structure for a given wage contract $(w, \bar{w})$ is given by proposition 1, where $\Delta v(\theta)$ is replaced by $\Delta u(\theta)$, and

(ii.) the principal offers the optimal contract that satisfies

$$(w^*, \bar{w}^*) \in \arg\max_{(w,\bar{w})} V(w, \bar{w}) = \mathbb{E}\left[m_{w,\bar{w}}(\theta)\{(\theta - w(\theta)) - (\theta_B - \bar{w})\}\right] + (\theta_B - \bar{w})$$

subject to:

$$w(\theta) \in [0, \theta], \bar{w} \in [0, \theta_B], \tag{LL}$$

$$\mathbb{E}[m_{w,\bar{w}}(\theta)(w(\theta) - \bar{w})] + \bar{w} - \mu c(m_{w,\bar{w}}) \geq 0. \tag{AP}$$

The constraint (AP) gives the participation constraint of the agent and the constraint (LL) is the limited liability constraint of both the principal and the agent. The limited liability of the agent is an important constraint as without this constraint, much like in other principal-agent models, the principal can always offer a contract to the agent such that his optimal signal structure and asset allocation decision is the same as that in the first best (see below). Also, given that $\bar{w} \geq 0$ and $w(\theta) \geq 0$ because of the limited liability of the agent, and that the agent can always choose not to incur any cost on information acquisition, the minimum utility the agent can get when he is choosing his optimal information structure is 0. Therefore, for any given contract the participation constraint will always be satisfied, hence we can ignore this constraint.

4.1 Implementing first-best when the agent does not have limited liability

Like in other principal-agent models, the first-best can be implemented if the agent does not have limited liability.

Lemma 1. *If the agent does not have limited liability, then the principal can choose a contract that implements the first-best information structure.*

The principal can implement the first-best in the following way. First, recall from the discussion above that the principal's and the agent's incentives are perfectly aligned, i.e., they will choose the same information structure if $\theta - \theta_B = w(\theta) - \bar{w}$. Now let us denote the expected utility of the agent in the first best situation by V_f. The principal can offer a contract to the agent such that

$$w(\theta) = \theta - V_f,$$

and

$$\bar{w} = \theta_B - V_f.$$

In this wage contract, $w(\theta) - \bar{w} = \theta - \theta_B$; therefore the agent will choose the same information structure that the principal chooses in the first-best case. Hence this contract gives the agent an expected utility of zero and the principal an expected utility of V_f. Therefore this contract implements the first best.

4.2 Optimal contracts without information acquisition

If the principal does not want the agent to acquire information then, she will offer a wage $w(\theta) = 0$ and $\bar{w} = 0$ to the agent and ask her to choose the asset which has the unconditional higher expected value. So, if $E[\theta] > \theta_B$, the principal asks the agent to choose the asset A, else she asks the agent to choose asset B. Since the outside option of the agent is 0, the agent accepts the contract and invests in the project asked by the principal without information acquisition.

4.3 Optimal contract with information acquisition

In this section, we characterize the optimal contract when the principal wants the agent to acquire information. Since the contract induces information acquisition, Proposition 1 implies that the wage contract $(w(\theta), \bar{w})$ must satisfy

$$\mathbb{E}[e^{\mu^{-1}\Delta u(\theta)}] > 1,$$

and

$$\mathbb{E}[e^{-\mu^{-1}\Delta u(\theta)}] > 1.$$

Also, note that if $\bar{w}$ is equal to 0, then the agent has no incentive to choose the safe project and will choose the risky project without information acquisition. Similarly, if $w(\theta)$ equals 0 for all θ, then the agent again does not acquire any information and chooses the safe project. This intuition gives us the following lemma.

Lemma 2. *Suppose the contract $(w, \bar{w})$ induces information acquisition by the agent, then, $\bar{w} > 0$ and $w(\theta) > 0$ for all $\theta \in A$, for some $A \subset \Theta$ that has a strictly positive lebesgue measure.*

Proof: See Appendix.

We now discuss some properties of the optimal response of the agent $m_{w,\bar{w}(\cdot)}$ given the contract $(w, \bar{w})$. If the contract induces the agent to acquire information then by proposition 1, $m_{w,\bar{w}}$ is unique and satisfies a condition analogous to equation 3, i.e.,

$$w(\theta) - \bar{w} = \mu \cdot [g'(m(\theta)) - g'(\mathbb{E}[m(\theta)]] \tag{4}$$

where

$$g'(x) = \ln(\frac{x}{1-x}). \tag{5}$$

Note that $g'(\cdot)$ is an increasing function. Therefore, for a given contract $(w, \bar{w})$, if $w(\theta_1) > w(\theta_2)$, then $m(\theta_1) > m(\theta_2)$. The result is intuitive as the agent prefers to choose the risky asset with a higher likelihood for a certain value of realization θ if he expects a higher reward from the realization of that θ. This also highlights the key trade-off that the principal faces when designing the optimal contract. The utility of the principal is given by

$$V(w, \bar{w}) = \mathbb{E}[m_{w,\bar{w}}[(\theta - w(\theta)) - (\theta_B - \bar{w})] + (\theta_B - \bar{w}).$$

Therefore, for a given $\bar{w}$, the principal wants the agent to choose asset A with a higher probability if true θ is higher. However, in order to do so, it must offer the agent a higher wage $w(\theta)$, which in turn reduces the payoff of the principal at the state θ. The principal balances this trade-off to arrive at the optimal contract.

The agent's optimal response also satisfies the intuitive property that for a fixed $w(\theta)$, as the reward from choosing the safe project $\bar{w}$ increases, the likelihood of choosing the risky project decreases pointwise.

Lemma 3. *Suppose the contract $(w, \bar{w})$ induces information acquisition by the agent, then*

$$\frac{\partial \mathbb{E}[m_{w,\bar{w}}(\theta)]}{\partial \bar{w}} < 0, \ \textit{and} \ \frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}} < 0 \ \textit{a.e.} \ ,$$

where $\frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}}$ is the partial derivative of $m_{w,\bar{w}}(\theta)$ with respect to $\bar{w}$ for each $\theta \in \Theta$.

Proof: See Appendix.

Having discussed some of the properties of the agent's optimal response for a given contract, we now characterize the optimal contract that the principal will offer. First, we show that $\bar{w}^* < \theta_B$.

Lemma 4. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition then, $\bar{w}^* < \theta_B$.*

Proof: See Appendix.

The intuition for this is as follows. If $\bar{w}^* = \theta_B$, then the principal's payoff if the agent chooses asset B is zero. Therefore, in equilibrium, there is no benefit of inducing the agent to acquire information. Hence $\bar{w}^*$ must be less than θ_B in an equilibrium with information acquisition.

We discussed above that the key trade-off in designing the optimal $w(\theta)$ is that while the principal wants higher $m_{w,\bar{w}}(\theta)$ for higher values of θ, he needs to pay a higher $w(\theta)$ for this to happen which reduces his profit at that θ. Despite this trade-off, we show that in the optimal contract that induces information acquisition, $w^*(\theta)$ and hence $m^*_{w,\bar{w}}(\theta)$ is weakly increasing in θ.

Lemma 5. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition then, w^* is weakly increasing on Θ a.e..*

Proof: See appendix.

This result ensures the efficiency of asset allocation. The principal compensates the agent more for a higher value of the outcome θ to incentivize him to acquire an efficient information structure.

Next, we use the variational approach to further characterize the optimal contract. In this approach, we perturb the contract $w(\theta)$ and impose the condition that for the optimal contract $w^*(\theta)$ any perturbation should not give the principal any higher utility than $w^*(\theta)$. This gives us the following lemma.

Lemma 6. *Let $(w, \bar{w})$ be a contract for a given $\bar{w} \in \mathbb{R}_+$. Let $\hat{w}(\theta) = w(\theta) + \alpha\epsilon(\theta)$ be a perturbation of w where $\alpha \in \mathbb{R}$ and ϵ is any real valued measurable function defined on Θ. The principal's marginal expected payoff from the perturbation is given by*

$$\frac{dV(w+\alpha\epsilon, \bar{w})}{d\alpha}\bigg|_{\alpha=0} = \begin{cases} \mathbb{E}\left[\epsilon(\theta)\cdot r(\theta)\right] \;\; \textit{if}\; \mathbb{E}K^{\mu^{-1}(w(\theta)-\bar{w})} > 1 \;\textit{and}\; \mathbb{E}K^{-\mu^{-1}(w(\theta)-\bar{w})} > 1, \\ \mathbb{E}\left[-\epsilon(\theta)\right] \;\; \textit{if}\; \mathbb{E}K^{-\mu^{-1}(w(\theta)-\bar{w})} < 1, \\ 0 \;\textit{if}\; \mathbb{E}K^{\mu^{-1}(w(\theta)-\bar{w})} < 1, \end{cases}$$

where

$$r(\theta) = -m_{w,\bar{w}}(\theta) + \frac{1}{\mu g''(m_{w,\bar{w}}(\theta))}\left[(\theta - w(\theta) - \theta_B + \bar{w}) + \beta_{\bar{w}}\right]$$

and

$$\beta_{\bar{w}} = \mathbb{E}\left(\frac{\theta - w(\theta) - \theta_B + \bar{w}}{g''(m_{w,\bar{w}}(\theta)}\right)\frac{\mathbb{E}(g''(m_{w,\bar{w}}(\theta))}{\mathbb{E}(g''(m_{w,\bar{w}}(\theta))) - g''(p_1)}$$

is a constant that is determined in equilibrium.

The term $r(\theta)$ is the Frechet derivative which gives the marginal change in the expected utility of the principal from perturbing $w(\theta)$. When we perturb $w(\theta)$, there is a direct effect on the utility of the principal, which is the change in utility without taking into account the induced change in $m(\theta)$; and an indirect effect which is the change in expected utility through the induced change in $m(\theta)$. The first term in the expression of $r(\theta)$ is $-m_{w,\bar{w}}(\theta)$, which is the direct effect on the expected utility from choosing the risky project as a consequence of changing $w(\theta)$. This term is obviously negative as increasing $w(\theta)$ reduces the utility of the principal. The second term is the indirect effect on the utility of the principal through change in $m(\theta)$. The expected value of the second term multiplied by the perturbation $\epsilon(\theta)$ must be positive. This trade-off pins down the optimal contract $w(\theta)$.

