Dynamic Incentives for Self-Monitoring

MUSTAFA DOGAN*

February 13, 2017

LATEST VERSION HERE.

Abstract

This paper studies a dynamic information acquisition problem within a regulation framework. Each period, the agent (he) would like to undertake a new project, which may cause social harm. He can acquire costly information about the type of the projects by self-monitoring, but the efforts spent on self-monitoring are only observed by him. Each period, the regulator (she) decides whether to ask the agent to self-monitor or not followed by the choice of project approval. There are no monetary transfers. Instead, the regulator uses future regulatory behavior for incentive provision. When the regulator has full commitment power, the regulator can induce costly self-monitoring and revelation of “bad news” in the initial phase of the optimal policy. During this phase, the agent is promised a higher continuation utility (in the form of future regulatory approval) each time he discloses “bad news”. Otherwise, he is downgraded to a lower continuation utility in order to incentivize information acquisition. If the regulator internalizes self-monitoring costs, the agent is either blacklisted or whitelisted in the long run. When she does not internalize these costs, blacklisting is replaced by a temporary probation state, and whitelisting becomes the unique long run outcome. This result suggests that whitelisting, which may appear to be a form of regulatory capture, may instead be a consequence of optimal policy. When the regulator has limited commitment power in that she cannot commit to a policy with a negative continuation value, the results change remarkably. If the expected social harm of a project is higher than its economic benefits, whitelisting disappears. In this case, if the regulator does not internalize the self-monitoring costs, the policy never reaches a stable outcome and fluctuates over time.

Keywords: Dynamic contracts, principal-agent, linking decisions, self-monitoring, regulation.

*Department of Economics, University of Pennsylvania. Email: mudogan@sas.upenn.edu.
I am grateful to my advisor George Mailath for his encouragement and insightful comments. I also thank my committee members Steven Matthews and Mallesh Pai. Finally, I thank Aislinn Bohren Ramazan Bora, David Dillenberger, Selman Erol, Daniel Hauser, Nick Janetos, Joonbae Lee, Sangmok Lee, Wojciech Olszewski, Andrew Postlewaite, Francisco Silva, Rakesh Vohra, Yuichi Yamamoto and Penn Micro Theory Seminar participants for their valuable comments. All remaining errors are mine.
1 Introduction

The U.S. Environmental Protection Agency has an auditing policy that encourages companies to monitor their ongoing and planned activities that may fall within its authority, and to voluntarily report their violations.\(^1\) This is an example of the framework that I explore in this paper. I am interested in understanding the behavior of regulators in environments where the regulated activity may result in bad outcomes and where there is significant uncertainty. Agents have an advantage in acquiring information about these activities because they have lower costs of monitoring. The regulator, in an efficient regulatory regime, would like to use agents’ self-monitoring. I study how regulators can induce economic agents to acquire and disclose costly information about the negative consequences of their activities through the use of future regulatory behavior without resorting to monetary transfers.

I show that, when the regulator has full commitment power, the optimal policy can induce self-monitoring only in an initial phase, which endures over a stochastic number of periods. When it ends, a terminal phase of the policy is initiated and self-monitoring stops. The outcome in this terminal phase is history-dependent and involves either blacklisting the agent or whitelisting him. When the regulator does not internalize self-monitoring costs, blacklisting is replaced by a temporary probation state. The unique long-run outcome is whitelisting in this case. This result suggests that whitelisting, which may appear to be a form of regulatory capture, may instead be a consequence of optimal policy. I also analyze the case in which the regulator’s commitment power is limited so that she cannot commit to policies with negative continuation values. In this case, if the expected cost of the social harm is larger than the economic benefits of the projects, then whitelisting never occurs in an optimal policy. Moreover, self-monitoring is sustained over the long term when the regulator does not internalize its costs.

In general, many enforcement authorities adopt self-monitoring practices for various regulatory purposes.\(^2\) A specific practice is the process of issuing licenses for activities with possible environmental consequences. A mining company, for example, in applying for a mining license, may be asked to submit an Environmental Impact Statement and sometimes other supplementary information that requires substantial and costly self-monitoring. The grant of the license empowers the company to operate and contributes to the aggregate economy. Yet, it may also cause some undesirable social consequences that the regulator needs to take into account.\(^3\) To make matters worse, these undesired outcomes oftentimes take a considerable amount of time to become apparent so that it is no more feasible to take ameliorating action.\(^4\) Therefore, investigating these potential

---

\(^1\)The auditing policy is defined in Environmental Protection Agency (2000), titled “Incentives for Self-Policing: Discovery, Disclosure, Correction and Prevention of Violations”.


