Ignorance is bliss

Latest version: June 10, 2019

Yufeng Sun[1]

Abstract

A principal can select a better project when she has more prior information, she will also hold a strong belief that the selected project is promising. However, given this strong belief, when an agent implements the project and submits a report with negative feedback, the principal will doubt the agent's competence rather than the project's quality and even dismiss the agent. The agent can anticipate this, so he has an incentive to manipulate his disclosure which includes less negative news to avoid being dismissed by the principal. The principal will then suffer a loss from the agent's information distortion, as she cannot immediately adjust the project. This paper shows that if the principal is initially ignorant about the project, the agent has less concern about being judged as incompetent and a greater incentive to provide more information about project progress, including negative news, then the wrong project is more likely to be adjusted immediately. In contrast to the classical principal-agent literature, where the principal can be better off having more information because less information rent is paid to the agent, this paper contends that in a bounded rationality environment, it is possible that for the principal, "ignorance is bliss." She may be better off intentionally choosing "not to know," as the agent then has an incentive to reveal more information which makes the project better-off.

Keywords: Principal-agent, Intentionally ignorance, Strategic disclosure
JEL Classification: D80, D82, D83, D91

[1]School of Public Economics and Administration, SUFE. **E-mail** sunyufeng2007@gmail.com.

"Do not judge, or you too will be judged." - New Jerusalem Bible, Matthew 7:1

1 Introduction

In many daily examples of principal-agent relationships (e.g., a VC and the CEO of a startup, a professor and a research assistant, or a leader and a follower), when the principal identifies a promising project, she will assign it to an agent to implement it. The agent will obtain immediate feedback about the project's quality as it progresses. When the agent reports this feedback, on the one hand, the principal will consider whether to maintain the initial project; on the other hand, the principal will also judge the agent's ability and consider whether the agent is competent to complete the project task. However, the agent can anticipate the principal's judgement, so the agent has an incentive to manipulate his report to avoid being dismissed by the principal. However, this information manipulation by the agent, driven by the principal's potential judgement, may generate distortion.

In business, when a VC invests in a promising project and assigns it to a professional manager to operate, if the manager submits a report that project experiences difficulties, the VC will attribute these difficulties to two main sources: On the one hand, the project itself may not be very good; on the other hand, the manager may be incompetent. The VC might replace the manager and maintain the initial project rather than adjust it. Thus, the manager has an incentive to manipulate his feedback on the project's progress. Verrecchia (1983) shows that the manager of a risky asset exercises discretion in information disclosure and strategically withholds unfavorable reports to convince the trader that he can provide a good service. Abrahamson and Park (1994) demonstrates that it is common the negative organizational outcomes are concealed from corporate officers.

In politics, when a political leader assigns a political project to a bureaucrat, based on the feedback received as the project progresses, the leader will evaluate the bureaucrat's ability and loyalty and decide whether to replace the the bureaucrat. The bureaucrat thus has an incentive to conceal bad news and report good news, even to the extent of inventing fake good news. A famous example is "The Great Leap Forward" in China from 1958 to 1961, when Chairman Mao held a strong belief and confidence in his plan for "The Great Leap Forward", while in practice, local governors pandered to Chairman Mao's will and concealed negative news; the result was that the incorrect direction of the plan was not immediately adjusted, eventually leading to a great famine (See Kung and Chen (2011); Meng et al. (2015); Fan et al. (2016)). Similar phenomena are common in autocratic and dictatorial regimes.

This paper develops a simple model to show that in a general environment, the principal's assessment of the agent's ability will induce strategic information disclosure by the agent to avoid being judged as incompetent. The agent's strategic information disclosure will generate an information distortion, then the project may be not adjusted efficiently due to the transmission of distorted information.

Then, a natural question is, how should one principal efficiently induce the agent to reveal informative project feedback signals ? This paper shows that if the principal can rationally remain naive or ignorant about a project and prevent prior bias, this will help to mitigate the agent's information distorting behavior and encourage the agent to efficiently reveal information. As the maxim goes, "ignorance is bliss"; this paper highlights the value of ignorance in the principal-agent

problem and is also consistent with the insights mentioned in Smithson (1993) that ignorance can serve as a prerequisite to learning or discovery and in McGoey (2012)

> *"..... ignorance serves as a productive asset, helping individuals and institutions to command resources, deny liability in the aftermath of crises, and to assert expertise in the face of unpredictable outcomes."*

In this model, in the first step, the principal can choose to acquire some prior information that indicates the potential benefit of a project. Based on this prior information, the principal selects one promising project and assigns it to an agent who has the professional skill to implement this project. The principal cares about the success of the project, and the agent cares about how to avoid being fired from implementing the project if possible.

In the second step, there is an experiment period during which the agent will run the project as an experiment and receive some soft feedback information related to project quality. The agent has two possible types: competent and incompetent. The competent agent is more likely to obtain feedback that includes strong evidence and reflects the project's quality, while the incompetent agent is more likely to receive noise that cannot indicate the project's quality. Based on the feedback signal, the agent will submit a report that recommends that the principal maintain or modify the initial plan for the selected project. Based on the agent's report, the principal will evaluate the project's quality and the agent's competence and then consider whether to change the project or replace the agent. Then, the principal ultimately determines the project and the implementer, and the project is formally implemented in the third step and the output realized.

The agent will consider how to report the feedback: If the agent reveals all of its details honestly, as the principal's prior information indicates the initially selected project is more likely to be correct, when the feedback is negative, the principal may attribute the negative feedback to the agent's ability rather than the project itself. As the principal will judge the agent's competence and the feedback information is soft, to avoid being being fired, the rational agent will disclose positive feedback, while negative feedback is more likely to be concealed.

In this model, prior project information plays two roles for the principal: The first is to guide the principal in learning about project quality and select the initial project, and better prior information helps to select a better project. The second role is to help the principal to judge the agent's competence, as a competent agent is more likely to receive the correct feedback signal; if the agent always tells the truth, more prior information can also help the principal to more easily determine whether the agent is competent. If the agent always discloses all of his information, the principal having more prior information should have positive effects on both roles. However, due to the agent's strategic disclosure driven by the principal's judgement, it is difficult for the principal to identify an incompetent agent through the agent's manipulated reports. Even worse, the more prior information that the principal has, the more distorted the information in the agent's report.

Thus, the principal faces the following trade-off: If the principal acquires more initial information, she will be more likely to select a good project, but because she believes that her selection is more likely to be correct in this case, her strong belief will depress the agent's incentive to report negative feedback, which will ultimately reduce her ability to adjust a project that is moving in the wrong direction. If the principal initially remains naive, it will be more difficult for her to select a good initial project, but the agent will not be concerned about the principal's judgement and will report

feedback immediately, meaning that a project on the wrong track could be adjusted efficiently. The tradeoff thus reduces to the following: better initial project selection based on prior information or more efficient project adjustment based on the agent's report. The paper shows that when the available prior information is bounded, it is optimal for the principal to remain naive and ignorant with less prior information, which encourages the agent to reveal negative feedback and induce efficient project adjustment.

2 Literature review

This paper relates to four main branches of the literature: the literature on the value of ignorance in principal-agent relationships, the information design literature, the leadership in organization and the related applied management and psychology literatures.

This paper provides a new perspective to understand the value of ignorance. Lewis and Sappington (1993) and Kessler (1998) show that an agent remaining intentionally uninformed or ignorant about the cost structure of a project before signing a contract can obtain more rent from the principal in the formal contract, as the agent's ignorance about the project creates greater uncertainty, which creates a larger information rent space, and thus the agent's ignorance provides him with a bargaining advantage. In Brocas and Carrillo (2007), the principal can control the public information flow available to the agent, and when the cumulative information is sufficient to induce the agent to choose the principal's preferred action, the principal will cancel all future information flow and create common ignorance among the public. In this situation, the principal creates the agent's ignorance to induce the agent to choose the the principal's preferred action.

In contrast to previous work, this paper demonstrates the rationality of the principal intentionally making herself ignorant. The main channel operates as follows: The principal's optimism based on having access to more information will depress the agent's incentive to reveal negative feedback, as negative feedback may induce the principal to doubt the agent's competence. Then the principal's rational ignorance acts as a tool to incentivize the agent to provide more informative feedback, especially negative news, which indicates that the principal needs to act to stop the loss.

In this paper, the principal will judge the agent's competence based on the agent's report about feedback received, and because the feedback is soft information that is difficult to verify, the agent will manipulate the collected feedback. Thus, similarly to Crawford and Sobel (1982), this paper discusses the strategic communication problem. In contrast to Crawford and Sobel (1982), where the receiver and sender have different first-best actions driven by differences in preferences, in this paper, the principal and agent share a common interest in the project's success. The main incentive for the agent to engage in information manipulation is to avoid being judged incompetent by the principal. In this situation, the agent faces a persuasion problem similar to the argument in Kamenica and Gentzkow (2011): The agent will attempt to manipulate the signal to persuade the principal to believe that he is more likely to be a competent type given the Bayesian plausibility constraint. As the survey in Garicano and Rayo (2016), the communication distortion in an organization where the principal and agent share some common interests has few discussion, this paper contributes the analysis from a new perspective.

In the classical principal-agent literature, many fundamental works, e.g.,Hölmstrom (1979); Grossman and Hart (1983); Holmstrom and Milgrom (1991), argue that when the principal's infor-

mation is inferior, she has to pay an information rent to the agent. In contrast to these classical works, this paper argues that the principal acquiring more prior information may backfire: The principal's prior judgement will depress the agent's incentive to honestly report the feedback signal. This paper demonstrates that in some cases, the principal pursuing a strategy of rational ignorance may produce better results.

This paper also helps to understand the art of leadership in an organization. Similar with the spirit of "Yes men" in Prendergast (1993) where the agent will guess and pander the principal's belief, this paper also shows the distortion generated as the agent's strategic report to keep alignment with the principal's prior belief. Bolton et al. (2012) emphasizes the importance of the leader's overconfidence in the organization to improve the follower's coordnation, this paper shows that the intentional ignorance of a leader can evokes the follower's informative information disclosure. Dessein et al. (2016) shows the value of rational inattention in the projects organization and this paper shows the value of ignorance in the project management.

From an applied perspective, the theory advanced in this paper is consistent with the empirical findings in the management and psychology literatures. Kluger and DeNisi (1996), Roberts (2013) and Hertwig and Engel (2016) show that principals' deliberate ignorance can serve as a performance-enhancing device. In this paper, ignorance enhances performance via encouraging the agent to informatively report feedback with less distortion.

The remainder of the paper is organized as follows. Section 3 introduces the model. Section 4 will complete the equilibrium analysis step by step following backward induction. Section 5 completes the welfare analysis and demonstrates the value of the principal's ignorance to the project's outcome. Section 6 presents three simple extensions, and section 7 concludes the paper. All omitted proofs are relegated to the appendix.

3 Model

3.1 Setup

3.1.1 Project selection

There are two possible states $\theta \in \{0, 1\}$ with the same prior likelihood. The goal of the principal ("she") is to select and implement one project that matches the true state. Initially, the principal receives an informative signal τ that indicates the true state.

$$P(\theta = \tau) = q \in (\frac{1}{2}, 1)$$

The principal can choose to read this signal or not: If the principal reads this signal, it is natural that the principal will select one project $\rho(\tau) = \tau \in \{0, 1\}$ following the informative signal τ. If the principal does not read this signal, the principal will select one project $\rho(\varnothing) \in \{0, 1\}$ randomly with same likelihood. Eventually, the initially selected project $\rho \in \{\rho(\tau), \rho(\varnothing)\}$.

There is an agent pool with infinite agents, and agents ("he") are of two types: competent and incompetent. The agent is a competent type with probability $p \in (\frac{1}{2}, 1)$. Every agent knows this competence type distribution but does not know his own type exactly.

The principal will select one agent from this pool at random and assign project ρ to this agent.

The difference in competence between the two types of agents has two dimensions: feedback collection ability during the experiment and operational ability in practice. We will describe the influence of the difference in agent competence in the following part.

The project is completed by the agent in two steps: the experimental step and the final implementation step.

3.1.2 Project experiment

In the experimental step, the agent attempts to operate the project through trial and error and collects some initial feedback that indicates the project's quality. After engaging in this trial and error, the agent receives a feedback signal for the project.

Specifically, the agent can collect feedback from the project experiment. The agent receives an informative signal s_I with probability ϕ and a noisy signal s_n with probability $1-\phi$. $\phi \in \{1, \epsilon\}$ measures the agent's competence: When the agent is competent, $\phi = 1$, and the received signal s is always informative. When the agent is incompetent, $\phi = \epsilon < 1$, and the received signal is a mixture of informative signal and noise.

$$s = \begin{cases} s_I & s \sim f(s) \text{ with probability } \phi \in \{1, \epsilon\} \\ s_n & s \sim g(s) \text{ with probability } 1-\phi \end{cases}$$

Throughout the paper, we assume that the agent does not know his competence ex ante and only has the common prior belief about the competence distribution.

The informative signal s_I comes from a distribution $F(s)$ with density $f(s)$ as follows:

$$f(s) = \begin{cases} f_+(s) & s \in [0, \bar{\omega}], \text{If the selected project is correct} \\ f_-(s) & s \in [\underline{\omega}, 0], \text{If the selected project is wrong} \end{cases}$$

where $\bar{\omega} > 0$ can be $+\infty$, and $\underline{\omega} < 0$ can be $-\infty$.