We now define the following three sets based on where $w^*(\theta)$ lies:

$$A_0 = \{\theta \in \Theta : w^*(\theta) = 0\},$$
$$A_1 = \{\theta \in \Theta : w^*(\theta) \in (0, \theta)\},$$
$$A_2 = \{\theta \in \Theta : w^*(\theta) = \theta\}.$$

From the Frechet derivative derived above, the first-order conditions of the principal's problem can be written as:

$$r^*(\theta) \begin{cases} \leq 0 \text{ a.e.} & \text{if } w^*(\theta) = 0, \text{ i.e., } \theta \in A_0; \\ = 0 \text{ a.e.} & \text{if } w^*(\theta) \in (0, \theta), \text{ i.e., } \theta \in A_1; \\ \geq 0 \text{ a.e.} & \text{if } w^*(\theta) = \theta, \text{ i.e., } \theta \in A_2. \end{cases}$$

We use the expression for $r^*(\theta)$ from Lemma 6 and the fact that $g''(x) = x^{-1}(1-x)^{-1}$ to rewrite the first-order condition as

$$(1 - m^*(\theta))(\theta - w^*(\theta) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*}) \begin{cases} \leq \mu \text{ a.e.} & \text{if } w^*(\theta) = 0, \text{ i.e., } \theta \in A_0; \\ = \mu \text{ a.e.} & \text{if } w^*(\theta) \in (0, \theta), \text{ i.e., } \theta \in A_1; \\ \geq \mu \text{ a.e.} & \text{if } w^*(\theta) = \theta, \text{ i.e., } \theta \in A_2. \end{cases} \tag{6}$$

These conditions help us characterize the shape of the optimal contract. In the optimal contract, there will be a region where limited liability constraints $0 \leq w(\theta) \leq \theta$ do not bind, i.e., the set A_1 has a positive measure.[7] We call this region the "unconstrained" part of the optimal $w^*(\theta)$ and denote it by $\tilde{w}(\theta)$. This unconstrained wage and the corresponding information structure, which we denote by $\tilde{m}_{w,\bar{w}}(\cdot)$, satisfy the following equations.

Lemma 7. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition and $\lambda(A_1) > 0$. Let m^* denote the agent's best response to $(w^*, \bar{w}^*)$. Then,*

$$\tilde{w}(\theta) - \bar{w} = \mu(g'(\tilde{m}_{w,\bar{w}}(\theta)) - g'(\mathbb{E}[m^*(\theta)])) \textit{ for all } \theta \in A_1, \tag{7}$$

and,

$$(1 - \tilde{m}(\theta))(\theta - \tilde{w}(\theta) - \theta_B + \bar{w}^* + \beta_{\bar{w}}) = \mu, \tag{8}$$

[7] This is formally proved later.

where $E[m^*(\theta)]$ and $\beta_{\bar{w}^*}$ are constants determined in equilibrium.

These conditions jointly determine the unconstrained part of the optimal contract and the corresponding information structure. Condition 7 come directly from equation 4. The second condition 8 comes from the first-order condition of the principal in the region where limited liability conditions are not binding. The next lemma shows that both $\tilde{m}_{w,\bar{w}}(\cdot)$ and $\tilde{w}(\theta)$ are increasing. Additionally $\tilde{w}(\theta)$ is concave with slope less than 1.

Lemma 8. *In an equilibrium with information acquisition, the unconstrained part of optimal wage and the corresponding information structure are given by equations*

$$\frac{\partial \tilde{m}(\theta)}{\partial \theta} = \mu^{-1}\tilde{m}(\theta)(1-\tilde{m}(\theta))^2, \tag{9}$$

and

$$\frac{\partial \tilde{w}(\theta)}{\partial \theta} = 1 - \tilde{m}(\theta), \text{ for all } \theta \in A_1.$$

Furthermore, $w^*(\theta)$ is strictly concave when $\theta \in A_1$.

From these lemmas, we can fully characterize the shape of the optimal contract. The next proposition shows that the optimal contract when asset A is chosen has two regions. First, the wage is zero below a cutoff value of realization of θ. We denote this cutoff by $\hat{\theta}$. After values of θ greater than $\hat{\theta}$, the contract is unconstrained and is given by $\tilde{w}(\theta)$.

Proposition 2. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition then there exists a cutoff $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ such that*

$$w^*(\theta) \begin{cases} = 0 \text{ if } \theta \leq \hat{\theta}; \\ = \tilde{w}(\theta) \text{ if } \theta > \hat{\theta}. \end{cases}$$

The optimal information structure $m^*_{w,\bar{w}}$ satisfies

$$\frac{dm^*_{w,\bar{w}}(\theta)}{d\theta} \begin{cases} = 0 \text{ if } \theta < \hat{\theta}; \\ > 0 \text{ if } \theta > \hat{\theta}. \end{cases}$$

The optimal $\bar{w}^* \in (0, \theta_B - \underline{\theta})$.

There are several things to unpack here. Let us do them one by one. First, note that since $\tilde{w}(\theta)$ is concave with a slope less than 1, it implies that $\theta - w(\theta)$, i.e., the principal's payoff is strictly increasing in θ. This result is intuitive. The principal gives an increasing wage to the agent to incentivize efficient asset allocation. However, the principal also wants to benefit from this, hence her reward is also increasing in equilibrium. Thus the optimal contract is dual monotone.[8]

Second, note that the optimal $\bar{w}^* < \theta_B - \bar{\theta}$, which in turn implies that

$$\theta_B - \bar{w}^* > \underline{\theta}.$$

This result is also intuitive. The term $\theta_B - \bar{w}^*$ is the payoff of the principal from the safe project, while $\underline{\theta}$ is the lowest payoff of the principal from the risky project since $w(\underline{\theta}) = 0$. Therefore, the expression implies that the payoff of the principal from the risky project cannot always be greater than the payoff from the safe project. If this were not true, then there would be no benefit of information acquisition by agent to the principal in equilibrium. The principal would simply prefer that the agent always chooses asset A.

This result has another important implication. Because of this result, combined with the limited liability of the agent, the principal is unable to implement the first best information structure. The relative gain (or loss) for the principal from choosing asset A over asset B given the true state θ is θ is $\theta - \theta_B$, which determines the first best information structure (see equation 3). Thus the maximum relative loss to the principal is $\underline{\theta} - \theta_B$. Similarly, the relative gain (or loss) to the agent from choosing asset A over asset B is given by $w(\theta) - \bar{w}$. However, given that $\bar{w}^* < \theta_B - \bar{\theta}$, and $w(\theta) > 0$ by the limited liability of the agent, the maximum relative loss to the agent is less than that of the principal.

Basically, given that $w^* < \theta_B - \bar{\theta}$, and that $w(\theta) \geq 0$, the principal's loss will be greater than the loss to the agent at low values of θ. To align the incentive of the agent as much as possible with the first best, the principal offers as low a $w(\theta)$ as possible at low values of θ. Hence in equilibrium, there exists a cutoff $\hat{\theta}$ below which she chooses $w(\theta) = 0$ for every $\theta \leq \hat{\theta}$.

Finally, $w(\theta)$ is a constant for $\theta \leq \hat{\theta}$, hence $m_{w,\bar{w}}(\theta)$ is also a constant. Therefore the information intensity $dm_{w,\bar{w}}(\theta)/d\theta$ is 0 in this region. The agent does not exert any effort to distinguish between these low states. This is different from what was happening in the first best when the principal was acquiring information about all θ.

We show numerically that the ratio of information acquired to the left of $\hat{\theta}$ and to the right of $\hat{\theta}$ is lower in the case of the principal-agent problem than in the first best problem. The Shannon's entropy of the random variable θ is given by

$$H(\theta) = -\int_{\underline{\theta}}^{\bar{\theta}} p(\theta)\ln(p(\theta))d\theta.$$

[8]Dual monotonicity is also obtained in the security design problem solved by Yang and Zeng (2019).

The expected posterior entropy after observing the signal s given a information structure $m(\cdot)$ is

$$H(\theta|m) = -\int_{\underline{\theta}}^{\bar{\theta}} [q(1)p(\theta|1)\ln(p(\theta|1)) + q(0)p(\theta|0)\ln(p(\theta|0))]d\theta,$$

where $q(0)$ and $q(1)$ are the probabilities of observing signal 0 and 1 respectively. The information cost incurred in learning about a set $[\theta_1, \theta_2]$ is the entropy reduction in that set and we denote it as $\Delta H(\theta_1, \theta_2|m)$. It is given by

$$\Delta H(\theta_1, \theta_2|m) = -\int_{\theta_1}^{\theta_2} [q(1)p(\theta|1)\ln(p(\theta|1)) + q(0)p(\theta|0)\ln(p(\theta|0)) - p(\theta)\ln(p(\theta))]d\theta$$

One way to define the left tail and the right tail region is to partition the set Θ around a cutoff. A natural cutoff can be thought of as θ_B in which case the left tail region is the interval $[\underline{\theta}, \theta_B]$ and the right tail region is the interval $[\underline{\theta}, \theta_B]$. We show numerically that

$$r_{f.b.} = \frac{\Delta H(\underline{\theta}, \theta_b|m_{f.b.})}{\Delta H(\theta_b, \bar{\theta}|m_{f.b.})} > \frac{\Delta H(\underline{\theta}, \theta_b|m_{s.b.})}{\Delta H(\theta_b, \bar{\theta}|m_{s.b.})} = r_{s.b.}, \tag{10}$$

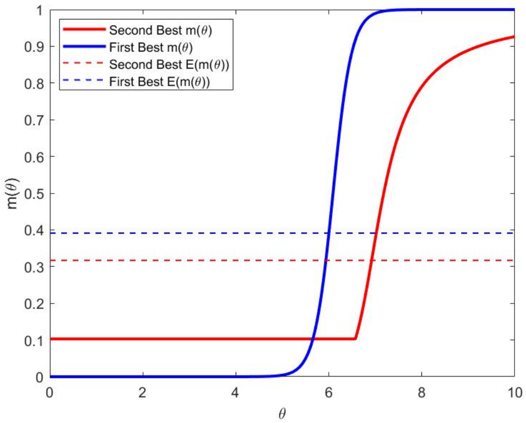
where $m_{f.b.}(\cdot)$ and $m_{s.b.}(\cdot)$ are the first-best and second-best information structures respectively. This implies that the ratio of entropy reduction in the left tail and right tail is lower in the second-best case than in the first-best case. Similar results hold if we choose the cutoff as $\hat{\theta}$.