\(^3\)For example, the mining area may have invisible connections to groundwater resources, in which case mining activities might lead to the production and the spread of hazardous material.

\(^4\)There are many cases, in the mining industry, for example, where the actual damage become apparent only
harms prior to making a licensing decision is the only convenient policy for the regulator. And these investigations are delegated to the applicant company through the request of Environment Impact Statement.

Incentive divergence is the most prominent feature of the aforementioned settings. The agents prefer to avoid suspension of their activities and also generally prefer to avoid monitoring due to its costs and the possibility of unfavorable signals that it might reveal. On the other hand, the regulators care about efficiency for which monitoring and suspending harm-producing activities are essential. Therefore, the regulator has a complicated policy problem that involves supervising the agents and at the same time incentivizing them to self-monitor.

To study the regulator’s problem, I construct a dynamic principal-agent model in which a stream of projects arrives over time, one for each period. The agent (he) wants to undertake the projects, but needs the approval of the regulator (she), who is the principal. The projects are ex-ante identical, yielding the same revenue for the agent. On the other hand, a project might either be harmless or it might result in social costs which would outweigh the value that the project generates. The agent can acquire information about whether the project has social costs by costly self-monitoring, but the efforts spent on self-monitoring are not directly observed by the principal. The principal has the objective of inducing socially preferable outcomes. Her preferences involve the economic benefits as well as the social costs resulting from these projects. Each period, the principal first decides whether or not to ask for self-monitoring, and then chooses whether to approve the project. There is no ex-post monitoring, and the realized harms are never observed. In many settings, the harms occur, or become evident, with a significant lag compared to the economic yields of the projects. To abstract from this reality, I analyze the case in which this lag goes to infinity.

There are no monetary transfers. In many situations, regulatory agencies are limited in their ability to use monetary transfers for various reasons. For example, in some industries there is a legal limitation on the size of the monetary fines that regulators can levy.\(^5\) As a result, monetary transfers are too weak to induce proper behavior, and the regulator needs to use other tools for incentive provision. This paper focuses on the extreme case of the regulators’ limited use of monetary transfers by ruling them out. In the absence of monetary transfers, the regulator provides incentives by linking her decisions over time.

The information structure governing the self-monitoring process takes the form of verifiable “bad news” which are publicly observed. More precisely, there is a unique verifiable signal perfectly revealing bad news and informing about the harm that will occur if the project is undertaken. In case the agent performs self-monitoring in a particular period, conditional on the project being harmful, the signal will be realized with some probability and will be publicly observed. If there is no decades after the operations took place.

\(^5\)In the oil and gas industry of the U.S., there is a daily limit on the maximum amount of fines, and this limit varies across states in the U.S. The total amount of fine the regulating authority collects is is negligible compared to the economic benefit that the companies receive, see for instance E&E Publishing, LLC (2011).
news, then there are two possibilities from the principal’s perspective. First, the agent shirked and did not monitor. Second, the agent acquired information; however, no signal was realized since the project is more likely not to cause harm. There is no direct signal indicating good news. In most of the settings that fit into this paper, the only good news is the absence of bad news. In other words, certifiably disclosing good news is not possible. On the contrary, bad news, in general, provides concrete evidence and detailed description of the harm that will occur if the project is undertaken. Conditional on this information structure, assuming that the signal is publicly observed is without loss of generality. As long as the agent prefers to monitor himself, he also prefers to disclose the signal in case it is realized. Otherwise, he could simply shirk in the first place and eliminate the cost of monitoring. The incentives that induce information acquisition automatically induces the disclosure contingent upon acquisition. Therefore, the signal remains public throughout the discussion in the paper, and hence self-disclosure exogenously occurs conditional on self-monitoring.