The noisy signal s_n comes from a distribution $G(s)$ with density $g(s)$, where $s \in [\underline{\omega}, \bar{\omega}]$. Thus, the sign of the informative signal s_I directly indicates the project selection.

The information structure of the feedback described above captures the first dimension of the difference in competence between competent and incompetent agents: The competent agent is more likely to receive the informative signal than the incompetent agent. Furthermore, when the signal is informative, i.e., $s \sim f(s)$, the magnitude of the feedback $|s|$ can be interpreted as the strength of the feedback or the evidence related to project quality. A larger $|s|$ indicates stronger positive evidence (if $s > 0$) to support the project or more negative evidence (if $s < 0$) in opposition to the project.

3.1.3 Agent's report and principal's judgement

After the experiment, the principal will require the agent to submit a report, and she will rely on the agent's report to assess the agent's competence and the project's quality.

Based on the feedback $s \in \Omega$, the agent will choose one report: A report π consists of a finite message space M and a family of distributions of $\{\pi(\cdot|s)\}_{s\in\Omega}$ over M, based on the feedback signal received by the agent. Specifically, this paper will focus on a special report form: a partition report. Similar with Crawford and Sobel (1982), we can define the agent's partition report as the following

Definition 1 (Partition report)**.** Let $s(n) = (s_0, \ldots, s_n)$ denote a partition of the signal space $\Omega = [\underline{\omega}, \bar{\omega}]$ with n steps and dividing the signal between $s_0, \ldots, s_n$, where $\underline{\omega} = s_0 < \cdots < s_n = \bar{\omega}$. The **partition report** π is

$$\pi(m_i|s) = \begin{cases} 1 & s \in \Omega_i \\ 0 & s \notin \Omega_i \end{cases}$$

where $\Omega_i = (s_i, s_{i+1}), i = 0, \cdots, n-1$, and $M = \cup_{i=0}^{n-1} m_i$.

The partition report is a degenerate form of the general report. In the partition report, the partition $s(n)$ is endogenously chosen by the agent, and the message m_i is associated with the signal interval $\Omega_i = (s_i, s_{i+1})$.

The the principal observes the agent's reporting choice and will receive a realized message $m \in M$; then based on m, she will form a belief vector $\mu(m) = (\hat{q}(m), \hat{p}(m))$, where $\hat{q}(m)$ is the posterior belief about the project's quality, and $\hat{p}(m)$ is the the posterior belief about the agent's competence.

In summary, based on the message m delivered by the agent, the principal will form the corresponding posterior belief about the project quality's and the agent's competence and complete two adjustment decisions:

(1) $\lambda(\mu) \in \{0, 1\}$: Maintain the initially selected project ($\lambda = 0$) or modify the project ($\lambda = 1$). Given the adjustment decision, the final project $\hat{\rho}(\lambda)$ is determined.

(2) $\sigma(\mu) \in \{0, 1\}$: Retain and trust the initially selected agent ($\sigma = 0$) or choose a new agent from the agent pool ($\sigma = 1$).

3.1.4 Project implementation

In the final implementation step, the agent faces two effort levels: $e_a \in \{0, 1\}$. The second dimension of the difference in competence between competent and incompetent agents is the difference in their project implementation efficiency. The final project outcome depends on the combination of the agent's effort and competence. The project outcome is denoted W.

When the project is correct and $e_a = 1$, the competent agent will perfectly implement the correct policy and generate a payoff of 1, while the incompetent agent will generate a discounted payoff κ. When the project is wrong and $e_a = 1$, the project will always generate a loss of -1 for certain regardless of the agent's competence. When the agent chooses $e_a = 0$, the project outcome is always 0.

In the implementation process, the agent will face a random implementation cost $c \in \{0, \eta\}$, where $c = \eta$ with probability ξ. The principal can always monitor the agent to take effort $e_a = 1$ to eventually implement the final project $\hat{\rho}(\lambda)$.

3.1.5 Payoff

The principal's payoff comes directly from the project outcome W; specifically, a good project outcome requires two elements: The final implemented project is good, and the project is implemented by the competent agent. The principal always attempts to improve project and implementer selection as much as possible.

$$\max_{\{\rho, \sigma, \lambda\}} E[W|\rho, \sigma(\mu), \lambda(\mu)] \tag{1}$$

The agent's payoff includes two main elements: First, the agent can enjoy a positive rent r during the project implementation process if he can manage the project ($\sigma(\mu) = 1$). Second, conditional on the agent not being fired ($\sigma(\mu) = 1$), the agent will enjoy a positive perk r from managing the project, and the project outcome also enters his payoff function with weight α. Thus, if the agent is not fired, he also cares about the final project outcome[2]. Once the agent is fired, his payoff from his outside option is normalized to 0. Thus, the agent's objective function is:

$$\max_{\pi} E\{\sigma(\mu(m)) \cdot (r + \alpha E[W|e_a, \lambda(\mu(m))]) - c \cdot 1(e_a = 1)\}$$

where r is the fixed rent that the agent can obtain from managing one project, α is the weight of the project outcome in the agent's payoff, and $c \in \{0, \eta\}$ is the random cost that is the agent's private signal when the agent exerts effort $e_a = 1$.

In the benchmark analysis, the principal has real and centralized authority, and she always encourages the agent to exert effort $e_a = 1$, and thus, the agent always faces an expected cost $E(c) = \xi\eta$. Dropping the fixed constant term, the agent's problem in the benchmark analysis is equivalent to

$$\max_{\pi} E\{\sigma(\mu(m)) \cdot (r + \alpha E[W|e_a = 1, \lambda(\mu(m))])\} \tag{2}$$

In summary, based on the above description, the timeline of this game is as follows.

Step 1: The principal chooses to read or ignore the initial signal τ; then she chooses one project $\rho \in \{\rho(\tau), \rho(\varnothing)\}$ and assigns it to a randomly selected agent to implement.

Step 2: In the experimental step, the agent will collect feedback information, and the principal will require the agent to submit a report, the agent commits a report choice $\{\pi(\cdot|s)\}_{s\in\Omega}$ over a measage space M: when the agent receives feedback s, he will draw a message m from $\pi(\cdot|s)$ and submit m to the principal.

Step 3: Given received signal s or message m, the principal will form a posterior belief and make the project adjustment decision $\lambda(\mu)$ to determine the final project and the agent adjustment decision $\sigma(\mu)$.

Step 4: The initial or new agent will implement the final project determined in step 3, and the project outcome W is realized.

3.2 Strategy and equilibrium

Our solution concept is perfect Bayesian equilibrium, which we refer to simply as an *equilibrium*. An equilibrium is characterized by a strategy-belief pair

$$\langle \sigma, \mu \rangle = \{(\rho, \pi, \lambda(\mu), \sigma(\mu)), \mu\}$$

where

- $\rho \in \{\rho(\tau), \rho(\varnothing)\}$ is the project initially selected by the principal.
- π is the report choice committed by the agent.
- $\mu = (\hat{p}, \hat{q})$ is the principal's belief about the agent's ability and project quality based on the received signal s or message m and the report π chosen by the agent.

[2]In practice, the principal may offer a bonus schedule associated with the project outcome; we use αW to simply capture this characteristic.

• $\lambda(\mu)$ is the principal's project adjustment decision, and $\sigma(\mu)$ is the principal's agent adjustment decision .

In equilibrium, the strategy-belief pair $\langle \sigma, \mu \rangle$ should satisfy the following conditions:

• Given the quality of the principal's initially selected project $\rho(\tau)$, when the agent receives a signal s, the agent will anticipate that reporting choice π will induce the principal to form a belief $\mu = (\hat{p}, \hat{q})$; the agent will choose a report π to maximize his expected payoff. (2)

• Given the agent's reporting strategy π, when the principal receives the signal s or message m, she will form the belief $\mu = (\hat{p}, \hat{q})$ following Bayes' rule, where $\hat{p}$ is the principal's belief about the agent's ability, and $\hat{q}$ is her belief about the project's quality.

• Given the agent's reporting choice π and the posterior belief $\mu = (\hat{p}, \hat{q})$, the principal will choose an optimal strategy $(\rho, \lambda(\mu), \sigma(\mu))$ to maximize the expected project outcome $E[W]$.

4 Equilibrium analysis

To facilitate the analysis, we introduce some restrictions on the agent's signal structure for the feedback below.

Assumption 1. Specifically, we impose the following assumption on the agent's feedback structure:

(1) $g(s)$ is symmetric around 0.

(2) $g(s)/f_+(s)$ is non-increasing in s and $g(0)/f_+(0) \geq 1$.

(3) $g(s)/f_-(s)$ is non-decreasing in s and $g(0)/f_-(0) \geq 1$.

In the above assumption, part (1) means that the incompetent agent may receive noise with probability $1 - \epsilon$ around 0 during the experimental step, while the signal that the competent agent receives is always correct in terms of direction. Parts (2) and (3) mean that the competent agent is more likely to receive a strong informative signal or "hard evidence" (larger $|s|$) than the incompetent agent, who only has probability ϵ of receiving an informative signal s_I from $f(s)$. This assumption will be applied throughout the analysis below.

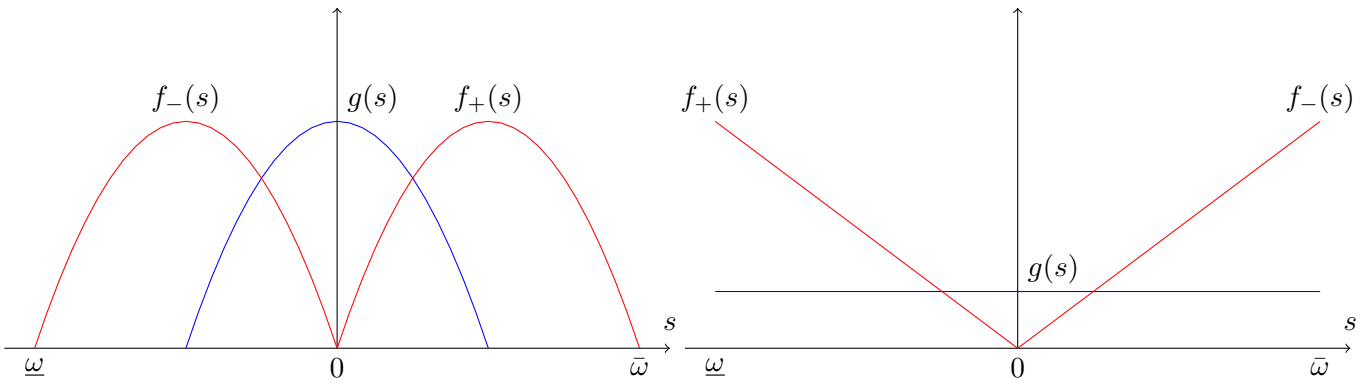

Figure 1: Examples of feedback signal structure

In the following analysis, without loss of generality, we focus on the case in which $\tau = 1$; the analysis of the case in which $\tau = 0$ is similar.

The entire analysis will follow backward induction, given the difference in implementation efficiency between the two types of agents.

4.1 Principal's optimal adjustment rule

After the experiment step, the principal will require the agent to deliver a message m, and conditional on the message m delivered by the agent, the principal's optimal adjustment decision about the project and the agent will be as follows.

(1) Project adjustment decision: $\lambda(m) = 0 \Leftrightarrow \hat{q}(m) \geq \frac{1}{2}$.

(2) Agent replacement decision: $\sigma(m) = 0 \Leftrightarrow \hat{p}(m) \geq p$.

If the principal decides to replace the initially selected agent, the probability that she selects a new and competent agent from the agent pool is p; given the outside option, the principal will not replace the agent if she infers that the initially selected agent is competent with probability not less than the average quality p of agents in the pool. This adjustment/replacement rule imposes a restriction on the agent's information disclosure.

We now return to the agent's optimization problem:

$$\max_{\pi} E\{\sigma(\mu(m)) \cdot (r + \alpha E[W|e_a, \lambda(\mu(m))] - c \cdot 1(e_a = 1))\}$$

To focus on the analysis of the most interesting case, we introduce the following assumption 2.

Assumption 2.

$$r - \alpha - \xi\eta > 0$$

Assumption 2 means that not being fired is the agent's primary concern. As the agent's primary concern is to avoid being fired, to explore the agent's optimal reporting choice, we define the set of so-called safe reports as follows:

$$\mathcal{F} = \{\pi : \sigma(\mu(m)) = 1, \forall m \in M\}$$

Given the the principal's adjustment rule $\sigma(\cdot)$, if $\mathcal{F} \neq \emptyset$, the report chosen from $\mathcal{F}$ dominates all other possible reporting choices. This report could guarantee that the agent is able to maintain his position and enjoy more benefits. Thus, if $\mathcal{F} \neq \emptyset$, the agent will always attempt to choose a report from this safe report set.

If the principal chooses $\sigma(m) = 1$ iff $\hat{p}(m) \geq p$, then the safe report set $\mathcal{F}$ can also be represented as

$$\mathcal{F} = \{\pi : \hat{p}(m) \geq p, \forall m \in M\}$$

which means the every possible message m under report π can induce the principal to form the belief that the current reporting agent's competence is not worse than the average level in the agent pool.