4.4 Numerical example and comparative statics

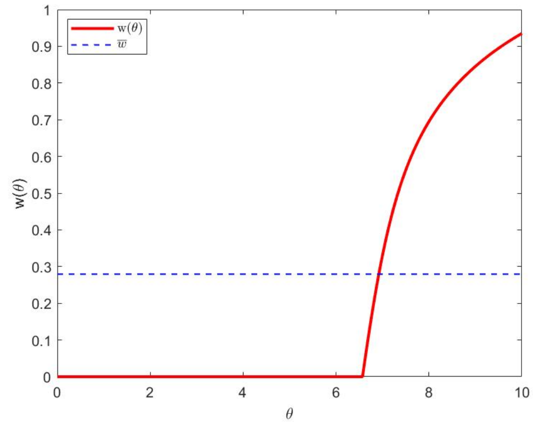
Here we first show a numerical example, and then discuss some comparative statics.

4.4.1 A numerical example

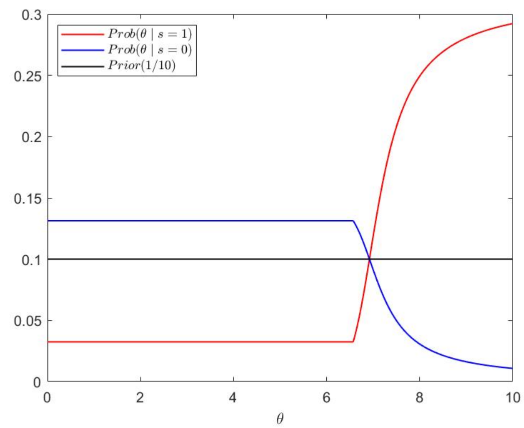
Figure 1 shows the optimal contract and the first-best and second-best information structures. We have chosen the following parameter values: $\theta \sim U[0, 10]$, $\theta_b = 6$ and $\mu = 0.2$. Notice that the asset choice is much more inefficient in the second best. Also, the information intensity is zero in the region below $\hat{\theta}$. This is because no wage is paid below $\hat{\theta}$, hence the agent does not exert effort to differentiate between different values in this region. Figure 1c shows the posterior probabilities after observing the signal realizations. The posterior probabilities in the left tail are very close to the prior probability while in the right tail, they are much apart from the prior. This shows that there is much more information acquisition on the right side of the distribution than on the left side.



(a) Optimal information structure



(b) Optimal wage



(c) Prior and posteriors

Figure 1: Optimal contract and information structure

4.4.2 Comparative statics

We numerically study the impact of change in information cost μ on the contract and the information structure. The parameter values chosen are $\theta \sim U[0, 10]$ and $\theta_b = 6$. We choose two different values of μ (see figure 2). First notice that as μ decreases from 0.3 to 0.2, the first-best information structure becomes more efficient. The likelihood of choosing asset A if $\theta > \theta_b = 6$ ($\theta < \theta_b$) increases (decreases). Similarly, the second-best information structure also becomes more efficient. Now let us see the change in the contracts. Since the expected payoff from asset B is higher, so as μ decreases, payment from asset B increases. Further, $\hat{\theta}$ also increases.

Figure 3 shows the impact of increasing the profitability of asset B compared to asset A. The parameter values are as follows: $\theta \sim U[0, 10]$, $\mu = 0.2$ and θ_B takes the values 5 and 6. As θ_B increases, the probability of choosing asset A decreases for all values of θ. To

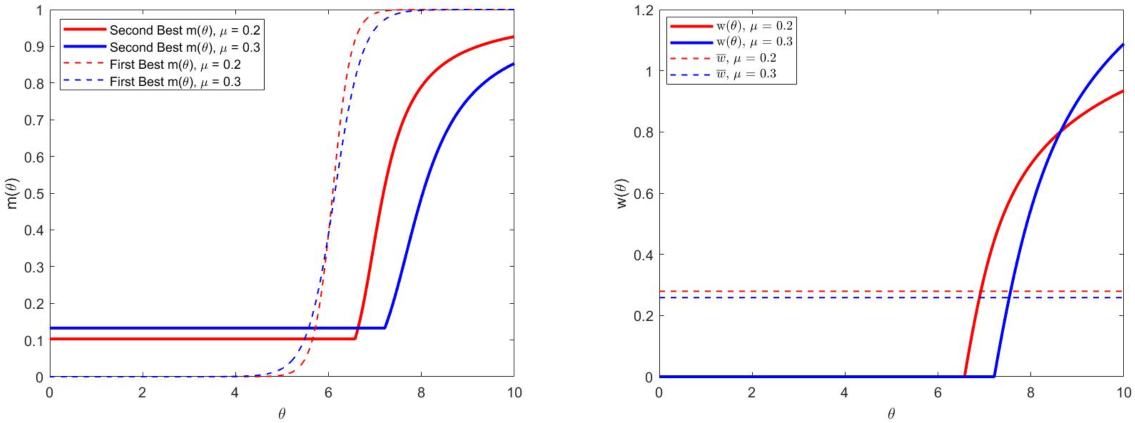


Figure 2: Impact of change in μ

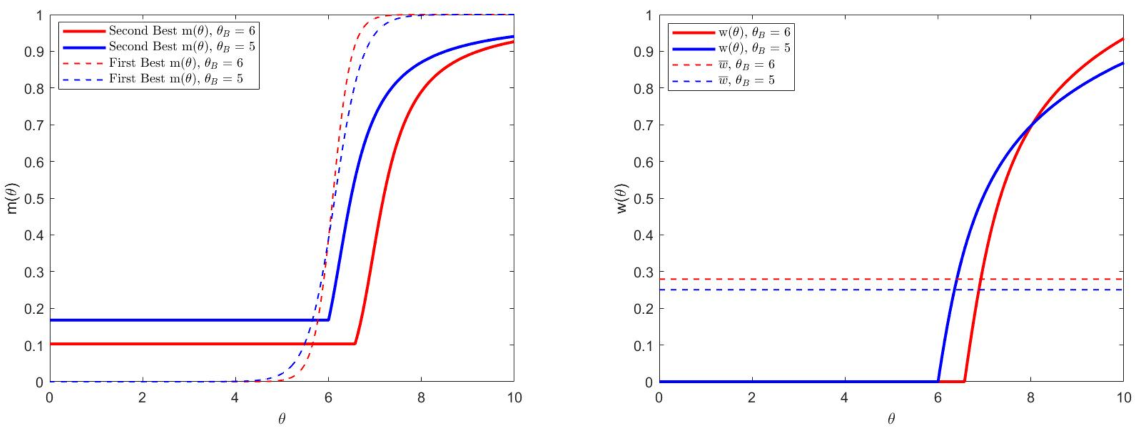


Figure 3: Impact of change in θ_B

incentivize the agent to choose asset A with lower probability, the principal offers a higher $\bar{w}$ when θ_B is higher.

Finally, Table I shows that the ratio of information acquired to the left of cutoff θ_B or $\hat{\theta}$ compared to the right of the cutoff is higher in the case of first-best than the second best. Thus in the second-best relatively less information is acquired about the left tail.

Welfare loss due to fat left tail: We have argued that as there is less information acquisition about the left tail, projects with thick left tails also get chosen. We show that as the thickness of the left tail increases, the welfare loss also increases (Figure 4). The figure is drawn with $\theta_B = 5$ and $\mu = 0.2$. θ is normally distributed with mean 5 and we vary the standard deviation and check the welfare loss for different values of standard deviation. We observe that as expected the welfare loss increases with standard deviation.

μ	θ_B	$\hat{\theta}$	$r_{f.b.}$	$r_{s.b.}$	$r_{f.b.}$	$r_{s.b.}$
			Cutoff $\hat{\theta}$		Cutoff θ_B	
0.2	5	6.00	1.27	0.55	1.00	0.42
0.2	6	6.57	0.93	0.52	0.84	0.45
0.3	5	5.75	1.10	0.55	1.00	0.45
0.3	6	7.21	1.19	0.38	0.84	0.30

Table I: Ratio of left tail vs. right tail information acquisition

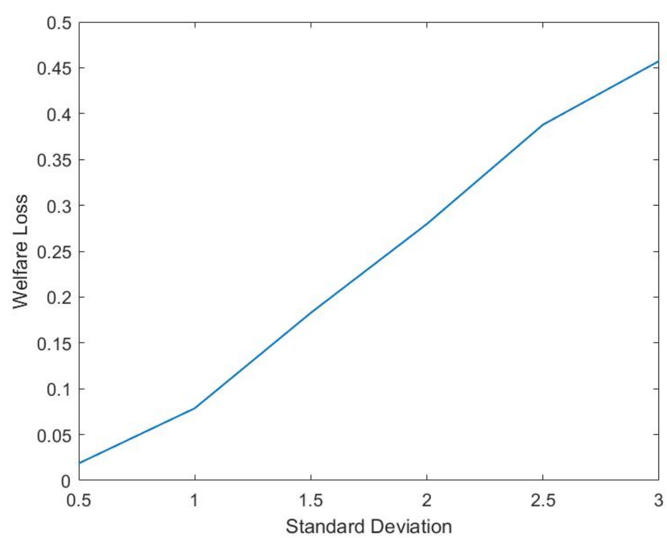


Figure 4: Impact of increasing standard deviation on welfare loss

5 Robustness: Asset B has non-degenerate distribution

So far we have assumed that asset B has a degenerate distribution. We now relax this assumption. Let the outcome of asset B be a random variable denoted by κ with probability distribution function $h(\kappa)$. Let us denote the expected value of asset B by $\bar{\kappa}$. If the agent chooses this asset and κ is realized, then the principal pays the agent a wage $w_B(\kappa)$.