Initially, I study the case in which the principal has full commitment power. At the beginning, she commits to history-contingent policy that specifies her decisions regarding self-monitoring requests, and the approval of the projects in each period. I show that self-monitoring is only induced in an initial phase of the optimal policy. During this initial phase, the agent is promised a higher continuation utility (in the form of future regulatory approval) each time he discloses bad news. His current project is less likely to get approved, but the regulator promises more frequent approval in the future. If he does not disclose any signal, he is downgraded to a lower continuation utility. His current project has higher chances of approval, yet he will be given less frequent approvals in the future. The duration of this phase is stochastic; when it ends, the policy reaches a second phase in which there is no more self-monitoring. The transitional dynamics between the phases and the long-run outcome of the optimal policy depends on whether the principal internalizes the cost of self-monitoring.

If the principal internalizes the costs of self-monitoring, the acquired information is always used in the approval decision. The agent’s continuation utility eventually reaches either its minimum or maximum and remains constant. In this stage, the principal permanently rejects or permanently approves projects, that is, the agent is either blacklisted or whitelisted in the long run. When the principal does not internalize the self-monitoring costs, the content of the information is not always used in the current approval decision, in contrast to the previous case. There is a probation state, which replaces blacklisting, wherein the agent acquires information, but the project is rejected regardless of the outcome. The probation occurs when the agent’s continuation utility reaches its minimum in consequence of the agent not disclosing bad news frequent enough. After being initiated, this probationary state repeats until the agent discloses some bad news. Leaving the probation state today does not rule out the possibility of facing it again in the future. The agent’s continuation utility eventually reaches its maximum which still puts permanent approval

---

6 EPA blacklists companies at times, by labeling them as ineligible for federal contracts, subcontracts, grants or loans. See Washington Post (1977). The outcome permanent rejection can be considered as the counterpart of blacklisting.
into action, in the long run, the agent is always whitelisted.

The above-mentioned difference in the principal’s preferences alters the set of effective incentive devices she is willing to use. When she does not internalize its costs, self-monitoring can purely be used to punish the agent. It is possible for the principal to use self-monitoring as punishment, because verifiability ensures that monitoring effort is taken. While the same channel was also feasible when the principal internalized the self-monitoring costs, she preferred not to punish this way because she cared about the cost.

In this model, whitelisting is an outcome of the optimal regulatory policy. Hence, I do not interpret it as a form of regulatory capture even though it shares some of its features. My paper, therefore, suggests that what has been described as regulatory capture in some cases may instead be an outcome of optimal regulatory policy.

Two situations give rise to inefficiencies in this framework. The first one occurs when the agent forgoes information acquisition, and the second one arises when the content of the information is not used efficiently. The first type only appears during the terminal phase of the optimal policy where there is no more self-monitoring. The second type, appears in the initial phase. Its occurrence triggers the terminal phase of the contract, when principal internalizes self-monitoring costs. Therefore, the inefficiencies are back-loaded in this case. However, when the principal does not internalize the self-monitoring costs, the second type of inefficiency occurs in a non-consecutive stochastic order, and its occurrence does not necessarily initiate the terminal phase. In this respect, efficiency will be lost and restored stochastically throughout the optimal policy.

I also study the situation in which the principal has limited commitment power, in that she cannot commit to a policy with a negative continuation value. The results change remarkably. If the expected cost of a project is higher than its economic benefit, the policy does not feature whitelisting. In this case, if the principal does not internalize self-monitoring costs, the policy never reaches a stable outcome and fluctuates over time. Finally, I analyze the case in which the principal can also monitor projects, but at a higher cost than the agent. Each period, prior to making an approval decision, she chooses either to monitor on her own, or delegate it to the agent, or completely avoid monitoring. For some parameter values, a randomized decision between non-delegated monitoring and direct approval replaces whitelisting.

Related Literature
Kaplow and Shavell (1994) is the first study analyzing self-monitoring and self-reporting. They introduce a self-reporting stage into the classical probabilistic law enforcement model of Becker (1968). By self-reporting a harmful act, the agent is granted a reduction in the sanctions he faces. In contrast to my model, the agent in their paper is initially endowed with the relevant information. Pfaff and Sanchirico (2000) introduce a more general framework in which the agency problem has
two tiers: testing for noncompliance and fixing it. There is no information asymmetry to begin with, and the agent needs to exert effort to acquire relevant information. Most of the papers in this literature focus on characterizing the optimal incentive scheme in a static framework. My paper, however, studies the dynamics of a regulatory regime incorporating self-monitoring. In a contemporaneous work, Wang et al. (2016) also study a similar dynamic environment. The main distinction is that the harms are already known to agent and monetary transfers are allowed in their framework. They show that the optimal regime, in order to induce the agent to disclose harms, incorporates a cyclical structure alternating between rewarding self-disclosure and initiating inspections. Departing from theirs, my paper provides some explanation for practices such as blacklisting which arise from dynamic consequences.