Given the principal's optimal adjustment strategy for the project and agent characterized in this subsection, the following subsections 4.2 and 4.3 will focus on the analysis of the agent's reporting strategy.

4.2 Full information disclosure

This subsection will show that full disclosure of the signal s received by the agent, i.e., $m = s$, cannot be a best response to the principal's optimal adjustment strategy, i.e., the principal will not replace the agent iff $\hat{p}(m) \geq p$.

4.2.1 Illustrated example

First, let us consider a simple example. Suppose that all feedback signals come from a uniform distribution: The informative feedback signal s_I comes from $f(s)$ as follows, and the competent agent will always receive $s_I \sim f(s)$.

$$f(s) = \begin{cases} \frac{1}{2} & s \in [0,2], \text{If the selected project is correct} \\ \frac{1}{2} & s \in [-2,0], \text{If the selected project is wrong} \end{cases}$$

and the noisy signal feedback comes from $g(s)$ as follows. For simplicity, this example only considers the extreme case in which $\epsilon = 0$, where the incompetent agent will always receive a noisy signal $s_n \sim g(s)$.

$$g(s) = \begin{cases} \frac{1}{2} & s \in [-1,1] \\ 0 & \text{Otherwise} \end{cases}$$

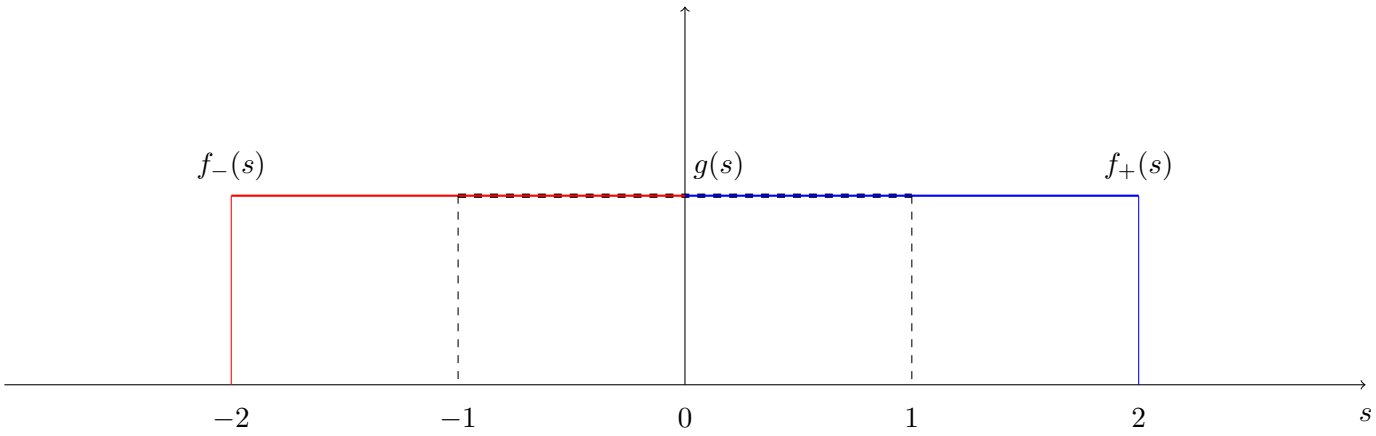

Figure 2: Uniform feedback signal

If the agent truthfully reports the signal s, i.e., $m = s$, then the principal will form her belief about the correctness of the project as follows:

$$P(\theta = 1 | m = s) = \begin{cases} 1, & \text{if } s \in (1,2] \\ \frac{q}{pq+(1-p)} > \frac{1}{2}, & \text{if } s \in [0,1] \\ \frac{q(1-p)}{p(1-q)+(1-p)}, & \text{if } s \in [-1,0] \\ 0, & \text{if } s \in [-2,-1) \end{cases}$$

Note that when $q < \frac{1}{2-p}$, provided that the agent truthfully reports the feedback signal, the principal believes that the initial project is more likely to be correct when $m = s > 0$ and believes that the initial project is incorrect only when $m = s < 0$.

Furthermore, the principal will also form a belief about the agent's competence as follows:

$$P(H|m=s) = \begin{cases} 1, & \text{if } s \in (1,2] \\ \frac{pq}{pq+(1-p)} < p, & \text{if } s \in [0,1] \\ \frac{p(1-q)}{p(1-q)+(1-p)} < p, & \text{if } s \in [-1,0] \\ 1, & \text{if } s \in [-2,-1) \end{cases}$$

When the principal receives the message $m = s \in [-1, 1]$, the feedback signal is weak, and the principal will believe that the agent is more likely to be incompetent, at least, that the agent's competence is lower than the average. As the agent's competence influences the project's implementation efficiency, it is optimal for the principal to replace the current agent with a new agent from the agent pool to implement the project. When the principal receives the message $m = s \in [-2, -1) \cup (1, 2]$, the signal is strong enough, and the principal will be certain that the agent is competent; in this situation, it is optimal for the principal to retain the current agent as the project implementer. Given the principal's concern about finding the optimal replacement, an agent who receives feedback $s \in [-1, 1]$ has an incentive to claim that he received feedback $s \in [-2, -1) \cup (1, 2]$, and thus, the agent will not always truthfully report his signal.

4.2.2 General case

In the general situation, we use the following lemma 1 and 2 to show the influence of feedback signal on the judgement about project quality and agent's competence.

Lemma 1. Given assumption 1, fixed $\forall \epsilon > 0$, when the agent receives $s > 0$, we have the following:

(1) The agent believes that the policy is more likely to be correct, i.e., $P(\theta|s) > \frac{1}{2}$.

(2) If the agent truthfully reports $s > 0$, the principal's inference about the agent's ability $\hat{p}(s) \geq p$ iff $s \geq \bar{s}(q) > 0$.

Given the agent always tell truth, the principal and agent will share the common information about the project. Any positive feedback will make the principal believe that the project is more likely to be profitable. However, driven by the assumption 1, the principal will only believe that the agent is more competent when the signal shows a strong evidence to support the project.

Lemma 2. Given assumption 1, when the agent receives signal $s < 0$: If the agent always truthfully reports $s < 0$, the principal will believe that the agent's ability is higher than the average level iff $s < \underline{s}(q) < 0$.

When the feedback signal is negative, as the principal may have prior bias in the project, the principal may still believe that the project is more likely to be correct if the feedback signal is weakly negative, so the weakly negative signal is not sufficient to falter the principal's confidence in the project, but the weakly nagetive signal will induce the principal's doubt about the agent's competence. The strongly negative signal is more likely to convince the principal to believe that the bad news dues to the wrong project rather than the agent's competence.

Proposition 1. When assumption 1 holds, given the principal's optimal adjustment choice, full information disclosure, i.e., $m = s$, is not a best response.

The proof is relegated to the appendix, we can describe the intuition as follows: As the agent does not know his competence type, the feedback he receives may be informative with probability $p+(1-p)\epsilon$ and may simply be noise with probability $(1-p)(1-\epsilon)$. Given assumption 1, the weak signal (when $|s|$ is small) is more likely to come from a noisy distribution $G(s)$, and the strong signal (when $|s|$ is large) is more likely to come from the informative distribution $F(s)$. When the agent receives a weak signal and discloses it truthfully, he will be judged to be of the incompetent type with high probability. Given the principal's optimal adjustment decision rule, an agent who reports a weak signal is more likely to be fired. When the agent receives a strong signal and discloses it truthfully, he will be judged to be of the competent type with high probability, and the agent could then continue to manage the project given the principal's adjustment rule. Thus, an agent who receives a weak signal has an incentive to claim that he received a strong signal, and thus, truthful reporting cannot be the agent's best response in all situations.

Given the principal's optimal adjustment rule, full information disclosure is not the agent's best response. We will explore the agent's best information disclosure strategy in the following subsection.

4.3 Informative disclosure

It is natural to observe that the babbling report is always a safe report given the principal's adjustment rule described above; the babbling disclosure corresponds to a trivial partition $s(1) = [\underline{\omega}, \bar{\omega}]$, where the agent always submits one message m_0 to the principal regardless of the feedback he receives. Given the agent's babbling disclosure, the principal's belief about the project's quality and agent's competence will not update, and then the agent will also not be replaced. Under babbling disclosure, the probability that the agent is not fired is always one. However, babbling disclosure is uninformative and thus not helpful to the project adjustment after the experimental step.

In the following part, we will show that given the principal's adjustment rule, a safe and informative report exists, and it is helpful for the principal to make the project adjustment.

4.3.1 Illustrated example

Following the previous uniform example, and the previous discussion, full information disclosure is impossible because of the agent's fear of being fired. As the agent's first concern is to avoid being fired, as the principal will replace the agent only when she infers that the agent's competence is lower than the average level, the agent will attempt to choose a report π such that

$$\hat{p}(m) \geq p$$

holds for $\forall m \in M_\pi$, where M_π is the message space under the report π.

In this special uniform case, we can construct a reporting choice as shown in the following figure 3:

When the agent receives a feedback signal $s \in [\frac{1-2q}{q}, 2]$, he delivers message $m_g =$ "Good news" to the principal.

When the agent receives a feedback signal $s \in [-2, \frac{1-2q}{q})$, he delivers message $m_b =$ "Bad news" to the principal.

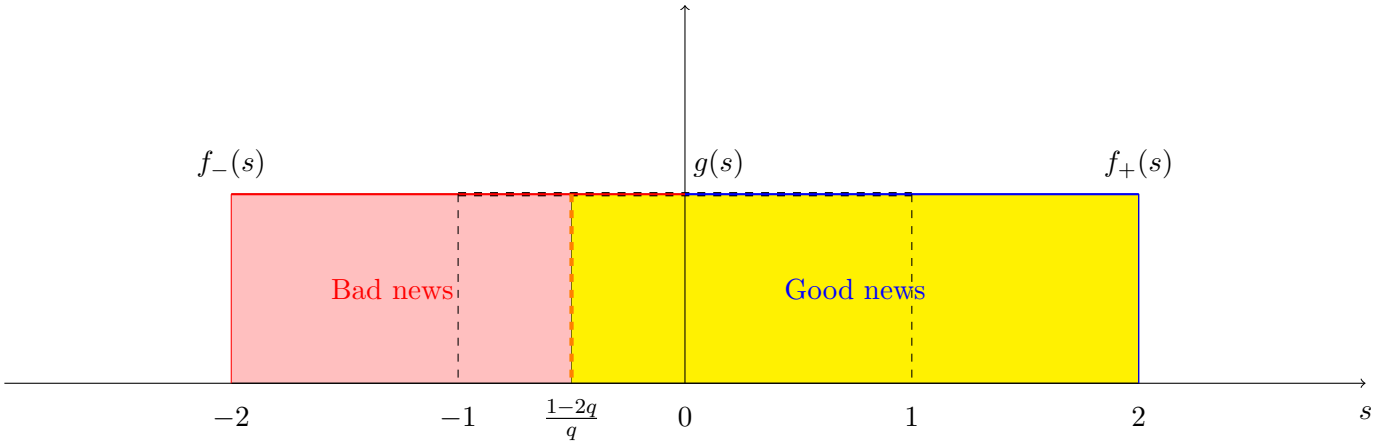

Figure 3: The agent's safe reporting choice

We can check that given the agent's above reporting choice, the principal's belief about the agent's competence is

$$P(H|m_g) = P(H|m_b) = p$$

Thus, the above report π is a safe report for the agent. Under this report, given the principal's replacement rule $\hat{p}(m) \geq p$, the agent will never be fired.

Furthermore, we can also check that the principal forms a belief about the project's quality based on the agent's message as follows:

$$P(\theta = 1|m_g) > \frac{1}{2}, P(\theta = 1|m_b) < \frac{1}{2}$$

Then, given the agent's reporting choice above, the principal will retain the current project when she receives "Good News" and modify the project when she receives "Bad News". The principal will never replace the current agent. Thus, the messages named "Good news" and "Bad news" share similar interpretations with Milgrom (1981).

We can see the distortion induced by the agent's reporting choice above: When the agent receives signal $s \in [\frac{1-2q}{q}, 0)$, he believes that the project is more likely to be incorrect, but if he discloses a feedback signal in this interval, the principal will also doubt his competence, and thus the agent chooses to submit a report with "Good News" when the signal is located in this interval; however, given the principal's belief about the project quality when she receives message "Good News", in this situation, the agent sending the message "Good News" means that the incorrect project cannot be modified immediately. The agent will deliver "Bad News" when the feedback signal is negative enough.

4.3.2 General result

Proposition 2 (Good-Bad news report)**.** Given the principal's adjustment decision rule, $\exists s^*(q) \leq 0$ such that the following best response exists.

(1) Good-Bad news report π:

$$\pi(m=\text{Bad News}|s)=\begin{cases}1 & s\in[\underline{\omega},s^*(q))\\ 0 & s\in[s^*(q),\bar{\omega}]\end{cases}\quad \pi(m=\text{Good News}|s)=\begin{cases}0 & s\in[\underline{\omega},s^*(q))\\ 1 & s\in[s^*(q),\bar{\omega}]\end{cases}$$

where the signal space partition is $s(2)=[\underline{\omega},s^*(q))\cup[s^*(q),\bar{\omega}]$.