First, let us see what happens to the principal problem. Her optimization problem can be written as

$$\max_{m} \mathbb{E}[m(\theta)(\theta - \bar{\kappa})] + \bar{\kappa} - \mu c(m). \tag{11}$$

This problem is exactly the same as the first best problem in section 3 except that θ_B has been replaced by $\bar{\kappa}$. Therefore, if the expected value of the cash flow from asset B is the same as before, i.e., $\bar{\kappa} = \theta_B$, then the solution to the first best problem remains the same.

Now let us analyze the second best. The agent's optimization problem for the wage (w, w_B) offered by the principal is given by

$$\max_{m} \mathbb{E}[m(\theta)w(\theta) + (1 - m(\theta))\mathbb{E}_h[w_B(\kappa)]] - \mu c(m) \tag{12}$$

Denote the solution to the agent's problem by $m_{w,w_B}(\cdot)$. The principal's problem can be written as

$$\max_{w(\theta), w_B(\kappa)} \mathbb{E}[m_{w,w_B}(\theta)(\theta - w(\theta)) + (1 - m_{w,w_B}(\theta))(\bar{\kappa} - \mathbb{E}_h[w_B(\kappa)])]$$

Therefore the principal's problem does not change except that the terms (θ_B) and $\bar{w}$ have been replaced by κ and $\mathbb{E}_h[w_B(\kappa)]$. Suppose the solution to the original problem is $(w*, \bar{w}^*)$. Therefore, the solution to this new problem will be that $w^*(\theta)$ remains the same and $\mathbb{E}_h[w^*(\kappa)] = \bar{w}^*$. The distribution of $w^*(\kappa)$ is irrelevant to the choice of optimal signal structure. Therefore the distribution of asset B is irrelevant.

6 Conclusion

The role of the financial intermediaries is to allocate capital to projects with the highest net present value. However, these intermediaries rely on managers to learn about cash flows of the new assets particularly those about which there is very little historical information. We show in a principal-agent setting that the optimal contract these managers are given incentivizes them not to learn about the tail risks of the asset they are investing in. This can result in even those projects being chosen which have a thick left tail and high likelihood of failure ex post. Thus, our model provides an explanation for why rational agents ignore the tail risk or crash risk in the investments they make.

Before the financial crisis of 2008, financial institutions held risky assets on their balance sheets. Broadly, there are two views on why banks did so. The first argument is that banks took this risk because they believed that they would be bailed out (Kelly et al. (2016), Bianchi (2016), Farhi and Tirole (2012), Acharya et al. (2010)). The other argument, as discussed above, is that managers at banks suffered from various behavioural biases which led to ignoring and underestimating the risks in their investments. Our paper provides an alternative explanation for why banks misunderstood the risks. The banks were investing in new kinds of innovative financial products such as ABSs and CDOs. The managers did not put effort into understanding the left tail of these investments because of agency problems related to exerting effort. Their contract simply did not incentivize them to do so. Thus we provide a novel explanation of the financial crisis.

References

Acharya, V., Pagano, M., Volpin, P., 2016. Seeking alpha: Excess risk taking and competition for managerial talent. The Review of Financial Studies 29, 2565–2599.

Acharya, V.V., Cooley, T., Richardson, M., Walter, I., et al., 2010. Manufacturing tail risk: A perspective on the financial crisis of 2007–2009. Foundations and Trends® in Finance 4, 247–325.

Agarwal, S., Ben-David, I., 2018. Loan prospecting and the loss of soft information. Journal of Financial Economics 129, 608–628.

Barberis, N., 2013. The psychology of tail events: progress and challenges. American Economic Review 103, 611–616.

Baron, M., Xiong, W., 2017. Credit expansion and neglected crash risk. The Quarterly Journal of Economics 132, 713–764.

Bénabou, R., Tirole, J., 2016. Bonus culture: Competitive pay, screening, and multitasking. Journal of Political Economy 124, 305–370.

Biais, B., Rochet, J.C., Woolley, P., 2009. The lifecycle of the financial sector and other speculative industries .

Bianchi, J., 2016. Efficient bailouts? American Economic Review 106, 3607–3659.

Bond, P., Glode, V., 2014. The labor market for bankers and regulators. The Review of Financial Studies 27, 2539–2579.

Bouvard, M., Lee, S., 2020. Risk management failures. The Review of Financial Studies 33, 2468–2505.

Cheng, I.H., Raina, S., Xiong, W., 2014. Wall street and the housing bubble. American Economic Review 104, 2797–2829.

Cole, S., Kanz, M., Klapper, L., 2015. Incentivizing calculated risk-taking: Evidence from an experiment with commercial bank loan officers. The Journal of Finance 70, 537–575.

Demski, J.S., Sappington, D.E., 1987. Delegated expertise. Journal of Accounting Research , 68–89.

Fahlenbrach, R., Prilmeier, R., Stulz, R.M., 2018. Why does fast loan growth predict poor performance for banks? The Review of Financial Studies 31, 1014–1063.

Farhi, E., Tirole, J., 2012. Collective moral hazard, maturity mismatch, and systemic bailouts. American Economic Review 102, 60–93.

Gennaioli, N., Shleifer, A., Vishny, R., 2012. Neglected risks, financial innovation, and financial fragility. Journal of financial economics 104, 452–468.

Glode, V., Lowery, R., 2016. Compensating financial experts. The Journal of Finance 71, 2781–2808.

Goldfarb, B., Kirsch, D.A., 2019. Bubbles and crashes: The boom and bust of technological innovation. Stanford University Press.

Gromb, D., Martimort, D., 2007. Collusion and the organization of delegated expertise. Journal of economic theory 137, 271–299.

Häfner, S., Taylor, C.R., 2022. On young turks and yes men: optimal contracting for advice. The RAND Journal of Economics 53, 63–94.

Heider, F., Inderst, R., 2012. Loan prospecting. The Review of Financial Studies 25, 2381–2415.

Jordà, Ò., Richter, B., Schularick, M., Taylor, A.M., 2021. Bank capital redux: solvency, liquidity, and crisis. The Review of economic studies 88, 260–286.

Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2016. Too-systemic-to-fail: What option markets imply about sector-wide government guarantees. American Economic Review 106, 1278–1319.

Kindleberger, C.P., 1978. Manias, Panics and Crashes: A History of Financial Crises. Palgrave Macmillan.

Lambert, R.A., 1986. Executive effort and selection of risky projects. The RAND Journal of Economics , 77–88.

Malmendier, U., Tate, G., 2005. Ceo overconfidence and corporate investment. The journal of finance 60, 2661–2700.

Malmendier, U., Tate, G., 2015. Behavioral ceos: The role of managerial overconfidence. Journal of Economic Perspectives 29, 37–60.

Payzan-LeNestour, E., Woodford, M., 2022. Outlier blindness: A neurobiological foundation for neglect of financial risk. Journal of Financial Economics 143, 1316–1343.

Reinhart, C.M., 2009. This time is different: Eight centuries of financial folly. Princeton University Press.

Shiller, R.J., 2000. Irrational Exuberance. Princeton University Press.

Sims, C.A., 2003. Implications of rational inattention. Journal of monetary Economics .

Thakor, A.V., 2012. Incentives to innovate and financial crises. Journal of Financial Economics 103, 130–148.

Veldkamp, L., 2023. Information choice in macroeconomics and finance. Princeton University Press.

Woodford, M., 2008. Inattention as a source of randomized discrete adjustment. unpublished, Columbia University .

Yang, M., 2020. Optimality of debt under flexible information acquisition. The Review of Economic Studies 87, 487–536.

Yang, M., Zeng, Y., 2019. Financing entrepreneurial production: security design with flexible information acquisition. The Review of Financial Studies 32, 819–863.

Appendix

A Proofs

A.1 Proof of Lemma 2

Suppose $\bar{w} = 0$ and $w(\theta) > 0$ for all $\theta \in A$, then the agent's best response is to acquire no information and choose $a = 1$. Next, suppose $\bar{w} > 0$ and $w(\theta) = 0$ for all $\theta \in \Theta$, except possibly for a subset of measure 0, then the agent's best response is to acquire no information and choose $a = 0$. Lastly, suppose that $\bar{w} = 0$ and $w(\theta) = 0$ for all $\theta \in \Theta$, except possibly for a subset of measure 0, then the agent's best response is to acquire no information and choose either $a = 0$ or $a = 1$.

From the agent's best response given by Equation 4 we know that,

$$m_{w,\bar{w}}(\theta) = f\left(\frac{w(\theta) + \bar{w}}{\mu} - g'(\mathbb{E}[m_{w,\bar{w}}])\right) \text{ where } f = g'^{-1}.$$

Taking the partial derivative with respect to $\bar{w}$, we get,

$$\frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}} = f'\left(\frac{w(\theta) - \bar{w}}{\mu} + g'(\mathbb{E}[m_{w,\bar{w}}])\right)\left(-\frac{1}{\mu} + g''(\mathbb{E}[m_{w,\bar{w}}])\frac{\partial \mathbb{E}[m_{w,\bar{w}}(\theta)]}{\partial \bar{w}}\right).$$

Taking expectations on both sides, noting that $f'(x) = \frac{1}{g''(f(x))}$, and then rearranging we get,

$$\frac{\partial \mathbb{E}[m_{w,\bar{w}}(\theta)]}{\partial \bar{w}} = -\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}{\mu(1 - g''(\mathbb{E}[m_{w,\bar{w}}(\theta)]))} = -\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - \mathbb{E}[(m_{w,\bar{w}}(\theta))^2]}{\mu}\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}{\mathbb{E}[(m_{w,\bar{w}}(\theta))^2] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}$$

and,

$$\frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}} = -\frac{m_{w,\bar{w}}(\theta) - (m_{w,\bar{w}}(\theta))^2}{\mu(1 - g''(\mathbb{E}[m_{w,\bar{w}}(\theta)]))} = -\frac{m_{w,\bar{w}}(\theta) - (m_{w,\bar{w}}(\theta))^2}{\mu}\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}{\mathbb{E}[(m_{w,\bar{w}}(\theta))^2] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}.$$

Further note that since the agent acquires information, $\mathbb{E}[m_{w,\bar{w}}(\theta)] \in (0,1)$ and $m_{w,\bar{w}}(\theta) < (m_{w,\bar{w}}(\theta))^2$ a.e., which implies that,

$$\frac{\partial \mathbb{E}[m_{w,\bar{w}}(\theta)]}{\partial \bar{w}} < 0, \text{ and } \frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}} < 0 \text{ a.e. }.$$

A.2 Proof of Lemma 3

We start by computing $\frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}}$.