In its use of non-monetary intertemporal incentives as a disciplining device, this paper relates to several different branches of literature. In mechanism design, Horner and Guo (2015) (HG) analyze a dynamic allocation problem in the absence of monetary transfers. The principal is interested in efficiency, which requires that the principal allocate the good only if the agent has a high valuation. The optimal mechanism follows a history-dependent rule which eventually converges to permanent allocation or permanent rejection of allocation. In the literature on relational contracts, Li et al. (2015) analyze the evolution of power inside organizations within the context of what they call a repeated trust game. The efficient equilibrium has a structure similar to that of HG, incorporating a bipolar long run outcome with permanent punishment and permanent rewards for the agent. Both of these papers assumes that the agent is initially informed about the state variable. In my paper, however, the state variable is initially unknown to both (the principal and the agent); but, the agent can acquire information about it at some cost. The effort spent on information acquisition is not observed by the principal, and the agency problem is moral hazard instead of adverse selection unlike HG. The fact that the relevant information comes with a cost changes the dynamic structure of the optimal contract/policy. More precisely, when the principal does not internalize the cost of information acquisition, the long-run outcome is unique, and permanent punishment is never a part of the optimal policy in contrast to HG and (Li et al., 2015). Moreover, if the principal has a limited commitment power, the optimal policy does not reach a stable outcome, instead it fluctuates over time.

Lipnowski and Ramos (2015) consider the repeated game version of HG. The efficient equilibria in their framework have a unique long-run outcome featuring a permanent punishment for the agent, for much the same reason permanent does not occur in the limited commitment section of my paper. In contrast to Lipnowski and Ramos (2015), I show that the optimal policy does not necessarily reach a stable outcome in this case, when the principal does not internalize the self-monitoring costs. Battaglini (2005) focuses on the same allocation problem as in HG without ruling out monetary transfers. The principal is a profit-maximizing monopolist. In this paper, the inefficiencies are entirely front-loaded. The use of monetary transfers played an important role on

---

7 Also see Short and Toffel (2008), Innes (1999a), and Innes (1999b).
this significant difference, as it alters the natural way of providing incentives.

This paper also relates to the literature on relationship formation and trading favors. Möbius (2001), Hauser and Hopenhayn (2008), and Abdulkadiroğlu and Bagwell (2013) are examples. Within a repeated game setting, players facilitate cooperation by providing favors to each other. The favor provision ability changes over time and is privately observed. Inter-temporal incentives are utilized to elicit proper behavior and sustain cooperative gains. Möbius (2001) suggests a simple chip mechanism which keeps track of the difference in the number of favors provided. There is a maximum number of chips that can be maintained in an equilibrium. Hauser and Hopenhayn (2008) improve on this by considering a more general set of mechanisms. The optimal policy in my paper can also be interpreted as a chips mechanism, in which several different factors affecting the amount of chips that the agent has for the next period. First, the quantity of the agent’s chips expands over time at a constant rate that is equal to the inverse of the discount factor. Second, the amount of chips that the agent has for the next period diminishes at an amount that is proportional to the approval rate in the current period. Finally, the agent receives a fixed amount of additional chips for each piece of bad news he discloses.

My paper also relates to the literature on linked decisions. Jackson and Sonnenschein (2007), within a static environment, showed that linking multiple independent decisions can help overcome incentive constraints. See also Cohn (2010), Hortala-Vallve (2010), and Fang and Norman (2006).

2 The Model

There is a principal (she) and an agent (he) interacting within a discrete time infinite horizon setting, and $\delta$ is the common discount factor. A stream of projects arrives over time, one for each period $t = 1, 2, \ldots, \infty$. The agent would like to undertake each of these projects for which he needs the approval of the principal.