(2) Given the agent's Good-Bad news report, based on the received message, the principal will infer the agent's ability as

$$\hat{p}(m)=p,\forall s\in[\underline{\omega},\bar{\omega}]$$

and form the belief about the project as $q(\theta=\tau|m=\text{G})>\frac{1}{2}$ and $q(\theta=\tau|\text{B})<\frac{1}{2}$

The intuition for proposition 2 is as follows: When the principal has a strong prior belief that the initially selected project's quality is good, if the agent receives negative but weak feedback and discloses it honestly, the principal will doubt the agent's competence rather than the project's quality. Thus, the agent has an incentive to report that the project is good when he receives weak negative feedback and only submit a bad news report when he receives strong negative feedback. Driven by the reputational concern of avoiding being considered incompetent, the agent's concealment of weakly negative feedback generates information distortion.

Reputational concerns driven by the principal's judgement induce the distortion of the agent's information disclosure, while the principal's prior knowledge about the project shapes the principal's judgement about the agent's competence type. The principal having less prior bias will help to mitigate the distortion affecting the agent's information disclosure. The following proposition 3 illustrates this point.

Proposition 3. The cut-off value $s^*(q)$ has the following properties.

(1) Range: $\underline{s}(q)<s^*(q)<0$

(2) Monotonicity: $ds^*(q)/dq<0$

Corollary 1. When $\frac{g(0)}{f_-(0)}<\frac{(1-q)(p+(1-p)\epsilon)}{(2q-1)(1-p)(1-\epsilon)}$ holds, a lower q can reduce the distortion in the agent's information disclosure.

The above analysis shows that the good-bad news report strategy can be the agent's best response given the principal's optimal replacement rule. However, equilibria with finer partitions may exist, and a natural question arises: Is there another strategic partition report with less information distortion than the good-bad news report? The following proposition will show that the good-bad news report is already is the most informative partition report with the least information distortion.

Definition 2. Given two reports π_1 and π_2, if

$$E_{\pi_1}\{P(\hat{\rho}=\theta|\lambda(\mu_1(m)))\}\geq E_{\pi_2}\{P(\hat{\rho}=\theta|\lambda(\mu_2(m)))\}$$

We will say the report π_1 is more informative than the report π_2. This definition of the informativeness is direct: The report which can lead to better expected project outcome is the more informative report.

Proposition 4. The Good-Bad news report is the most informative partition report in $\mathcal{F}$.

Proposition 4 shows that the Good-Bad news report is most informative of all possible partition reports that are in the safe set given the principal's adjustment rule. Thus, the optimal equilibrium welfare among the possible partition equilibria can be achieved in the Good-Bad partition equilibrium.

5 Welfare analysis

5.1 Illustrated example

Still following the example in the uniform case, for simplicity, let $\epsilon = 0$ and $\kappa \to 1$; we can calculate the expected project outcome under the *Good-bad news equilibrium* as follows:

$$W_E(q) = 2(1-q)p + \frac{(2q-1)^2}{q}$$

The first observation is that the principal's rational ignorance is indeed helpful to improve the expected project outcome in some situations. We can calculate $dW_E(q)/dq$ as follows:

$$\frac{dW_E(q)}{dq} = -2p - \frac{1}{q^2} + 4 = \begin{cases} < 0 & \text{if} q \in (\frac{1}{2}, \frac{1}{\sqrt{2(2-p)}}) \\ \geq 0 & \text{if } q \in (\frac{1}{\sqrt{2(2-p)}}, 1) \end{cases}$$

Thus, $W_E(q)$ is decreasing in q when $q \in (\frac{1}{2}, \frac{1}{\sqrt{2(2-p)}})$, and we also note that

$$W_E(q) < W_E(\frac{1}{2})$$

holds when $q \in (\frac{1}{2}, \frac{1}{2-p})$. When q is bounded, it is better for the principal to remain naive and ignore the initial signal τ rather than read it.

The second observation is that the experiment step is indeed helpful to improve the expected project outcome due to the collection of feedback.

If there is no experiment step, the expected project outcome will be $2q-1$. We can observe that

$$W_E(q) - (2q-1) = 2p + 2(1-p)q + \frac{1}{q} - 3 \geq 0$$

holds for $\forall q \in [\frac{1}{2}, 1]$ and $\forall p \in [\frac{1}{2}, 1]$.

We can also compare the equilibrium project outcome and the project outcome under full information disclosure. We can take the expected project outcome under the agent's full information disclosure as the benchmark, and then:

$$W_{\text{Full information disclosure}} = \begin{cases} p & \text{if } q \in (\frac{1}{2}, \frac{1}{2-p}) \\ q - (1-q)(1-p) & \text{if } q \in (\frac{1}{2-p}, 1) \end{cases}$$

We observe that the equilibrium expected project outcome is always worse than the project outcome under full information disclosure. This result is intuitive because there is information distortion in

equilibrium.

$$W_E(q) \leq W_{\text{Full information disclosure}}$$

Figure 4 summarizes the comparison of the equilibrium project outcome and the expected project outcome under full information disclosure. We can observe the distortion generated by the agent's strategic disclosure behavior: The information distortion is maximized when $q = \frac{1}{\sqrt{2(2-p)}}$ and gradually disappears as $q \to 1$.

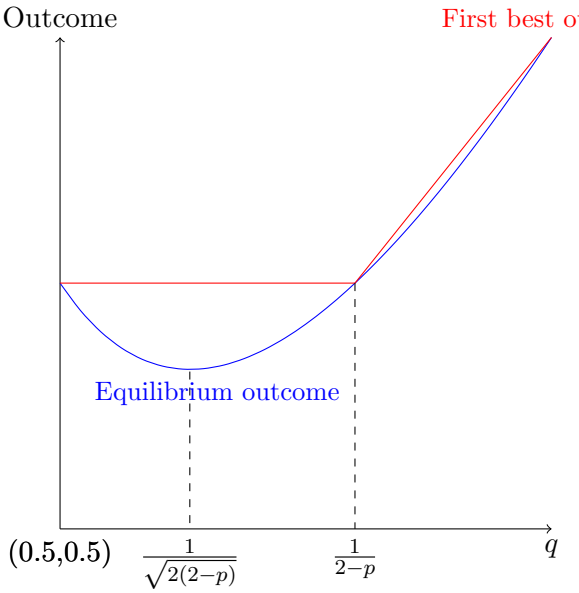

Figure 4: Comparison of project outcomes

The final observation is, as figure 4 shows, when q is large enough, it is still better for the principal to read the initial signal, as the signal is informative enough, and the positive influence of prior knowledge, which is beneficial for the project, dominates the negative influence from the agent's distortion behavior.

The following subsection will show that the above observations hold in a more general environment.

5.2 Welfare

Given the above analysis of the agent's information disclosure choice in the presence of reputational concerns, the principal having less prior bias is helpful to mitigate the distortion in the agent's information disclosure; however, the former having less prior knowledge also leads to worse initial project selection. Thus, obviously, the principal faces a trade-off between initial selection and ex post adjustment: More prior knowledge helps select a better initial project but induces greater distortion in the agent's information disclosure and inefficient ex post adjustment based on feedback; in other words, less prior knowledge induces less distortion in agent reports but worse initial project selection.

This section will show how the principal's prior knowledge and the agent's corresponding reporting choice influence the final project outcome. As proposition 4 demonstrates that the *Good-bad news report* is already the most informative report, the following proposition will focus on the welfare analysis in the equilibrium with the agent's *Good-bad news report*.

Proposition 5. In the *Good-bad news Equilibrium*, we observe that the following properties of social welfare hold for $\forall \epsilon \in (0,1), \forall \kappa \in (0,1)$.

(1) The expected project outcome in the *Good-bad news Equilibrium* is always better than the expected social welfare without an experimental step:

$$W(q,\epsilon,\kappa) \geq q[p+(1-p)\kappa] - (1-q)$$

(2) $\exists\ \underline{q} \in (\frac{1}{2},1)$ such that when $q \leq \underline{q}$ the expected project when the principal reads the initial signal is worse than that when principal initially remains naive

$$W(q,\epsilon,\kappa) < W(\frac{1}{2},\epsilon,\kappa)$$

(3) $\exists\ \bar{q} \in (\underline{q},1)$ such that when $q \in (\bar{q},1)$, then

$$W(q,\epsilon,\kappa) > W(\frac{1}{2},\epsilon,\kappa)$$

The main finding in proposition 5 can be summarized as follows: Part (1) shows that while the experiment is always helpful, feedback will make the expected project outcome better than it would have been under direct implementation. However, the information collected during the experiment may not be fully disclosed because the agent is worried about the principal's judgement. Part (2) shows that when q is limited, the distortion in the agent's disclosure induced by the principal's prior knowledge dominates the improved project selection accuracy derived from the principal having prior knowledge, and the expected project outcome when the principal remains naive is better than that when the principal reads the initial signal and obtains more information. When q is large enough, as part (3) shows, the advantage in selection accuracy generated by the principal having very informative prior knowledge dominates the distortion in the agent's disclosure, and thus it is best for the principal to read the initial signal.

Based on the above analysis, we can see the principal's choice to remain ignorant of the initial signal can improve the expected project outcome when the initial signal's accuracy or quality is limited. We complete the last step of the backward induction and characterize the entire equilibrium of the above game in the following proposition.

Proposition 6. In the above model, the equilibrium can be characterized as follows.

(1) When the available prior knowledge is bounded, i.e., q is limited, we will have the following:

• The principal will ignore the initial signal to remain naive and select one project $\rho = \rho(\varnothing)$, i.e., a project from $\{0,1\}$ at random with probability $\frac{1}{2}$.

• The agent will choose a report π as follows:

$$\pi(m = \text{Bad News}|s) = \begin{cases} 1 & , s \in [\underline{\omega},0) \\ 0 & , s \in [0,\bar{\omega}] \end{cases} \quad \pi(m = \text{Good News}|s) = \begin{cases} 0 & , s \in [\underline{\omega},0) \\ 1 & , s \in [0,\bar{\omega}] \end{cases}$$

where the signal space partition is $s(2) = [\underline{\omega},0) \cup [0,\bar{\omega}]$.

• The principal's adjustment decision:

$$\lambda(\mu(m)) = 0 \Leftrightarrow \hat{\rho}(m) \geq \frac{1}{2}, \sigma(\mu(m)) = 0 \Leftrightarrow \hat{p}(m) \geq p$$

(2) When the the available prior knowledge is enough, i.e., q is large enough, we will have the following:

• The principal will read the initial signal and select the project $\rho = \rho(\tau) = \tau$ follow the informative signal τ.

• The agent will choose a report π as follows:

$$\pi(m = \text{Bad News}|s) = \begin{cases} 1 & , s \in [\underline{\omega}, s^*(q)) \\ 0 & , s \in [s^*(q), \bar{\omega}] \end{cases} \quad \pi(m = \text{Good News}|s) = \begin{cases} 0 & , s \in [\underline{\omega}, s^*(q)) \\ 1 & , s \in [s^*(q), \bar{\omega}] \end{cases}$$

where the signal space partition is $s(2) = [\underline{\omega}, s^*(q)) \cup [s^*(q), \bar{\omega}]$.

• The principal's adjustment decision

$$\lambda(\mu(m)) = 0 \Leftrightarrow \hat{\rho}(m) \geq \frac{1}{2}, \sigma(\mu(m)) = 0 \Leftrightarrow \hat{p}(m) \geq p$$

6 Extension

6.1 Optimal information acquisition effort

In the baseline model, the accuracy of the prior signal is exogenous. We can allow the accuracy of this signal to be endogenously determined by the principal's effort, where $e \in [0, \bar{e}]$ and $\bar{e}$ can be any positive real number, even $+\infty$, $q(0) = \frac{1}{2}, \lim_{e \to \bar{e}} q(e) = 1, q^{'}(\cdot) > 0, q^{''}(\cdot) < 0$, and the corresponding information acquisition cost is $c(e)$, where $c(0) = 0, c^{'}(\cdot) > 0, c^{''}(\cdot) \geq 0$.

Denote $\hat{e}$ as the solution to

$$\frac{g(0)}{f_{-}(0)} = \frac{(1 - q(e))(p + (1-p)\epsilon)}{(2q(e) - 1)(1-p)(1-\epsilon)}$$

Then, the following proposition provides a sufficient condition to characterize the principal's optimal information acquisition effort level.

Proposition 7 (Optimal information acquisition effort)**.** The principal's optimal information acquisition effort level is:

(1) When $c(\hat{e}) > \frac{1}{2}(1-p)(1-\epsilon)(1+\kappa)$. The principal's optimal effort is then

$$e^* = 0$$

(2) When $\lim_{e \to \bar{e}} c(e) < \frac{1}{2}(1-p)(1-\epsilon)(1+\kappa)$, then the principal's optimal effort is strictly positive

$$e^* > \hat{e} > 0$$

Proof. From the analysis in proposition 5, we know when $e \leq \hat{e}$, the expected project outcome is worse than the expected project outcome, where $e = 0$.