From the agent's best response given by Equation 4 we know that,

$$m_{w,\bar{w}}(\theta) = f\left(\frac{w(\theta)+\bar{w}}{\mu} - g'(\mathbb{E}[m_{w,\bar{w}}])\right) \text{ where } f = g'^{-1}.$$

Taking the partial derivative with respect to $\bar{w}$, we get,

$$\frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}} = f'\left(\frac{w(\theta)-\bar{w}}{\mu} + g'(\mathbb{E}[m_{w,\bar{w}}])\right)\left(-\frac{1}{\mu} + g''(\mathbb{E}[m_{w,\bar{w}}])\frac{\partial \mathbb{E}[m_{w,\bar{w}}(\theta)]}{\partial \bar{w}}\right).$$

Taking expectations on both sides, noting that $f'(x) = \frac{1}{g''(f(x))}$, and then rearranging we get,

$$\frac{\partial \mathbb{E}[m_{w,\bar{w}}(\theta)]}{\partial \bar{w}} = -\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}{\mu(1-g''(\mathbb{E}[m_{w,\bar{w}}(\theta)]))} = -\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - \mathbb{E}[(m_{w,\bar{w}}(\theta))^2]}{\mu}\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}{\mathbb{E}[(m_{w,\bar{w}}(\theta))^2] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]}$$

and,

$$\frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}} = -\frac{m_{w,\bar{w}}(\theta) - (m_{w,\bar{w}}(\theta))^2}{\mu(1-g''(\mathbb{E}[m_{w,\bar{w}}(\theta)]))} = -\frac{m_{w,\bar{w}}(\theta) - (m_{w,\bar{w}}(\theta))^2}{\mu}\frac{\mathbb{E}[m_{w,\bar{w}}(\theta)] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}{\mathbb{E}[(m_{w,\bar{w}}(\theta))^2] - [\mathbb{E}[m_{w,\bar{w}}(\theta)]]^2}.$$

Further note that since the agent acuires information, $\mathbb{E}[m_{w,\bar{w}}(\theta)] \in (0,1)$ and $m_{w,\bar{w}}(\theta) < (m_{w,\bar{w}}(\theta))^2$ a.e., which implies that,

$$\frac{\partial \mathbb{E}[m_{w,\bar{w}}(\theta)]}{\partial \bar{w}} < 0, \text{ and } \frac{\partial m_{w,\bar{w}}(\theta)}{\partial \bar{w}} < 0 \text{ a.e. }.$$

A.3 Proof of Lemma 4

Let m^* denote the best response of the agent to $(w^*, \bar{w}^*)$. From Lemma **??** we know that

$$\left.\frac{\partial V(w^*,\bar{w}^*)}{\partial \bar{w}^*}\right|_{\bar{w}^*=\theta_B} = \mathbb{E}\left[\underbrace{\left.\frac{\partial m^*(\theta)}{\partial \bar{w}}\right|_{\bar{w}^*=\theta_B}}_{<0 \text{ a.e.}}\{\underbrace{(\theta - w^*(\theta))}_{\geq 0} - \underbrace{(\theta_B - \bar{w}^*)}_{=0}\} - \underbrace{(1-m^*(\theta))}_{>0 \text{ a.e.}}\right] < 0.$$

A.4 Proof of Lemma 5

For the sake of contradiction, suppose that w^* is decreasing on $E \subseteq \Theta$ and there exist $E_1, E_2 \subset E$ such that E_1, E_2 have positive lebesgue measure and $w^*(\theta') > w^*(\theta'')$ whenever $\theta' \in E_1$ and $\theta'' \in E_2$. Let $\underline{e} = \inf E$ and $\bar{e} = \sup E$. We construct a new contract $(\hat{w}, \hat{\bar{w}})$ by rearranging the values of w^* on E such that $\hat{w}$ is is increasing on E. Towards that we define a CDF on E given by $G(\theta) := \int_{\underline{\theta}}^{\theta} \mathbb{1}_E dF(\theta)$ and a rearrangement function given by $R(\theta) : E \to E$, defined implicitly by the equation $G(\theta) = G(\bar{e}) - G(R(\theta))\ \forall\ \theta \in E$. Note that $G(\underline{e}) = 0$ and $G(\bar{e})$ is equal to the probability that $\theta \in E$. Furthermore, R maps E onto E and since F is absolutely continuous with respect to the Lebesgue measure and has full support, R is strictly decreasing on E, and therefore R is a bijection, and therefore a rearrangement. Additionally, note that $d(G(\theta)) = d(G(R(\theta)))$, and therefore R is a measure preserving rearrangement.[9,10]

We define $(\hat{w}, \hat{\bar{w}})$ as follows

$$\hat{w}(\theta) = \begin{cases} w^*(\theta) \text{ if } \theta \notin E, \\ w^*(R(\theta)) \text{ if } \theta \in E. \end{cases}$$
$$\hat{\bar{w}} = \bar{w}^*.$$

Since $R(\theta)$ is a measure preserving rearrangement, by the definition of $\hat{w}$ we have

$$\int_E \hat{w}(\theta) dG(\theta) = \int_E w^*(R(\theta)) dG(R(\theta)) = \int_E w^*(\theta) dG(\theta).$$

Recall that the agent's best response m^* to the contract $(w^*, \bar{w}^*)$, is given by

$$\frac{w^*(\theta) - \bar{w}^*}{\mu} = g'(m^*(\theta)) - \mathbb{E}[g'(m^*(\theta))].$$

Since $\hat{w}$ is a measure preserving rearrangement of w^* and $\hat{\bar{w}} = \bar{w}^*$, the agent's best response $\hat{m}$ to the contract $(\hat{w}, \hat{\bar{w}})$ satisfies $\mathbb{E}[\hat{m}(\theta)] = \mathbb{E}[m^*(\theta)]$, which implies that $(\hat{w}, \hat{\bar{w}})$ induces information acquisition from the agent. In light of this note that $\hat{m}$ is a measure preserving rearrangement of m^*, i.e.,

$$\hat{m}(\theta) = \begin{cases} m^*(\theta) \text{ if } \theta \notin E, \\ m^*(R(\theta)) \text{ if } \theta \in E. \end{cases}$$

[9]To be precise $d(G(\theta)) = -d(G(R(\theta)))$ due to the reverse orientation of $R(\theta)$ from the definition. We ignore the negative sign and reverse the integration limits accordingly.

[10]By "measure preserving rearrangement" we mean that θ and $R(\theta)$ are assigned the same "mass" since $d(G(\theta)) = d(G(R(\theta)))$.

Let V^* and $\hat{V}$ be the payoff of the principal under contracts $(w^*, \bar{w}^*)$ and $(\hat{w}, \hat{\bar{w}})$ respectively.

$$\begin{aligned}
\hat{V} - V^* = &\int_E [\hat{m}(\theta)(\theta - \hat{w}(\theta) - \theta_B + \hat{\bar{w}})]dG(\theta) + (\theta_B - \hat{\bar{w}}) \\
&- \int_E [m^*(\theta)(\theta - w^*(\theta) - \theta_B + \bar{w}^*)]dG(\theta) + (\theta_B - \bar{w}^*) \\
= &\int_E [\hat{m}(\theta)(\theta - \hat{w}(\theta)) - m^*(\theta)(\theta - w^*(\theta))]dG(\theta) \\
= &\int_E [m^*(R(\theta))(\theta - w^*(R(\theta))) - m^*(\theta)(\theta - w^*(\theta))]dG(\theta) \text{ (using the definition of } \hat{m}) \\
= &\int_E [m^*(\theta)(R(\theta) - w^*(\theta)) - m^*(R(\theta))(R(\theta) - w^*(R(\theta)))]dG(R(\theta)) \text{ (switching from } \theta \text{ to } R(\theta)) \\
= &\frac{1}{2}\int_E [m^*(R(\theta))(\theta - w^*(R(\theta))) - m^*(\theta)(\theta - w^*(\theta)) \\
&+ m^*(\theta)(R(\theta) - w^*(\theta)) - m^*(R(\theta))(R(\theta) - w^*(R(\theta)))]dG(\theta) \text{ (since } dG(\theta) = dG(R(\theta))) \\
= &\frac{1}{2}\int_E [m^*(R(\theta))(\theta - R(\theta)) - m^*(\theta)(\theta - R(\theta))]dG(\theta) \\
= &\frac{1}{2}\int_E [(m^*(R(\theta)) - m^*(\theta))(\theta - R(\theta))]dG(\theta)
\end{aligned}$$

Finally note that by since by assumption w^* is decreasing on E, which implies that m^* is decreasing on E (using the agent's best response) which implies that if $\theta - R(\theta) > 0$ then $m^*(R(\theta)) - m^*(\theta) \geq 0$ and if $\theta - R(\theta) < 0$ then $m^*(R(\theta)) - m^*(\theta) \leq 0$, which implies that $(m^*(R(\theta)) - m^*(\theta))(\theta - R(\theta)) \geq 0$ for all $\theta \in E$. Further notice that since $w^*(\theta)$ is strictly greater on E_1 compared to E_2, $m^*(R(\theta)) - m^*(\theta) \neq 0$ on a strictly positive measure subset of E. We also know that $\{\theta : \theta \in E, \theta = R(\theta)\}$ is either empty or a singleton set because $R(\theta)$ is strictly decreasing on E and therefore $\{\theta : \theta \in E, \theta \neq R(\theta)\}$ has the same measure as that of E and therefore $\hat{V} - V^* > 0$, a contradiction to the optimality of $(w^*, \bar{w}^*)$.

A.5 Proof of Lemma 6

We first look at the change in the best response of the agent when the principal's strategy is perturbed. Recall from Proposition 1 that the best response of the agent can take three forms $(i) p_1 = 0$ $a.s.$, $(ii) p_1 = 1$ $a.s.$, and $(iii) p_1 \in (0, 1)$. We consider the case (iii) when $p_1 \in (0, 1)$, i.e., the agent's best response is an interior point of the set $\mathcal{M}$ and we know that $\mathbb{E}K^{\mu^{-1}(w(\theta)-\bar{w})} > 1$ and $\mathbb{E}K^{-\mu^{-1}(w(\theta)-\bar{w})} > 1$. In this case the best response of the agent to a contract $(w, \bar{w})$ offered by the principal is given by

$$w(\theta) - \bar{w} = \mu \cdot [g'(m_{w,\bar{w}}(\theta)) - g'(p_1)] \tag{A.1}$$

In particular for the perturbation $\epsilon(\theta)$ scaled by α,

$$\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0} = [g''(m_{w,\bar{w}}(\theta))]^{-1}\left[\frac{\epsilon(\theta)}{\mu} + g''(p_1)\left.\frac{dp_1}{d\alpha}\right|_{\alpha=0}\right].$$

Since $p_1 = \mathbb{E}m_{w,\bar{w}}(\theta)$, we have $\frac{dp_1}{d\alpha} = \mathbb{E}\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}$. Taking expectations on both sides of the above equation we get

$$\left.\frac{dp_1}{d\alpha}\right|_{\alpha=0} = \frac{1}{\mu}\mathbb{E}\left[\frac{\epsilon(\theta)}{g''(m_{w,\bar{w}}(\theta))}\right]\frac{\mathbb{E}[g''(m_{w,\bar{w}}(\theta))]}{\mathbb{E}[g''(m_{w,\bar{w}}(\theta))] - g''(p_1)}.$$

Plugging back $\left.\frac{dp_1}{d\alpha}\right|_{\alpha=0}$ in the expression of $\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0}$ we get

$$\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0} = [g''(m_{w,\bar{w}}(\theta))]^{-1}\left[\frac{\epsilon(\theta)}{\mu} + \frac{g''(p_1)}{\mu}\mathbb{E}\left[\frac{\epsilon(\theta)}{g''(m_{w,\bar{w}}(\theta))}\right]\frac{\mathbb{E}[g''(m_{w,\bar{w}}(\theta))]}{\mathbb{E}[g''(m_{w,\bar{w}}(\theta))] - g''(p_1)}\right] \tag{A.2}$$

We now turn to calculate the marginal change in payoff for the principal from applying the perturbation $\epsilon(\theta)$.

$$\begin{aligned}\left.\frac{dV(w+\alpha\epsilon,\bar{w})}{d\alpha}\right|_{\alpha=0} &= \int_\Theta \left[\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0}(\theta - w(\theta) - \theta_B + \bar{w}) - m_{w,\bar{w}}(\theta)\epsilon(\theta)\right] dP \\ &= \mathbb{E}\left[\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0}(\theta - w(\theta) - \theta_B + \bar{w}) - m_{w,\bar{w}}(\theta)\epsilon(\theta)\right]\end{aligned}$$

Plugging $\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0}$ derived above and after simplifying the expression we get,

$$\left.\frac{dV(w+\alpha\epsilon,\bar{w})}{d\alpha}\right|_{\alpha=0} = \mathbb{E}\left[\epsilon(\theta)\cdot r(\theta)\right],$$

where

$$r(\theta) = -m_{w,\bar{w}}(\theta) + \frac{1}{\mu g''(m_{w,\bar{w}}(\theta))}\left[(\theta - w(\theta) - \theta_B + \bar{w}) + \beta_{\bar{w}}\right]$$

and

$$\beta_{\bar{w}} = \mathbb{E}\left(\frac{\theta - w(\theta) - \theta_B + \bar{w}}{g''(m_{w,\bar{w}}(\theta)}\right)\frac{\mathbb{E}(g''(m_{w,\bar{w}}(\theta))}{\mathbb{E}(g''(m_{w,\bar{w}}(\theta))) - g''(p_1)}$$

is a constant that is determined in equilibrium.

Finally, consider the case when $p_1 = 1$ *a.s.*, i.e., $m_{w,\bar{w}}(\theta) = 1$ *a.s.*, which, from Proposition **??** implies that $\mathbb{E}K^{\mu^{-1}(w(\theta)-\bar{w})} \leq 1$. Consider the case when the inequality is strict. Since $\lim_{\alpha\to 0}\mu^{-1}\alpha\epsilon(\theta) = 0\ \forall\theta\in\Theta$, $\exists\delta > 0$, such that $\forall\alpha\in(-\delta,\delta)$ we have $\mathbb{E}K^{\mu^{-1}(w(\theta)+\alpha\epsilon(\theta)-\bar{w})} < 1$

and $m_{w+\alpha\epsilon,\bar{w}} = 1$ and therefore $\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0} = 0$. This further implies that

$$\begin{aligned}\left.\frac{dV(w+\alpha\epsilon,\bar{w})}{d\alpha}\right|_{\alpha=0} &= \int_{\Theta}\left[\left.\frac{dm_{w,\bar{w}}(\theta)}{d\alpha}\right|_{\alpha=0}(\theta - w(\theta) - \theta_B + \bar{w}) - m_{w,\bar{w}}(\theta)\epsilon(\theta)\right]dP \\ &= \mathbb{E}\left[-m_{w,\bar{w}}(\theta)\epsilon(\theta)\right] \\ &= \mathbb{E}\left[-\epsilon(\theta)\right], \text{ since } m_{w,\bar{w}}(\theta) = 1, a.s..\end{aligned}$$

Similarly when $p_1 = 0$ $a.s.$, i.e., $m_{w,\bar{w}}(\theta) = 0$ $a.s.$, and $\mathbb{E}K^{-\mu^{-1}(w(\theta)-\bar{w})} < 1$

$$\begin{aligned}\left.\frac{dV(w+\alpha\epsilon,\bar{w})}{d\alpha}\right|_{\alpha=0} &= \mathbb{E}\left[-m_{w,\bar{w}}(\theta)\epsilon(\theta)\right] \\ &= 0, \text{ since } m_{w,\bar{w}}(\theta) = 0, a.s..\end{aligned}$$

A.6 Proof of Proposition 2

Our proof builds on a series of lemmas given below.

Lemma 9. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition, then, at least one of $\lambda(A_0)$ or $\lambda(A_2)$ is equal to 0.*

Proof. For the sake of contradiction suppose that $\lambda(A_0) > 0$ and $\lambda(A_2) > 0$. Let $\theta_\alpha \in A_0$ and $\theta_\beta \in A_2$. Note that $w^*(\theta_\beta) > w^*(\theta_\alpha) = 0$ and from the agent's best response given by equation A.1, we know that $m^*(\theta_\beta) > m^*(\theta_\alpha)$. From the principal's first order necessary condition for optimality given by equation 6, we get

$$\begin{aligned}(1 - m^*(\theta_\beta))(\theta_\beta - w^*(\theta_\beta) - \theta_B + \bar{w} + \beta_{\bar{w}^*}) \geq \mu, \\ (1 - m^*(\theta_\alpha))(\theta_\alpha - w^*(\theta_\alpha) - \theta_B + \bar{w} + \beta_{\bar{w}^*}) \leq \mu.\end{aligned}$$

Since $\mu > 0$, we must have $1 - m^*(\theta_\beta) > 0$ and $\theta_\beta - w^*(\theta_\beta) - \theta_B + \bar{w} + \beta_{\bar{w}^*} > 0$. But that means that

$$1 - m^*(\theta_\alpha) > 1 - m^*(\theta_\beta) > 0,$$

and,

$$\theta_\alpha - w^*(\theta_\alpha) - \theta_B + \bar{w} + \beta_{\bar{w}^*} > \theta_\beta - w^*(\theta_\beta) - \theta_B + \bar{w} + \beta_{\bar{w}^*} > 0,$$

which implies that

$$(1 - m^*(\theta_\alpha))(\theta_\alpha - w^*(\theta_\alpha) - \theta_B + \bar{w} + \beta_{\bar{w}^*}) > \mu,$$

which is a contradiction to the optimality of $(w^*, \bar{w}^*)$ since $\theta_\alpha \in A_0$ and this inconsistency holds for a positive measure of points in A_0 and A_2. $\square$

Lemma 10. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition, then there exist cutoffs $\theta_0, \theta_1 \in \Theta$ such that*

$$\lambda([\underline{\theta}, \theta_0] \triangle A_0) = 0$$
$$\lambda([\theta_0, \theta_1] \triangle A_2) = 0$$
$$\lambda([\theta_1, \bar{\theta}] \triangle A_1) = 0$$

where λ represents the lebesgue measure set function and $\triangle$ represents the symmetric difference operation on sets.

Proof. Let us define $\bar{a}_i = \inf\{x : x \in \Theta, \lambda(A_i \cap [x, \bar{\theta}]) = 0\}$ and $\underline{a}_i = \sup\{x : x \in \Theta, \lambda(A_i \cap [\underline{\theta}, x]) = 0\}$ for $i \in \{0, 1, 2\}$. Note that the interval $[\underline{a}_i, \bar{a}_i]$ contains A_i except for zero measure subsets of A_i. If $\underline{a}_i \geq \bar{a}_i$, then it means that A_i has zero measure.