If the principal makes an effort to collect prior information, the expected project outcome is improved by at most

$$W(1, \epsilon, \kappa) - W(\frac{1}{2}, \epsilon, \kappa) = [p + (1-p)\kappa] - [p + (1-p)\epsilon\kappa - \frac{1}{2}(1-p)(1-\epsilon)(1-\kappa)]$$

$$= \frac{1}{2}(1-p)(1-\epsilon)(1+\kappa)$$

while the induced cost is at least $c(\hat{e})$.

Thus, when $c(\hat{e}) > \frac{1}{2}(1-p)(1-\epsilon)(1+\kappa)$, the benefit from the positive effort can never cover the corresponding cost, so the optimal effort of the principal is to choose

$$e^* = 0$$

However, when $\lim_{e\to\infty} c(e) < \frac{1}{2}(1-p)(1-\epsilon)(1+\kappa)$, which means that the benefit generated from the positive effort level might cover the effort cost, then exerting a high enough effort can generate a better result. Then, the optimal effort is strictly positive. Furthermore, if $e \leq \hat{e}$ is insufficient, then $e > \hat{e}$.

□

6.2 Multiple agents

Would having multiple agents improve the efficiency of feedback ? We consider the simplest case with two agents. If the principal hires two agents to independently manage the same project experiment, will the principal obtain more informative reports than in the single-agent case?

As the two agents could obtain more feedback, signals s_1 and s_2 independently, ideally, if two agents were to truthfully disclose their own feedback signals, more information would be available to help adjust the project.

However, when the two agents simultaneously submit their reports, it is possible for the principal to compare s_1 and s_2 to separately judge the two agents' competence. Even when the principal initially remains naive, the reports $r_1(s_1)$ and $r_2(s_2)$ allow the principal to use one report to check the other and then judge the agents' competence.

In particular, when the two agents' reports are inconsistent, for example, when $r_1(s_1)$ suggests that the principal should maintain the project and $r_2(s_2)$ suggests that the principal modify the project, from the principal's perspective, it is more likely that at least one agent is incompetent. This is the so-called *double check effect*. This effect emerges when many agents submit reports simultaneously and makes it less likely that an agent will report an informative signal.

Example 1. Following the numerical example in section 5.2, suppose that the principal has already chosen to initially be naive, i.e., $q = 0.5$; then, we have the following:

(1) If there is only one agent, following the *Good-bad news* reporting choice, he will report good news when he receives signal $s > 0$ and report bad news when he receives signal $s < 0$ in equilibrium.

(2) If there are two agents, suppose that both of them follow the *Good-bad news* reporting choice: Agent i will report good news iff $s_i > 0, i = 1, 2$, and when $r_1(s_1) = \mathrm{G}, r_2(s_2) = \mathrm{B}$, the principal will infer the two agents' abilities based on both of their reports $(r_1(s_1), r_2(s_2))$:

$$P(1 = H | r_1 = \mathrm{G}, r_2 = \mathrm{B}) = \frac{\frac{p}{2}q(1-p)}{\frac{p}{2}q(1-p) + \frac{p}{2}(1-q)(1-p) + \frac{1}{4}(1-p)^2} = \frac{p}{p + \frac{1-q}{q} + \frac{1-p}{2q}} = \frac{p}{2} < p$$

$$P(2 = H | r_1 = \mathrm{G}, r_2 = \mathrm{B}) = \frac{\frac{p}{2}(1-q)(1-p)}{\frac{p}{2}q(1-p) + \frac{p}{2}(1-q)(1-p) + \frac{1}{4}(1-p)^2} = \frac{p}{p + \frac{q}{1-q} + \frac{1-p}{2(1-q)}} = \frac{p}{2} < p$$

$$P(1=H|r_1=\mathrm{G},r_2=\mathrm{G})=P(2=H|r_1=\mathrm{G},r_2=\mathrm{G})=\frac{\frac{1}{2}p[p+\frac{1}{2}(1-p)]}{\frac{1}{2}p[p+\frac{1}{2}(1-p)]+\frac{1}{4}(1-p)[p+\frac{1}{2}(1-p)]+\frac{1}{8}(1-p)^2}$$

$$P(1=H|r_1=\mathrm{B},r_2=\mathrm{B})=P(2=H|r_1=\mathrm{B},r_2=\mathrm{B})=\frac{\frac{1}{2}p[p+\frac{1}{2}(1-p)]}{\frac{1}{2}p[p+\frac{1}{2}(1-p)]+\frac{1}{4}(1-p)[p+\frac{1}{2}(1-p)]+\frac{1}{8}(1-p)^2}$$

In this situation, both agents 1 and 2 will be judged as more likely to be incompetent when their reports are inconsistent, and they will be judged as more likely to be competent only when their reports are consistent. Thus, if both agent 1 and agent 2 follow the single agent's equilibrium reporting choice based on their received signals, they will find that they always face a positive probability of being replaced. However, if agent 1 follows the single agent's equilibrium reporting choice and if agent 2 always babbles, then the principal will not update her belief about agent 2, and agent 2 can avoid being replaced with probability 1.

The above example demonstrates that the single agent's equilibrium reporting choice will not be applied by two agents simultaneously. More generally, the following proposition 8 shows that in the uniform example, there is no equilibrium in which the two agents apply the same informative reporting choice.

Proposition 8. In the uniform example, provided that the principal is initially naive, if there are two agents, except for the babbling equilibrium, a symmetric equilibrium in which the two agents use the same reporting strategy does not exist.

Proof. Suppose that both of them follow the same *Good-bad news* reporting choice: Agent i will report good news iff $s_i > s_*, i = 1, 2$, and when $r_1(s_1) = \mathrm{G}, r_2(s_2) = \mathrm{B}$, the principal will infer the two agents' abilities based on both of their reports $(r_1(s_1), r_2(s_2))$:

When $-1 < s_* < 0$,

$$P(2=H|r_1(s_1)=\mathrm{G},r_2(s_2)=\mathrm{B})=\frac{\frac{(2+s_*)p}{2}\left[\frac{(1-s_*)(1-p)}{2}-\frac{s_*p}{2}\right]}{\frac{(2+s_*)p}{2}\left[\frac{(1-s_*)(1-p)}{2}-\frac{s_*p}{2}\right]+\frac{(1-s_*)(1-p)}{2}\left[\frac{(1-s_*)(1-p)}{2}+p(\frac{1}{2}-\frac{s_*}{2})\right]}$$

Note that this expression being less than p is equivalent to

$$(1-s_*)^2>(2+s_*)(1-s_*)(1-p)-(2+s_*)s_*p \Leftrightarrow 1-2p-2s_*^2+(1-p)s_*<0$$

and this inequality always holds when $-1 < s_* < 0$.

When $0 < s_* < 1$,

$$P(1=H|r_1(s_1)=\mathrm{G},r_2(s_2)=\mathrm{B})=\frac{\frac{(2-s_*)p}{2}\left[\frac{(1+s_*)(1-p)}{2}+\frac{s_*p}{2}\right]}{\frac{(2-s_*)p}{2}\left[\frac{(1+s_*)(1-p)}{2}+\frac{s_*p}{2}\right]+\frac{(1+s_*)(1-p)}{2}\left[\frac{(1+s_*)(1-p)}{2}+p(\frac{1}{2}+\frac{s_*}{2})\right]}$$

Note that this expression being less than p is equivalent to

$$(1+s_*)^2>(2-s_*)(1+s_*)(1-p)+(2-s_*)s_*p \Leftrightarrow 2s_*^2+2p-1+(1-p)s_*>0$$

and this inequality always holds when $0 < s_* < 1$.

When $s_* \geq 1$, it is trivial that when the principal receives $r_1 = \mathrm{G}, r_2 = \mathrm{B}$, he will judge agent 2 to be the incompetent type for certain.

When $s_* \leq -1$, it is trivial that when the principal receives $r_1 = \mathrm{G}, r_2 = \mathrm{B}$, he will judge agent 1 to be the incompetent type for certain.

In this example, if the two agents follow the same reporting strategy, although the principal is naive, when she receives two reports with the opposite recommendations, she will judge at least one of the agents to be more likely to be incompetent, and the probability that this agent is not replaced is less than 1. □

In this example, although an equilibrium in which the two agents apply the same informative reporting strategy does not exist, there is a babbling equilibrium in which the two agents apply the same reporting strategy. Furthermore, at a minimum, a trivially informative equilibrium in which the two agents apply the asymmetric reporting strategy exists: One agent chooses to remain silent regardless of the signal he receives, and the other agent chooses the *Good-Bad news* report as described in proposition 2.

Proposition 8 shows that in general, due to the existence of the double-check effect, it may not be helpful for the principal to hire more than one agent to collect the information and submit a report. The double-check effect will offset the additional informational advantage from having more agents collecting information. This finding is consistency with Krishna and Morgan (2001) which argues that the principal can not be beneficial to consult two agents with same direction of interest. In the environment of this paper, both of the two agents have the common interest to keep the position to avoid being kicked out, it is difficult for the principal to obtain more informative report via multiple agents.

6.3 Delegation of power

Will delegating power improve efficiency? When the principal has the full authority, on the one hand, the agent has to submit a report to the principal, and because of the principal's judgement, the agent will disclose information with distortion. On the other hand, the principal can encourage the agent to exert maximal effort to guarantee efficient implementation. What would be the result of delegating the authority to an agent who has sufficient feedback information?

Assumption 3.

$$\eta - \alpha > 0$$

Assumption 3 means that the cost η is high enough and the agent will exert effort $e_a = 0$ when he faces the realized cost η and can choose the effort level himself. Provided that assumption 3 holds, when the principal delegates the authority to the agent, the agent can immediately adjust the project based on the feedback he receives, but the potential for loss in this case comes from the fact that the agent may shirk when he faces a significant effort cost in the implementation process. For the principal, the question of whether to delegate represents a trade-off between information distortion and the loss of control, which is similar in spirit to Aghion and Tirole (1997) and Dessein (2002).

Thus, when the authority is delegated to the agent, with probability ξ, the agent will shirk. In general, centralizing authority in the principal offers an implementation advantage relative to delegating authority to the agent, while delegating authority to the agent has an informational

advantage over centralized authority. The following proposition 9 identifies the condition under which centralization (no delegation) dominates delegation.

Proposition 9. Comparison between centralization and delegation

(1) $\exists \underline{q}$, when $q < \underline{q}$, for $\forall \xi > 0$, there will be

$$W_{\text{Centralization}} > W_{\text{Decentralization}}$$

(2) $\exists \bar{q}$, when $q > \bar{q}$, for $\forall \xi > 0$, there will be

$$W_{\text{Centralization}} > W_{\text{Decentralization}}$$

The intuition for proposition 9 is as follows:

(1) When q is low, the information loss under centralization is small, while the implementation efficiency under centralized authority dominates that under delegated authority.

(2) When q is high enough, the informational advantage under decentralization is not significant, and the implementation efficiency under centralized authority dominates that under delegated authority.

Furthermore, the only condition under which delegated authority could dominate centralized authority is when q is at an intermediate level. In this situation, the information distortion under centralized authority is large and the informational advantage under delegated authority may dominate the implementation efficiency under centralized authority. The following part provides a numerical example to compare centralized authority and delegated authority in the uniform case.

Example 2. In above uniform case, we can calculate that if the principal delegates all power to the agent, then the expected project outcome will be

$$W_D(q) = \begin{cases} (1-\xi)p & \text{if } q \in (\frac{1}{2}, \frac{1}{2-p}) \\ (1-\xi)[q-(1-q)(1-p)] & \text{if } q \in (\frac{1}{2-p}, 1) \end{cases}$$

Under centralized authority, as in the previous calculation, the expected project outcome is

$$W_E(q) = 2(1-q)p + \frac{(2q-1)^2}{q}$$

It is obvious that centralized authority is optimal when $q < \frac{1}{2-p}$; then, when $q > \frac{1}{2-p}$

$$\Delta W = W_E(q) - W_D(q) = (1+\xi)(2-p)q + \frac{1}{q} - (2-p)(1+\xi) - (1-\xi)$$

Then,

$$\frac{d\Delta W}{dq} = -2p + 4 - \frac{1}{q^2} - (1-\xi)(2-p) = 0 \Rightarrow q = \frac{1}{\sqrt{(1+\xi)(2-p)}}$$

While $q > \frac{1}{2-p}$ requires that $\xi + p < 1$; then, the minimum value of ΔW is achieved when

$$q = \frac{1}{\sqrt{(1+\xi)(2-p)}}$$

The minimum value is:

$$2\sqrt{(1+\xi)(2-p)} - (2-p)(1+\xi) - (1-\xi) < 0 \Leftrightarrow p < 1 - \frac{2\sqrt{\xi}}{1+\xi}$$

When $p < 1 - \frac{2\sqrt{\xi}}{1+\xi}$, denote the two solutions of

$$(1+\xi)(2-p)q^2 - [(2-p)(1+\xi) - (1-\xi)]q + 1 = 0$$

as q_1, q_2.

In summary, we have the following:

(1) When $\frac{1}{2} \leq q \leq q_1$, centralized authority is better than delegated authority.

(2) When $q_1 < q < q_2$, delegated authority is better than centralized authority.

(3) When $q_2 < q < 1$, centralized authority is better than delegated authority.

The following figure 5 depicts the above results.