From Lemma **??** we know that w^* is weakly increasing on Θ a.e., and therefore, $\bar{a}_0 \leq \underline{a}_1$ and $\bar{a}_0 \leq \underline{a}_2$. We further claim that $\bar{a}_2 \leq \underline{a}_1$. To see this suppose that $\bar{a}_2 - \underline{a}_1 = \delta > 0$. Define $B_1 = [\underline{a}_1, \underline{a}_1 + \frac{\delta}{3}] \cap A_1$ and $B_2 = [\bar{a}_2 - \frac{\delta}{3}, \bar{a}_2] \cap A_2$. Note that B_1 and B_2 both have positive measure and $\sup B_1 < \inf B_2$. Let m^* denote the best response of the agent when offered the contract $(w^*, \bar{w}^*)$. From equations A.1 and 6 we know that for any $\theta_\alpha \in B_2$ we must have

$$w^*(\theta_\alpha) - \bar{w}^* = \mu(g'(m^*(\theta_\alpha)) - g'(\mathbb{E}(m^*(\theta))), \text{ and,}$$

$$(1 - m^*(\theta_\alpha))(\theta_\alpha - w^*(\theta_\alpha) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*}) \geq \mu$$

Similarly we know that for any $\theta_\beta \in B_1$ we must have

$$w^*(\theta_\beta) - \bar{w}^* = \mu(g'(m^*(\theta_\beta)) - g'(\mathbb{E}(m^*(\theta))), \text{ and,}$$

$$(1 - m^*(\theta_\beta))(\theta_\beta - w^*(\theta_\beta) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*}) = \mu$$

Since $\theta_\alpha \in B_2$, we have $w^*(\theta_\alpha) = \theta_\alpha$ and since $\theta_\beta \in B_1$, we have $w^*(\theta_\beta) < \theta_\beta$. Furthermore, we have $(\theta_\alpha - w^*(\theta_\alpha) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*}) < (\theta_\beta - w^*(\theta_\beta) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*})$. This implies that $(1 - m^*(\theta_\alpha)) > (1 - m^*(\theta_\beta)) \implies m^*(\theta_\alpha) < m^*(\theta_\beta)$. However, equation A.1 tells us that the agent's best response at any $\theta \in \Theta$, $m^*(\theta)$, is strictly increasing in $w^*(\theta)$, and since $w^*(\theta_\beta) < w^*(\theta_\alpha)$ we must have $m^*(\theta_\beta) < m^*(\theta_\alpha)$. Since this inconsistency is for any $\theta_\alpha \in B_2$ and any $\theta_\beta \in B_1$, and B_1 and B_2 have strictly positive measures, we get a contradiction to the assertion that $(w^*, \bar{w}^*)$ is an optimal contract. Therefore $\bar{a}_2 \leq \underline{a}_1$.

Finally note that since $\bar{a}_0 \leq \underline{a}_1$, $\bar{a}_0 \leq \underline{a}_2$, $\bar{a}_2 \leq \underline{a}_1$ and that $\sum_{i=0}^{2} \lambda([\underline{a}_i, \bar{a}_i]) = \lambda(\Theta)$, we must have $\bar{a}_0 = \underline{a}_2$ and $\bar{a}_2 = \underline{a}_1$ and therefore, there exist $\theta_0 := \bar{a}_0$ and $\theta_1 := \bar{a}_2$ satisfying the claim. □

Lemma 11. *(Temporary. We will have a more precise result)Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition, then there exists a cutoff $\hat{\theta} \in \Theta$, such that either*

$$w^*(\theta) \begin{cases} = 0 \ \ a.e. \quad if\ \theta < \hat{\theta}, \\ \in (0, \theta) \ \ a.e. \quad if\ \theta > \hat{\theta}, \end{cases}$$

or,

$$w^*(\theta) \begin{cases} = \theta \ \ a.e. \quad if\ \theta < \hat{\theta}, \\ \in (0, \theta) \ \ a.e. \quad if\ \theta > \hat{\theta}. \end{cases}$$

Proof. Follows from Lemma 9 and Lemma 10. □

Lemma 12. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition, then, $\lambda(A_0) > 0$.*

Proof. Suppose that $\lambda(A_0) = 0$. Let m^* be the best response of the agent to $(w^*, \bar{w}^*)$. From Lemma 11 we know that such a contract will have the following structure for some $\hat{\theta} \in \Theta$

$$w^*(\theta) \begin{cases} = \theta \text{ a.e.} \quad \text{if } \theta < \hat{\theta}, \\ \in (0, \theta) \text{ a.e.} \quad \text{if } \theta > \hat{\theta}. \end{cases}$$

From Lemma 8, we know that w^* is strictly increasing on A_1 and A_2 since $m^*(\theta) \in (0, 1)$ a.e., which implies that w^* is strictly increasing on Θ a.e.. Without loss of generality, we will assume that $\underline{\theta} \in A_1 \cup A_2$. We consider two possibilities separately.

- Case 1: $\underline{\theta} > 0$: Note that since $\underline{\theta} \in A_1 \cup A_2$, we have $w^*(\underline{\theta}) > 0$. Importantly, note that $\bar{w}^* > w^*(\underline{\theta})$, otherwise $w^*(\theta) \geq \bar{w}^*$ a.e. and therefore, the agent chooses $a = 1$ without acquiring information. Consider the perturbed contract given by

$$\hat{w}(\theta) = w^*(\theta) - \epsilon,$$
$$\hat{\bar{w}} = \bar{w}^* - \epsilon,$$

where ϵ is chosen in such as way that $w^*(\underline{\theta}) - \epsilon > 0$, i.e., $\underline{\theta} - \epsilon > 0$ and $\bar{w}^* - \epsilon > 0$.[11] Note

[11] It is possible that that $(\hat{w}, \hat{\bar{w}})$ does not satisfy the limited liability constraints of the players on a set of measure zero, but that doesn't affect the result since we can redefine $(\hat{w}, \hat{\bar{w}})$ to satisfy these constraints on this measure zero set without affecting the payoffs of any player.

that it is possible to find such an ϵ since from Lemma 2 we know that $\bar{w}^* > 0$. Further note that since $\hat{w}(\theta) - \hat{\bar{w}} = w^*(\theta) - \bar{w}^*$ a.e. on Θ, from Proposition **??** and Equation A.1 we know the agent's best response remains unchanged under the perturbed contract. However, the principal is strictly better off in the perturbed contract and therefore $(w^*, \bar{w}^*)$ cannot be optimal.

- Case 2: $\underline{\theta} = 0$: In this case $w^*(\underline{\theta}) = 0$. Recalling that $\bar{w}^* > 0$, let $\epsilon_1 > 0$ be such that $\bar{w}^* - \epsilon_1 > 0$. Let us define $\hat{\theta}_1$ such that $w^*(\hat{\theta}_1) = \epsilon_1$. Since w^* is strictly increasing, $\hat{\theta}_1$ is unique. We will construct a perturbed contract $(\hat{w}, \hat{\bar{w}})$ that improves on $(w^*, \bar{w}^*)$.

  We now define the following three sets:

$$X_1 = [\underline{\theta}, \hat{\theta}_1),$$
$$X_2 = (\hat{\theta}_2, \hat{\theta}_2 + \epsilon_2),$$
$$X_3 = (\hat{\theta}_1, \hat{\theta}_2) \cup (\hat{\theta}_2 + \epsilon_2, \bar{\theta}]$$

  and the perturbed contract

$$\hat{w}(\theta) = \begin{cases} 0 \text{ if } \theta \in X_1, \\ w^*(\theta) - \epsilon_1 - \delta \text{ if } \theta \in X_2, \\ w^*(\theta) - \epsilon_1 \text{ if } \theta \in X_3, \end{cases}$$
$$\hat{\bar{w}} = \bar{w}^* - \epsilon_1,$$

  for some $\hat{\theta}_2 > \hat{\theta}_1$ and $\delta > 0$ such that $\hat{w}(\theta) \in [0, \theta]$. Note that we can find such a δ because $\exists \gamma \in [\underline{\theta}, \bar{\theta})$ such that for all $\theta \geq \gamma$ we have $w^*(\theta) > \bar{w}^*$, since otherwise, the agent will choose $a = 0$ without acquiring information. Let us denote the best response of the agent to the contract $(\hat{w}, \hat{\bar{w}})$ by $\hat{m}$. The best response of the agent to a contract $(w, \bar{w})$ is given by the following equation which is simply a reformulation of Equation A.1.

$$m(\theta) = f(\mu^{-1}(w(\theta) - \bar{w}) + g'(\mathbb{E}[m(\theta)]).$$

  We define an auxiliary function $t(\hat{w}(\theta))$ as

$$t(\hat{w}(\theta), \hat{\bar{w}}) = f(\mu^{-1}(\hat{w}(\theta) - \hat{\bar{w}}) + g'(\mathbb{E}[m^*(\theta)])). \quad \text{(A.3)}$$

Substituting the definition of $\hat{w}(\theta)$ and $\hat{\bar{w}}$ gives

$$t(\hat{w}(\theta)) = \begin{cases} f(\mu^{-1}(w^*(\theta) - \bar{w}^* + \epsilon_1) + g'(\mathbb{E}[m^*(\theta)])) \text{ if } \theta \in X_1, \\ f(\mu^{-1}(w^*(\theta) - \bar{w}^* - \delta) + g'(\mathbb{E}[m^*(\theta)])) \text{ if } \theta \in X_2, \\ f(\mu^{-1}(w^*(\theta) - \bar{w}^*) + g'(\mathbb{E}[m^*(\theta)])) \text{ if } \theta \in X_3. \end{cases}$$

Now $\mathbb{E}[t(\hat{w}(\theta)) - m^*(\theta)]$ can be written as

$$\begin{aligned}
&\mathbb{E}[t(\hat{w}(\theta)) - m^*(\theta)] \\
&= \int_{\underline{\theta}}^{\hat{\theta}_1(\epsilon_1)} [f(\mu^{-1}(w^*(\theta) - \bar{w}^* + \epsilon_1) + g'(\mathbb{E}[m^*(\theta)])) - f(\mu^{-1}(w^*(\theta) - \bar{w}^*) + g'(\mathbb{E}[m^*(\theta)]))] dP(\theta) \\
&\quad + \int_{\hat{\theta}_2}^{\hat{\theta}_2+\epsilon_2} [f(\mu^{-1}(w^*(\theta) - \bar{w}^* - \delta) + g'(\mathbb{E}[m^*(\theta)])) - f(\mu^{-1}(w^*(\theta) - \bar{w}^*) + g'(\mathbb{E}[m^*(\theta)]))] dP(\theta)
\end{aligned}$$