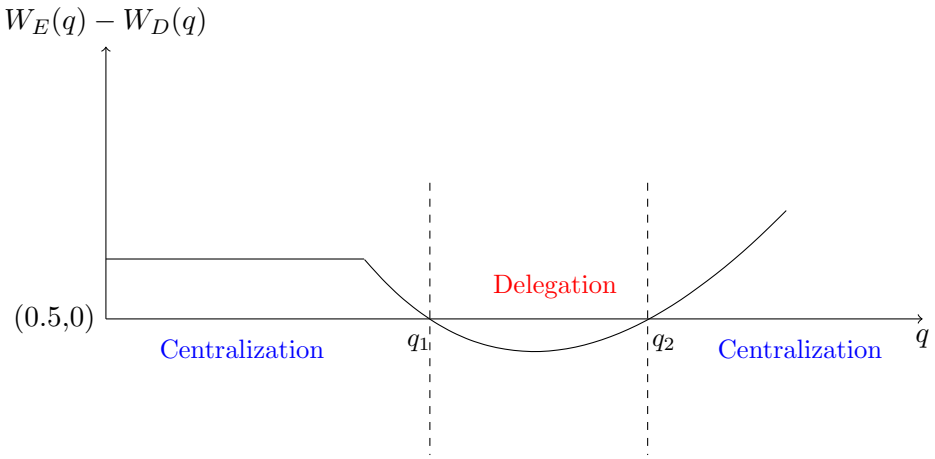

Figure 5: Optimal power allocation in the uniform example, $p = 0.55, \xi = 0.04$

7 Conclusion

This paper shows that because the agent anticipated that the principal will infer his competence based on the information he discloses, he has an incentive to manipulate his report to minimize his likelihood of being fired. In this case, the principal making the rational choice to remain ignorant could encourage the agent to disclose more informative information.

Note that the some conditions are important for the conclusions of this paper:

First, similar to Crawford and Sobel (1982), in this paper, the feedback signal is soft and difficult to verify. Thus, for the principal, other than the prior information about the project, her source of information is the message delivered by the agent based on the feedback signal. If the principal had effective tools to verify the agent's report, the story would be different.

Second, the principal has no commitment power. If the principal could commit to never replacing the agent regardless of the agent's report, then the agent would disclose the full feedback signal.

The agent will worry about the principal's judgement and thus have an incentive to manipulate his report only when the principal cannot commit to maintaining the agent's position.

Third, the agent's primary concern of retaining his position dominates his other concerns; his desire to retain his position may come from the benefits from implementing the project (wage, bonus or other incentive schedule offered by the principal) or the benefit of avoiding the cost of losing his job and searching for a new one. This condition guarantees that the agent's reporting choice is always located in the safe report set. If the agent cared more about the project outcome and always told the whole truth about the feedback result, there would be no distortion.

In this paper, although the principal and agent still share a common interest in the project as both of them will be better off if the project is successful, there is still information transimission distortion driven by the interaction between the principal and the agent, as the principal always attempts to judge the agent's competence agent through the latter's message and replaces an incompetent agent with a more competent one to improve project implementation efficiency. The agent always attempts to retain his position by manipulating information. The co-existence of some common interests and some conflicts of interest is also the principal-agent relationship feature in this paper.

Consistent with classical principal-agent literature, in an environment with asymmetric information, from the agent's perspective, an agent with an informational advantage can obtain information rent through strategic disclosure; in particular, an incompetent agent could also retain his position. However, in the classical principal-agent literature, from the principal's perspective, the principal's informational disadvantage under asymmetric information between the principal and agent always makes the principal suffer a loss in the form of the agent's information rent; the significant departure of this paper is to demonstrate that the principal might be better off to choosing to know less information and thereby induce the agent to make a more informative information disclosure when a project adjustment based on efficient information disclosure is sufficiently important for the final outcome in a bounded rationality environment.

References

Abrahamson, Eric and Choelsoon Park, "Concealment of negative organizational outcomes: An agency theory perspective," *Academy of management journal*, 1994, *37* (5), 1302–1334.

Aghion, Philippe and Jean Tirole, "Formal and real authority in organizations," *Journal of political economy*, 1997, *105* (1), 1–29.

Bolton, Patrick, Markus K Brunnermeier, and Laura Veldkamp, "Leadership, coordination, and corporate culture," *Review of Economic Studies*, 2012, *80* (2), 512–537.

Brocas, Isabelle and Juan D Carrillo, "Influence through ignorance," *The RAND Journal of Economics*, 2007, *38* (4), 931–947.

Crawford, Vincent P and Joel Sobel, "Strategic information transmission," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1431–1451.

Dessein, Wouter, "Authority and communication in organizations," *The Review of Economic Studies*, 2002, *69* (4), 811–838.

_ , Andrea Galeotti, and Tano Santos, "Rational inattention and organizational focus," *American Economic Review*, 2016, *106* (6), 1522–36.

Fan, Ziying, Wei Xiong, and Li-An Zhou, "Information Distortion in Hierarchical Organizations: A Study of China?s Great Famine," Technical Report, Working Paper 2016.

Garicano, Luis and Luis Rayo, "Why organizations fail: models and cases," *Journal of Economic Literature*, 2016, *54* (1), 137–92.

Grossman, Sanford J and Oliver D Hart, "An analysis of the principal-agent problem," *Econometrica: Journal of the Econometric Society*, 1983, pp. 7–45.

Hertwig, Ralph and Christoph Engel, "Homo ignorans: Deliberately choosing not to know," *Perspectives on Psychological Science*, 2016, *11* (3), 359–372.

Hölmstrom, Bengt, "Moral hazard and observability," *The Bell journal of economics*, 1979, pp. 74–91.

Holmstrom, Bengt and Paul Milgrom, "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design," *JL Econ. & Org.*, 1991, *7*, 24.

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

Kessler, Anke S, "The value of ignorance," *The Rand Journal of Economics*, 1998, pp. 339–354.

Kluger, Avraham N and Angelo DeNisi, "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory.," *Psychological bulletin*, 1996, *119* (2), 254.

Krishna, Vijay and John Morgan, "A model of expertise," *The Quarterly Journal of Economics*, 2001, *116* (2), 747–775.

Kung, James Kai-Sing and Shuo Chen, "The tragedy of the nomenklatura: Career incentives and political radicalism during China's Great Leap famine," *American Political Science Review*, 2011, *105* (1), 27–45.

Lewis, Tracy R and David EM Sappington, "Ignorance in agency problems," *Journal of Economic Theory*, 1993, *61* (1), 169–183.

McGoey, Linsey, "The logic of strategic ignorance," *The British Journal of Sociology*, 2012, *63* (3), 533–576.

Meng, Xin, Nancy Qian, and Pierre Yared, "The institutional causes of China's great famine, 1959–1961," *The Review of Economic Studies*, 2015, *82* (4), 1568–1611.

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

Prendergast, Canice, "A theory of *yes men*," *The American Economic Review*, 1993, pp. 757–770.

Roberts, Joanne, "Organizational ignorance: Towards a managerial perspective on the unknown," *Management Learning*, 2013, *44* (3), 215–236.

Smithson, Michael, "Ignorance and science: Dilemmas, perspectives, and prospects," *Knowledge*, 1993, *15* (2), 133–156.

Verrecchia, Robert E, "Discretionary disclosure," *Journal of accounting and economics*, 1983, *5*, 179–194.

Appendices

A Omitted Proof

Proof of proposition 1

We will complete the proof of proposition 1 through the proof of the following two lemmas 1 and 2.

Lemma 1. For $\forall$ fixed $\epsilon > 0$, when the agent receives $s > 0$

(1) The agent believes that policy is more likely to be correct, i.e., $P(\theta|s) > \frac{1}{2}$

(2) If the agent reports $s > 0$ truthfully, the principal's inference about the agent's ability $\hat{p}(s) \geq p$ iff $s \geq \bar{s}(q) > 0$.

Proof. Note that for $\forall$ fixed $e \geq 0$, $q \geq \frac{1}{2}$, if the agent receives signal $s > 0$, his belief about $\theta = 1$ is:

$$P(\theta = 1|s) = \frac{q[(p + (1-p)\epsilon)f_+(s) + (1-p)(1-\epsilon)g(s)]}{q(p + (1-p)(1-\epsilon))f_+(s) + (1-p)(1-\epsilon)g(s)} \geq \frac{1}{2}$$

So the agent will believe that the policy is more likely to be corrected when he receives signal $s > 0$.

Furthermore, when the agent reports signal $s > 0$ truthfully, the principal will judge the agent's ability as following:

$$P(H|s) = \frac{qpf_+(s)}{qpf_+(s) + (1-p)[q\epsilon f_+(s) + (1-\epsilon)g(s)]}$$

The principal will judge the agent's ability higher than average level only when

$$qf_+(s) \geq q\epsilon f_+(s) + (1-\epsilon)g(s) \Leftrightarrow \frac{g(s)}{f_+(s)} \leq q$$

In particularly, the threshold signal s such that $P(H|s) = p$ holds is denoted as $\bar{s}(q)$.Given assumption 1, $\bar{s}(q) > 0$. □

The lemma 1 shows: when the agent receives a positive signal s and discloses it to the principal truthfully, the principal will always think that the project is more likely to be correct, driven by the assumption 1, this feedback maybe correct informative signal or just a noise, she will judge that the agent's ability is higher than the average level of the agent pool only when the positive signal is strong enough as the competent agent is more likely to receive a strong feedback.

Lemma 2. When the agent receives signal $s < 0$. If the agent always reports $s < 0$ truthfully, the principal will believe that the agent's ability is higher than the average level iff $s < \underline{s}(q) < 0$.

Proof. If the agent receives signal $s < 0$, his belief about $\theta = 1$ is:

$$P(\theta = 1|s) = \frac{q(1-p)(1-\epsilon)g(s)}{(1-q)pf_-(s) + (1-p)(\epsilon(1-q)f_-(s) + (1-\epsilon)g(s))}$$

The agent will believe that the policy is more likely to be wrong if

$$(1-q)[p + \epsilon(1-p)]f_-(s) > (2q-1)(1-p)(1-\epsilon)g(s) \Leftrightarrow \frac{g(s)}{f_-(s)} < \frac{(1-q)[p + \epsilon(1-p)]}{(2q-1)(1-p)(1-\epsilon)}$$

There are two possibities:

If $\frac{g(0)}{f_-(0)} < \frac{(1-q)[p+\epsilon(1-p)]}{(2q-1)(1-p)(1-\epsilon)}$, the agent who receives signal $s < 0$ will always thinks that the policy 1 is more likely to be wrong.

If $\frac{g(0)}{f_-(0)} > \frac{(1-q)[p+\epsilon(1-p)]}{(2q-1)(1-p)(1-\epsilon)}$, then $\exists \tilde{s}(q) \in (\underline{s}(q), 0)$ ($\underline{s}(q)$ will be defined in the following) such that the agent believes the policy 1 is more likely to be correct when $s \in (\tilde{s}(q), 0)$ and believes that the policy 1 is more likely to be wrong when $s < \tilde{s}(q)$.

Furthermore, when the agent reports signal $s < 0$ truthfully, the principal will judge the agent's ability as following:

$$P(H|s) = \frac{(1-q)pf_-(s)}{(1-q)pf_-(s) + (1-p)[(1-q)\epsilon f_-(s) + (1-\epsilon)g(s)]}$$

The principal will judge that the agent's ability is lower than the average level if

$$(1-q)f_-(s) < (1-q)\epsilon f_-(s) + (1-\epsilon)g(s) \Leftrightarrow \frac{g(s)}{f_-(s)} > 1-q$$

In particularly, the threshold signal s such that $P(H|s) = p$ is denoted as $\underline{s}(q)$. □

The lemma 2 shows: when the agent receives a negative signal $s < 0$ and disclose the signal truthfully, the principal will judge the agent's ability is not less than the average level in the agent pool if and only if the s seems to be negative enough, as the competent agent is more likely to receive a strong feedback when the project is bad.

Based on analysis of above two lemmas: When the agent receives signal $s \in (\underline{s}(q), \bar{s}(q))$ and disclose s truthfully, he will be judged to be less than the average competence level of the outside agent pool and be replaced by the principal. When the agent receives signal $s \geq \bar{s}(q)$ or $s \leq \underline{s}(q)$ and disclose it truthfully, he will be judged as to be more competent and continue to handle the project. Then the agent who receives signal $s \in (\underline{s}(q), \bar{s}(q))$ has incentive to deviate the truthful report claim that he recieves a signal larger than $\bar{s}(q)$ or less than $\underline{s}(q)$. Then the truthful report can not be the agent's best response given the agent's optimal adjustment decision rule.

Proof of proposition 2

Proof. First part, let's check the agent's Good-bad news disclosure is indeed a best response to the principal's adjustment rule.

Suppose the agent follows a report strategy: Report bad news when $s < s^*(q)$ and report good news when $s > s^*(q)$.

If the cut-off value $s^*(q) < 0$, then if the principal receives signal "Bad news", he will infer the agent's ability as the following.

$$P(H|B) = \frac{(1-q)pF_-(s^*(q))}{(1-q)pF_-(s^*(q)) + (1-p)[\epsilon(1-q)F_-(s^*(q)) + (1-\epsilon)G(s^*(q))} \geq p$$

which is equivalent to

$$(1-q)F_-(s^*(q)) \geq \epsilon(1-q)F_-(s^*(q)) + (1-\epsilon)G(s^*(q)) \tag{3}$$

If the principal receives signal "Good news", he will infer the agent's ability as the following:

$$P(H|G) = \frac{p[(1-q)(1-F_-(s^*(q))) + q]}{(p+(1-p)\epsilon)[(1-q)(1-F_-(s^*(q))) + q] + (1-p)(1-\epsilon)(1-G(s^*(q))]} \geq p$$

which is equivalent to

$$(1-q)(1-F_-(s^*(q))) + q \geq \epsilon[(1-q)(1-F_-(s^*(q))) + q] + (1-\epsilon)(1-G(s^*(q)))$$

After simple algebraic, it becomes:

$$(1-q)F_-(s^*(q)) \leq G(s^*(q)) \tag{4}$$

Combining the inequality (3) and (4), then we obtain that:

$$(1-q)F_-(s^*(q)) = G(s^*(q)) \tag{5}$$

Note that:

$$G(0) = \frac{1}{2}, F_-(0) = 1, \frac{G(0)}{F_-(0)} = \frac{1}{2} > 1-q$$

Combined with the assumption 1, monotone likelihood ratio gurantees that $\frac{G(s)}{F_-(s)}$ is increasing in s, so the equation 5 has unique solution $s^*(q)$.

So given the agent's report strategy, the principal will always judge the agent's ability not less than the average level in the agent pool and will not replace current agent.

Second part, we will show the principal's belief about the agent's ability and project quality after receiving agent's report given agent's report strategy, if the leader receives a good or bad report, his inference about the true state will be:

$$P(\theta = 1|G) = \frac{q[(p+(1-p)\epsilon) + (1-p)(1-\epsilon)(1-G(s^*(q))]}{q(p+(1-p)\epsilon) + (1-p)(1-\epsilon)(1-G(s^*(q))} > \frac{1}{2}$$
$$P(\theta = 1|B) = \frac{q(1-p)(1-\epsilon)G(s^*(q))}{(1-p)(1-\epsilon)G(s^*(q)) + (1-q)(p+(1-p)\epsilon)F_-(s^*(q))} < \frac{1}{2}$$

always hold for any $p, q > \frac{1}{2}$. So the principal's best response is: Sticking to initial policy when receiving "Good News" and modifying initial policy when receiving "Bad News". □

Proof of proposition 3

Proof. First part, in order to show $s^*(q) < \underline{s}(q)$, we can compare

$$\frac{g(\underline{s}(q))}{f_-(\underline{s}(q))} = 1-q, \frac{G(s^*(q))}{F_-(s^*(q))} = 1-q$$

Let

$$R(s) = \frac{G(s)}{F_-(s)}$$

Then:

$$\frac{dR(s)}{ds} = \frac{F_-(s)g(s) - G(s)f_-(s)}{F_-^2(s)} = \frac{F_-(s)f_-(s)[\frac{g(s)}{f_-(s)} - \frac{G(s)}{F_-(s)}]}{F_-^2(s)} \tag{6}$$

Based on the property of the monotone likelihood ratio property, given assumption (1), we can obtain that:

$$\frac{g(s)}{f_-(s)} \geq \frac{G(s)}{F_-(s)}$$

Then, back to the differention

$$\frac{dR(s)}{ds} \geq 0$$

Furthermore

$$R(s^*(q)) = \frac{G(s^*(q))}{F_-(s^*(q))} = 1 - q = \frac{g(\underline{s}(q))}{f_-(\underline{s}(q))} \geq \frac{G(\underline{s}(q))}{F_-(\underline{s}(q))} = R(\underline{s}(q))$$

As $R(s)$ is increasing in s, so $s^*(q) \geq \underline{s}(q)$.

In order to show the second part, recall the equation 5

$$(1-q)F_-(s^*(q)) = G(s^*(q))$$

Based on the implicit function theorem, we can obtain that:

$$\frac{ds^*(q)}{de} = \frac{ds^*(q)}{dq} = \frac{F_-(s^*(q))}{(1-q)f_-(s^*(q)) - g(s^*(q))}$$

From above part, we know that

$$s^*(q) > \underline{s}(q)$$

Combined with assumption 1, we know:

$$\frac{g(s^*(q))}{f_-(s^*(q))} \geq \frac{g(\underline{s}(q))}{f_-(\underline{s}(q))} = 1 - q \Rightarrow g(s^*(q)) \geq (1-q)f_-(s^*(q))$$

Then

$$\frac{ds^*(q)}{dq} < 0$$

is shown. □

Proof of proposition 4

Proof. In general, the whole signal space can be partitioned as the following $\Omega = \cup_{k=0}^{n-1}[s_k, s_{k+1}]$ where $s_0 = \underline{\omega}$ and $s_n = \bar{\omega}$ and the agent can submit different reports in different signal interval.

Case 1: $\exists m$ such that $s_m = 0$

In this case, there is no signal interval with different signs. If this signal partition can support an equilibrium, then:

$$P(H|[s_k, s_{k+1}])$$
$$= \frac{(1-q)(F_-(s_{k+1}) - F_-(s_k))p}{(1-q)(F_-(s_{k+1}) - F_-(s_k))p + [\epsilon(1-q)(F_-(s_{k+1}) - F_-(s_k)) + (1-\epsilon)(G(s_{k+1}) - G(s_k))](1-p)}$$

$\geq p$

holds for $0 \leq k \leq m-1$.

It is equivalent to

$$(1-q)(F_-(s_{k+1}) - F_-(s_k)) \geq G(s_{k+1}) - G(s_k)$$

Sum the inequalities from $k=0$ to $k=m-1$, there will be:

$$1 - q \geq \frac{1}{2}$$

which contradicts with $q > \frac{1}{2}$. So there will be at least one k such that

$$(1-q)(F_-(s_{k+1}) - F_-(s_k)) < G(s_{k+1}) - G(s_k)$$

Then the agent's survival probability will be less than 1 in this information partition.

Case 2: $s_m \neq 0$ **for** $m = 1 \cdots n$

In this case, there exists m such that $s_m < 0$ and $s_{m+1} > 0$.

When $0 \leq k \leq m-1$, then

$$P(H|[s_k, s_{k+1}])$$
$$= \frac{(1-q)(F_-(s_{k+1}) - F_-(s_k))p}{(1-q)(F_-(s_{k+1}) - F_-(s_k))p + [(1-q)\epsilon(F_-(s_{k+1}) - F_-(s_k)) + (1-\epsilon)(G(s_{k+1}) - G(s_k))](1-p)} \geq p$$

holds for $1 \leq k \leq m-1$ which is equivalent to

$$(1-q)(F_-(s_{k+1}) - F_-(s_k)) \geq G(s_{k+1}) - G(s_k)$$

Sum from $k=0$ to $k=m-1$, there will be

$$(1-q)F_-(s_m) \geq G(s_m) \tag{7}$$

When $k=m$, then

$$P(H|[s_k, s_{k+1}])$$
$$= \frac{[(1-q)(1-F_-(s_k)) + qF_+(s_{k+1})]p}{[(1-q)(1-F_-(s_k)) + qF_+(s_{k+1})](p + (1-p)\epsilon) + (1-\epsilon)(G(s_{k+1}) - G(s_k))(1-p)} \geq p$$

which is equivalent to

$$(1-q)(1-F_-(s_m)) + qF_+(s_{m+1}) \geq G(s_{m+1}) - G(s_m) \tag{8}$$

When $m+1 \leq k \leq n-1$, then

$$P(H|[s_k, s_{k+1}])$$
$$= \frac{q(F_+(s_{k+1}) - F_+(s_k))p}{(1-q)(F_+(s_{k+1}) - F_+(s_k))p + [\epsilon q(F_+(s_{k+1}) - F_+(s_k)) + (1-\epsilon)(G(s_{k+1}) - G(s_k))](1-p)} \geq p$$

holds for $m+1 \leq k \leq n$ which is equivalent to

$$q(F_+(s_{k+1}) - F_+(s_k)) \geq G(s_{k+1}) - G(s_k)$$

Sum from $k = m+1$ to $k = n-1$, there will be

$$q(1 - F_+(s_{m+1})) \geq 1 - G(s_{m+1}) \tag{9}$$

If either inequality (7) or (9) strictly holds, there will be:

$$(1-q)F_-(s_m)+q(1-F_+(s_{m+1})) > 1-G(s_{m+1})+G(s_m) \Leftrightarrow (1-q)(1-F_-(s_m))+qF_+(s_{m+1}) < G(s_{m+1})-G(s_m)$$

which is contradicts with inequality (8).

So the inequality (7) and (9) should keep equality strictly and then the inequality (8) also keep the equality strictly.

The good-bad signal partial informative equilibrium satisfies above constraint

$$s_m = s^*(q), s_{m+1} = 1$$

Step 3: Other equilibrium will not be more informative

In this step, we will show that other equilibrium if exists can not provide more information than the good-bad signal equilibrium. Note given above reputation constraint, if the agent recieves signal $s \in [s_k, s_{k+1}], 1 \leq k \leq m-1$, then the principal will infer the project quality as:

$$\begin{aligned} P(\theta = 1|[s_k, s_{k+1}]) &= \frac{q(1-p)(1-\epsilon)(G(s_{k+1}) - G(s_k))}{(p+(1-p)\epsilon)(F_-(s_{k+1}) - F_-(s_k)) + (1-p)(1-\epsilon)(G(s_{k+1}) - G(s_k))} \\ &\leq \frac{q(1-p)(1-\epsilon)}{\frac{p+(1-p)\epsilon}{1-q} + (1-p)(1-\epsilon)} = \frac{1}{1 + \frac{(p+(1-p)\epsilon)q}{(1-p)(1-\epsilon)(1-q)}} < \frac{1}{2} \end{aligned}$$

So the principal will always modify the initial project when he recieves signal $s \in [s_k, s_{k+1}], 0 \leq k \leq m-1$.

If the agent recieves signal $s \in [s_m, s_{m+1}]$, then the principal will infer the project quality as:

$$\begin{aligned} &P(\theta = 1|s \in [s_m, s_{m+1}]) \\ &= \frac{q[(p+(1-p)\epsilon)F_+(s_{m+1}) + (1-p)(G(s_{m+1}) - G(s_m))]}{(p+(1-p)\epsilon)(F_+(s_{m+1}) - F_-(s_m)) + (1-p)(1-\epsilon)(G(s_{m+1}) - G(s_m))} > \frac{1}{2} \end{aligned}$$

So the principal will always stick to the initial project when he recieves signal $s \in [s_m, s_{m+1}]$.

When the agent recieves signal $s \in [s_k, s_{k+1}], m+1 \leq k \leq n-1$, as $s_k > 0$ in these intervals, then the principal will always stick to the initial selected project.

Eventually, given the reputation restriction, if other partition equilibria exist, no other equilibrium provide more information than the good-bad binary equilibrium. So the good-bad signal partial informative equilibrium is the most informative equilibrium. □

Proof of proposition 5

Proof. We will show the three results one by one.

First step, we can calculate the expected outcome under above *Good-Bad News equilibrium* $W(q, \epsilon, \kappa)$.

As the benchmark, when there is no experimental step, the expected project outcome will be:

$$q[p + (1-p)\kappa] - (1-q)$$

Then let's consider the calculation in the *Good-Bad news equilibrium*: If the initial project is correct, the correct project will continue with probability $p + (1-p)\epsilon + (1-p)(1-\epsilon)(1-G(s^*(q)))$, futhermore, it will be implemented by the competent agent with probability p and implemented by the incompetent agent with probability $(1-p)\epsilon + (1-p)(1-\epsilon)(1-G(s^*(q)))$, while it will be modified to be wrong with probability $(1-p)(1-\epsilon)G(s^*(q))$ which is completely implemented by the incompetent agent. Then the expected outcome conditional on correct initial selected project will be:

$$p + (1-p)\epsilon\kappa + (1-p)(1-\epsilon)[\kappa - (1+\kappa)G(s^*(q))]$$

If the initial selected project is wrong, the wrong project will be modified to be correct with probability $[p + (1-p)\epsilon]F_-(s^*(q)) + (1-p)(1-\epsilon)G(s^*(q))$, furthermore, it is implemented by the competent agent with probability $pF_-(s^*(q))$ and implemented by the incompetent agent with probability $(1-p)\epsilon F_-(s^*(q)) + (1-p)(1-\epsilon)G(s^*(q))$. While the wrong policy will continue with probability $[p + (1-p)\epsilon](1 - F_-(s^*(q))) + (1-p)(1-\epsilon)(1-G(s^*(q)))$. Then the expected outcome conditional on wrong initial selected project will be:

$$p(2F_-(s^*(q)) - 1) + (1-p)\epsilon[(1+\kappa)F_-(s^*(q)) - 1] - (1-p)(1-\epsilon)(1-(1+\kappa)G(s^*(q)))$$

The total expected outcome will be:

$$\begin{aligned} W(q,\epsilon,\kappa) &= q\{p + (1-p)\epsilon\kappa + (1-p)(1-\epsilon)(\kappa - (1+\kappa)G(s^*(q)))]\} + (1-q)\{p(2F_-(s^*(q)) - 1) \\ &\quad + (1-p)\epsilon[(1+\kappa)F_-(s^*(q)) - 1] - (1-p)(1-\epsilon)[1-(1+\kappa)G(s^*(q))]\} \\ &= q[p + (1-p)\kappa] - (1-q) + 2\{[p + \frac{1}{2}(1-p)\epsilon(1+\kappa)] - \frac{1}{2}(2q-1)(1-p)(1-\epsilon)(1+\kappa)\}G(s^*(q)) \end{aligned}$$

From above first equality to the second equality, we apply

$$(1-q)F_-(s^*(q)) = G(s^*(q))$$

While $G(s^*(q)) \geq 0$, furthermore

$$[p + \frac{1}{2}(1-p)\epsilon(1+\kappa)] - \frac{1}{2}(2q-1)(1-p)(1-\epsilon)(1+\kappa) = p - \frac{1}{2}(1+\kappa)[2q(1-\epsilon) - 1](1-p) > 0$$

as $1 - p < p$ and $\frac{1}{2}(1+\kappa)[2q(1-\epsilon) - 1] < 1$.