Denote the first integral on the right-hand side, which is a function of $\hat{\theta}_1$ and therefore a function of ϵ_1 by $T(\epsilon_1)$ and the second term, which is a function of ϵ_2 by $S(\epsilon_2)$. Also note that since f is strictly increasing, $T(\epsilon_1)$ and $S(\epsilon_2)$ are continuous functions with $T(0) = S(0) = 0$ and $T'(\epsilon_1) > 0$ and $S'(\epsilon_2) < 0$. This implies that both T and S are invertible functions in the vicinity of 0. Now we choose $\epsilon_2 = S^{-1}(-T(\epsilon_1))$, which means that $\mathbb{E}[t(\hat{w}(\theta)) - m^*(\theta)] = 0$, which implies that given $\epsilon_1, \hat{\theta}_2$, δ, and with this choice of ϵ_2 we have $\mathbb{E}[\hat{m}(\theta))] = \mathbb{E}[m^*(\theta)]$. Since ϵ_1 can be made arbitrarily small, such $\hat{\theta}_1, \delta$ and ϵ_2 exist. We further note that

$$\left| \lim_{\epsilon_1 \downarrow 0} \frac{\delta}{\epsilon_1} \right| < \zeta \in (0, \infty), \tag{A.4}$$

since f is bounded and differentiable, and P is absolutely continuous and has full support. Next note that since f is differentiable, using the first order Taylor expansion, we have

$$\begin{aligned}
\hat{m}(\theta) - m^*(\theta) =& f(\mu^{-1}(\hat{w}(\theta) - \hat{\bar{w}}) + g'(\mathbb{E}[\hat{m}(\theta)]) - f(\mu^{-1}(w^*(\theta) - \bar{w}^*) - g'(\mathbb{E}[m^*(\theta)]) \\
=& f(\mu^{-1}(\hat{w}(\theta) - \hat{\bar{w}}) + g'(\mathbb{E}[m^*(\theta)]) - f(\mu^{-1}(w^*(\theta) - \bar{w}^*) - g'(\mathbb{E}[m^*(\theta)]) \\
=& - f'(\mu^{-1}(w^*(\theta) - \bar{w}^*) + g'(\mathbb{E}[m^*(\theta)])(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*)) \\
& + o(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*))
\end{aligned}$$

This implies that

$$
\begin{aligned}
|\hat{m}(\theta) - m^*(\theta)| =& | - f'(\mu^{-1}(w^*(\theta) - \bar{w}^*) + g'(\mathbb{E}[m^*(\theta)])(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*)) \\
&+ o(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*))| \\
\leq& |f'(\mu^{-1}(w^*(\theta) - \bar{w}^*) + g'(\mathbb{E}[m^*(\theta)])| \cdot |(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*))| \\
&+ |o(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*))| \\
\leq& \eta|(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*))| + |o(\mu^{-1}(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*))|,
\end{aligned}
\tag{A.5}
$$

for some $\eta > 0$ since f' is bounded above in the relevant domain. Now we calculate the change in the payoff of the principal from offering the perturbed contract

$$
\begin{aligned}
V(\hat{w}, \hat{\bar{w}}) - V(w^*, \bar{w}^*) =& \int_{\underline{\theta}}^{\bar{\theta}} \left[\hat{m}(\theta)(\theta - \hat{w}(\theta) - \theta_B + \hat{\bar{w}}) - m^*(\theta)(\theta - w^*(\theta) - \theta_B + \bar{w}^*)\right] dP(\theta) \\
&+ (\theta_B - \hat{\bar{w}}) - (\theta_B - \bar{w}^*). \\
=& I_1 + I_2 + I_3 + \epsilon_1,
\end{aligned}
$$

where

$$
I_i = \int_{X_i} \left[\hat{m}(\theta)(\theta - \hat{w}(\theta) - \theta_B + \hat{\bar{w}}) - m^*(\theta)(\theta - w^*(\theta) - \theta_B + \bar{w}^*)\right] dP(\theta).
$$

We will now evaluate each of these integrals separately.

$$
\begin{aligned}
I_1 =& \int_{\underline{\theta}}^{\hat{\theta}_1} \left[\hat{m}(\theta)(\theta - \hat{w}(\theta) - \theta_B + \hat{\bar{w}}) - m^*(\theta)(\theta - w^*(\theta) - \theta_B + \bar{w}^*)\right] dP(\theta) \\
=& \int_{\underline{\theta}}^{\hat{\theta}_1} \left[m^*(\theta)(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*) + (\hat{m}(\theta) - m^*(\theta)(\theta - \hat{w}(\theta) - \theta_B + \hat{\bar{w}}))\right] dP(\theta)
\end{aligned}
$$

Note that when $\theta \in X_1$, $|w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*| \leq \epsilon_1$ because Lemma 8 tells us that $\frac{dw^*(\theta)}{d\theta} \leq 1$. Further, using Equation A.5 we have

$$
\frac{|I_1|}{\epsilon_1} \leq \int_{\underline{\theta}}^{\hat{\theta}_1} \left[|m^*(\theta)| + \eta\mu^{-1} + \frac{|o(\mu^{-1}\epsilon_1)|}{\epsilon_1}\right] dP(\theta).
$$

Since $f \in C^\infty$ the integrand is bounded. Further, $\lim_{\epsilon_1 \downarrow 0} \hat{\theta}_1 = \underline{\theta}$, we have $\lim_{\epsilon_1 \downarrow 0} \frac{|I_1|}{\epsilon_1} = 0$ and therefore, $I_1(\epsilon_1) = o(\epsilon_1)$.

Similarly,

$$I_2 = \int_{\hat{\theta}_2}^{\hat{\theta}_2+\epsilon_2} \left[m^*(\theta)(w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*) + (\hat{m}(\theta) - m^*(\theta)(\theta - \hat{w}(\theta) - \theta_B + \hat{\bar{w}})) \right] dP(\theta)$$

Note that when $\theta \in X_2$, $|w^*(\theta) - \hat{w}(\theta) + \hat{\bar{w}} - \bar{w}^*| = \delta$. Further, using Equation A.4 we have

$$\frac{|I_2|}{\epsilon_1} \leq \int_{\hat{\theta}_2}^{\hat{\theta}_2+\epsilon_2} \left[|m^*(\theta)\zeta| + \eta\mu^{-1} + \frac{|o(\mu^{-1}\delta)|}{\epsilon_1} \right] dP(\theta).$$

Since $f \in C^\infty$, integrand is bounded above and $\lim_{\epsilon_1 \downarrow 0} \epsilon_2 = 0$, we have $\lim_{\epsilon_1 \downarrow 0} \frac{|I_2|}{\epsilon_1} = 0$ and therefore, $I_2(\epsilon_1) = o(\epsilon_1)$. Since $m^* = \hat{m}$ and $\theta - \hat{w}(\theta) - \theta_B + \hat{\bar{w}} = \theta - w^*(\theta) - \theta_B + \bar{w}^*$ on X_3, we have $I_3 = 0$.

To complete the argument recall that

$$\begin{aligned} V(\hat{w}, \hat{\bar{w}}) - V(w^*, \bar{w}^*) =& I_1 + I_2 + I_3 + \epsilon_1 \\ =& I_1 + I_2 + \epsilon_1. \end{aligned}$$

Since $I_1 = o(\epsilon_1)$ and $I_2 = o(\epsilon_1)$, there exists an $\hat{\epsilon}_1 > 0$ such that $I_1(\epsilon_1) + I_2(\epsilon_1) + \epsilon_1 > 0$ for all $0 < \epsilon_1 < \hat{\epsilon}_1$, which is a contradiction to the optimality of $(w^*, \bar{w}^*)$.

□

Lemma 13. *Suppose $(w^*, \bar{w}^*)$ is an optimal contract that induces information acquisition, then there exists $\theta^* \in (\underline{\theta}, \bar{\theta})$ such that*

$$\begin{aligned} w^*(\theta) = 0 \text{ if } \theta < \theta^*, \text{ and },\\ \frac{dw^*}{d\theta} = 1 - m^*(\theta) \text{ if } \theta > \theta^*. \end{aligned}$$

Furthermore, $\lim_{\theta \uparrow \theta^*} w^*(\theta) = \lim_{\theta \downarrow \theta^*} w^*(\theta)$, i.e., $w^*(\theta)$ is continuous at θ^*.

Proof. The first part of the lemma is a straightforward consequence of Lemma 9, Lemma 10, and Lemma 12.

To see that $w^*(\theta)$ is continuous at θ, let us assume that it is not, i.e., $\lim_{\theta \uparrow \theta^*} w^*(\theta) < \lim_{\theta \downarrow \theta^*} w^*(\theta)$. Note that the reverse inequality is not possible owing to the fact that w^* is increasing. From the first order necessary condition of optimality given by Equation 6 we know that

$$(1 - m^*(\theta))(\theta - w^*(\theta) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*}) \begin{cases} \leq \mu \text{ a.e.} & \text{if } w^*(\theta) = 0, \text{ i.e., } \theta \in A_0, \\ = \mu \text{ a.e.} & \text{if } w^*(\theta) \in (0, \theta), \text{ i.e., } \theta \in A_1, \end{cases}$$

i.e.,

$$\lim_{\theta \uparrow \theta^*} (1 - m^*(\theta))(\theta - w^*(\theta) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*}) \leq \lim_{\theta \downarrow \theta^*} (1 - m^*(\theta))(\theta - w^*(\theta) - \theta_B + \bar{w}^* + \beta_{\bar{w}^*}) = \mu.$$

Since $\lim_{\theta \uparrow \theta^*} w^*(\theta) < \lim_{\theta \downarrow \theta^*} w^*(\theta)$, using the agent's best response given by Equation A.1, we have $\lim_{\theta \uparrow \theta^*}(1 - m^*(\theta)) > \lim_{\theta \downarrow \theta^*}(1 - m^*(\theta))$, which implies that $\lim_{\theta \uparrow \theta^*}(\theta - w^*(\theta)) < \lim_{\theta \downarrow \theta^*}(\theta - w^*(\theta))$, which implies that $\lim_{\theta \uparrow \theta^*} w^*(\theta) > \lim_{\theta \downarrow \theta^*} w^*(\theta)$, a contradiction. □

These lemmas combined together give the proof of the proposition.