Then there will be :

$$W(q) \geq q[p + (1-p)\kappa] - (1-q)$$

Though there is information distortion in the transmission, the agent's report still makes the project better than the situation without experiment.

Second step, we will show $W(q) < W(\frac{1}{2})$ when $\frac{g(0)}{f_-(0)} < \frac{(1-q)(p+(1-p)\epsilon)}{(2q-1)(1-p)(1-\epsilon)}$ and $q > \frac{1}{2}$

Note that the parameter condition $\frac{g(0)}{f_-(0)} < \frac{(1-q)(p+(1-p)\epsilon)}{(2q-1)(1-p)(1-\epsilon)}$ implies that the agent believes that the policy should change when he receives negative signal and the policy should continue when he receives positive signal.

Given $\frac{g(0)}{f_-(0)} < \frac{(1-q)(p+(1-p)\epsilon)}{(2q-1)(1-p)(1-\epsilon)}$ holds and $q > \frac{1}{2}$, we can obtain

$$f_-(0) > \frac{(2q-1)(1-p)(1-\epsilon)}{(1-q)(p+(1-p)\epsilon)} g(0)$$

As $g(s)/f_-(s)$ is increasing in $s \in \Omega_-$, so

$$f_-(0) > \frac{(2q-1)(1-p)(1-\epsilon)}{(1-q)(p+(1-p)\epsilon)} g(0) \Rightarrow f_-(s) > \frac{(2q-1)(1-p)(1-\epsilon)}{(1-q)(p+(1-p)\epsilon)} g(s), \text{for } \forall s \in \Omega_-$$

Furthermore, we can note that

$$\begin{aligned} F_-(s^*(q)) &= 1 - \int_{s^*(q)}^{0} f_-(s)ds \\ &< 1 - \frac{(2q-1)(1-p)(1-\epsilon)}{(1-q)(p+(1-p)\epsilon)} \int_{s^*(q)}^{0} g(s)ds \\ &= 1 - \frac{(2q-1)(1-p)(1-\epsilon)}{(1-q)(p+(1-p)\epsilon)} (\frac{1}{2} - G(s^*(q))) \end{aligned}$$

While

$$F_-(s^*(q)) = \frac{G(s^*(q))}{1-q}$$

So

$$\begin{aligned} &\frac{G(s^*(q))}{1-q} < 1 - \frac{(2q-1)(1-p)(1-\epsilon)}{(1-q)(p+(1-p)\epsilon)} (\frac{1}{2} - G(s^*(q))) \\ &\Leftrightarrow G(s^*(q)) < \frac{1}{2} \frac{2(1-q)[p+(1-p)\epsilon] - (2q-1)(1-p)(1-\epsilon)}{[p+(1-p)\epsilon] - (2q-1)(1-p)(1-\epsilon)} \end{aligned}$$

Note that $G(s^*(q)) \geq 0$ always holds and

$$(p+(1-p)\epsilon) - (2q-1)(1-p)(1-\epsilon) > 1 - 2(1-p)q > 0$$

So from above expected outcome function $W(q, \epsilon, \kappa)$, we can obtain

$$W(q, \epsilon, \kappa) = q[p+(1-p)\kappa] - (1-q) + 2\{[p + \frac{1}{2}(1-p)\epsilon(1+\kappa)] - \frac{1}{2}(2q-1)(1-p)(1-\epsilon)(1+\kappa)\} G(s^*(q))$$

$$W(\frac{1}{2}, \epsilon, \kappa) = p + (1-p)\epsilon\kappa - \frac{1}{2}(1-p)(1-\epsilon)(1-\kappa)$$

Define

$$\Delta(q, \epsilon, \kappa) = W(q, \epsilon, \kappa) - W(\frac{1}{2}, \epsilon, \kappa)$$

Note that $\Delta(q, \epsilon, \kappa)$ is a linear function in κ, so

$$\max_{\kappa \in [0,1]} \Delta(q, \epsilon, \kappa) = \max\{\Delta(q, \epsilon, 0), \Delta(q, \epsilon, 1)\}$$

Firstly, we note

$$\begin{aligned}\Delta(q,\epsilon,1) &= 2q-1+2\{[p+(1-p)\epsilon]-(2q-1)(1-p)(1-\epsilon)\}G(s^*(q))-[p+(1-p)\epsilon]\\ &< 2q-1+2(1-q)[p+(1-p)\epsilon]-(2q-1)(1-p)(1-\epsilon)-[p+(1-p)\epsilon]=0\end{aligned}$$

The inequaility applies above inequality of $G(s^*(q))$.

Secondly, we note

$$\begin{aligned}\Delta(q,\epsilon,0) &= (1+p)q-1+2\{[p+\frac{1}{2}(1-p)\epsilon]-\frac{1}{2}(2q-1)(1-p)(1-\epsilon)\}G(s^*(q))-[p-\frac{1}{2}(1-p)(1-\epsilon)]\\ &< -(1-q)(1+p)+\frac{1}{2}(1-p)(1-\epsilon)+\frac{1}{2}[2(1-q)[p+(1-p)\epsilon]-(2q-1)(1-p)(1-\epsilon)]+\\ &\frac{p}{2}\frac{2(1-q)[p+(1-p)\epsilon]-(2q-1)(1-p)(1-\epsilon)}{[p+(1-p)\epsilon]-(2q-1)(1-p)(1-\epsilon)}=-\frac{1}{2}\frac{(2q-1)^2(1-p)(1-\epsilon)}{[p+(1-p)\epsilon]-(2q-1)(1-p)(1-\epsilon)}p<0\end{aligned}$$

The inequaility applies above inequality of $G(s^*(q))$.

So we can see

$$\Delta(q,\epsilon,\kappa)<0\Leftrightarrow W(q,\epsilon,\kappa)<W(\frac{1}{2},\epsilon,\kappa)$$

holds for $\forall\epsilon\in(0,1),\kappa\in(0,1)$ when q satisifes

$$\frac{g(0)}{f_-(0)}<\frac{(1-q)p}{(2q-1)(1-p)}$$

Define $\underline{q}$ is the solution of

$$\frac{g(0)}{f_-(0)}=\frac{(1-q)p}{(2q-1)(1-p)}$$

As $\frac{(1-q)p}{(2q-1)(1-p)}$ is decreasing in q, furthermore, it tends to $+\infty$ when $q\to\frac{1}{2}$ and tends to 0 when $q\to 1$, so there is an unique solution $\underline{q}\in(\frac{1}{2},1)$. Then when $q\in(\frac{1}{2},\underline{q})$, $\frac{g(0)}{f_-(0)}<\frac{(1-q)p}{(2q-1)(1-p)}$ always holds.

Eventually, we can conclude that when the available prior information $q\in(\frac{1}{2},\underline{q})$ which is limited enough, rational ignorance about the informative signal and keep naive can lead to better social welfare compared to knowing the informative signal.

Third step, when $q>\frac{1+p+(1-p)\epsilon\kappa-\frac{1}{2}(1-p)(1-\epsilon)(1-\kappa)}{1+p+(1-p)\kappa}$, from above analysis, we will obtain that

$$W(q,\epsilon,\kappa)\geq q[p+(1-p)\kappa]-(1-q)>p+(1-p)\epsilon\kappa-\frac{1}{2}(1-p)(1-\epsilon)(1-\kappa)=W(\frac{1}{2},\epsilon,\kappa)$$

holds when

$$q>\frac{1+p+(1-p)\epsilon\kappa-\frac{1}{2}(1-p)(1-\epsilon)(1-\kappa)}{1+p+(1-p)\kappa}$$

Define

$$\bar{q}=\frac{1+p+(1-p)\epsilon\kappa-\frac{1}{2}(1-p)(1-\epsilon)(1-\kappa)}{1+p+(1-p)\kappa}$$

So $\exists\bar{q}$, when $q>\bar{q}$, better prior information generates better expected project outcome. □

Proof of proposition 9

Proof. Given assumption 3, if the agent faces cost η, he will take an effort $e_a = 0$ and the outcome will be always 0. If the agent faces cost 0, he will always take an effort $e_a = 1$.

If the initial project is correct, this project is implemented by the competent agent with probability p. This project keeps correct and is implemented by the incompetent agent with probability $(1-p)\epsilon + (1-p)(1-\epsilon)(1-G(s(q)))$. The project is modified to be wrong and implemented by the incompetent agent with probability $(1-p)(1-\epsilon)G(s(q))$. Then the expected project outcome is

$$p + [(1-p)\epsilon + (1-p)(1-\epsilon)(1-G(s(q)))]\kappa - (1-p)(1-\epsilon)G(s(q))$$

If the initial project is incorrect, this project is modified to be correct and implemented by the competent agent with probability $pF_-(s(q))$, this project is modified to be correct and implemented by the incompetent agent with probability $(1-p)\epsilon F_-(s(q)) + (1-p)(1-\epsilon)G(s(q))$, this project keeps wrong and is implemented by the incompetent agent with probability $(p+(1-p)\epsilon)(1-F_-(s(q))) + (1-p)(1-\epsilon)(1-G(s(q)))$. Then the expected project outcome is

$$pF_-(s(q)) + (1-p)\epsilon F_-(s(q))\kappa + (1-p)(1-\epsilon)G(s(q))\kappa - [(p+(1-p)\epsilon)(1-F_-(s(q))) + (1-p)(1-\epsilon)(1-G(s(q)))]$$

Then the total expected project outcome is:

$$\begin{aligned}
&q\{p + [(1-p)\epsilon + (1-p)(1-\epsilon)(1-G(s(q)))]\kappa - (1-p)(1-\epsilon)G(s(q))\} + (1-q)\{pF_-(s(q)) + \\
&(1-p)\epsilon F_-(s(q))\kappa + (1-p)(1-\epsilon)G(s(q))\kappa - [(p+(1-p)\epsilon)(1-F_-(s(q))) + (1-p)(1-\epsilon)(1-G(s(q)))]\} \\
=&(1-q)[2p + (1-p)(1+\kappa)\epsilon]F_-(s(q)) - (2q-1)(1-p)(1-\epsilon)(1+\kappa)G(s(q)) + q[p + (1-p)\kappa] - (1-q)
\end{aligned}$$

When q satisfies

$$\frac{g(0)}{f_-(0)} < \frac{(1-q)[p+(1-p)\epsilon]}{(2q-1)(1-p)(1-\epsilon)}$$

Then

$$W_D(q) = (1-\xi)[p + (1-p)\epsilon\kappa - \frac{1}{2}(1-p)(1-\epsilon)(1-\kappa)]$$

The power delegation is worse than the centralization with initial ignorance for $\forall \xi > 0$.

When $q > \bar{q}$, $W_D(q) \leq (1-\xi)(p + (1-p)\kappa)$. The centralization could induce better result when q is large enough and $\xi > 0$.

The centralization regime could be better than the decentralization regime when q is small or q is large. It is possible that decentralization is better than the centralization only when q locates in an intermediate level.

□

B Discussion of other equilibrium candidates

Except for the equilibrium characterized in proposition 6, there are other equilibrium candidates. This part will rule out or refine these equilibrium candidates.

(1) A strategy profile where the principal always replace the agent regardless of the agent's report $\lambda = 1$

When the principal's replacement rule is $\sigma(m) = 1$ iff $\hat{p} > p$. This means that the agent will also be replaced when the principal's belief about the agent's competence is $\hat{p} = p$. Keep the strategy profile except that the principal's replacement rule is modified to be $\lambda = 1$, this strategy profile can also be an equilibrium.

Here gives more remarks, it seems that the agent will be indifferent between full disclosure and other report choice in this situation. However, if we consider a strategy profile where the agent discloses full information and the principal chooses $\lambda = 1$ regardless of the agent's disclosure, we can see: For principal, $\lambda = 1$ regardless of the agent's disclosure does not satisify the sequential rationality. When the agent discloses full information, it will be optimal for the principal to keep the agent who is more likely to be the competent.

(2) A strategy profile where the principal always keep the initial agent without replacement regardless of the agent's report $\lambda = 0$.

The principal's committment that there is no replacement is not promising. Sequential rationality constraint in PBE will requires that the principal replace the agent who is more likely to be the incompetent. So the strategy profile with $\lambda = 0$ can never be a PBE.