Approving a project yields a positive value $v \in (0,1)$ to the agent. In addition to this value, each project may cause a social harm depending on its type $\theta$, which takes its values from the binary space $\Theta = \{\theta_g, \theta_b\}$. If $\theta = \theta_b$, then the project is “bad”, producing harm with a magnitude normalized to 1. Otherwise, if $\theta = \theta_g$, then the project is “good”, and does not produce any harm. The type of the project is initially unknown to the principal and the agent, and $\mu = P(\theta = \theta_b)$ is the common prior about it. The project types are independently and identically distributed over time; hence, the project arriving at the beginning of each period is believed to be a bad one with probability $\mu \in (0, 1)$.

The principal is interested in efficiency, and wants to maximize the social surplus in her decision. The surplus resulting from the approval of the project is equal to $v$ or $v - 1$ depending on whether the project is good or bad, respectively. The agent, on the other hand, only cares about the value that the project generates for him, hence his utility increases by $v$ each time a project is approved.
irrespective of its type. Rejecting the project causes a loss due to the forgone value $v$, yet, at the same time, prevents the production of probable harms. Therefore, from an ex-post point of view, she wants to grant an approval for a project only if it is a good one.

At each period, the agent can acquire information about the type of the project at cost $c$ prior to the principal’s approval decision. The information acquisition process, which is also referred as self-monitoring, is governed by the following information structure. There is a unique verifiable signal “$s$” which perfectly reveals “bad news” about the type of the project. Conditional on the project being bad, the signal is realized with probability $\lambda \leq 1$. If the project is good, then the signal is never realized. More precisely:

$$P(s|\theta_b) = \lambda, \quad P(s|\theta_g) = 0.$$  

I assume that the signal is publicly observed whenever it is realized.\textsuperscript{8} Due to this publicity, “self-reporting”, which refers to the event of signal realization, exogenously follows conditional on self-monitoring.

The event of no signal realization following the information acquisition, besides being informative, does not perfectly reveal the type of the project (unless $\lambda = 1$). Conditional on information acquisition, the posterior beliefs after signal realization and no signal realization are denoted by $\mu_s$ and $\mu_{ns}$ respectively, which satisfy:

$$\mu_s = 1$$
$$\mu_{ns} = \frac{\mu(1-\lambda)}{1-\mu\lambda}.$$  

In the ex-ante stage the expected cost of a project is equal to $\mu = \mu_1 + (1-\mu)0$. Therefore, the ex-ante expected surplus that arises from the project approval is $v - \mu$. There is no assumption imposed on the sign of this value, hence both approval and rejection can be the optimal uninformed decision. The assumptions on the parameters that are maintained throughout the entire paper are defined as follows:

ASSUMPTIONS.

i) $v > \mu_{ns}$.

ii) $(1-\mu\lambda)(v-\mu_{ns}) - c > \max(v-\mu, 0)$.

iii) $\delta > \frac{1}{1+\mu\lambda v}$.

The first assumption asserts that, from the principal’s perspective, it is optimal to approve the

\textsuperscript{8} This publicity assumption is without loss of any generality. All of the results would also follow in a more general framework, where the agent privately observes the signal and then decides whether or not to disclose it to the principal. This stems from the fact that, as long as the agent prefers to monitor himself, he also prefers to disclose the signal in case it is realized. Otherwise, he could simply shirk in the first place and eliminate the cost of monitoring. Of course, this property crucially depends on the signal structure governing the information acquisition.
project in case no signal realization takes place as a result of information acquisition. The second assumption guarantees that the information acquisition is efficient, hence the problem is not a trivial one. The third assumption asserts that the discount factor large enough and players are sufficiently patient.

The effort spent on self-monitoring is not observed by the principal, and this generates the moral hazard component which constitutes the main source of the agency problem. If the incentives are not provided properly, then the agent would shirk rather than monitoring the projects. By this, the agent can get rid of the cost of self-monitoring, and at the same time prevent the revelation of bad news and hence the suspension of his projects. Nonetheless, because the information acquisition is efficient, the principal wants to design an incentive scheme to motivate the agent towards this end. Note that the verifiability of the signal plays a crucial role. It would never be feasible to induce self-monitoring under an information structure that comprises only non-verifiable soft information.

There are no monetary transfers. In many situations, regulatory agencies has limited ability to use monetary transfers, which leads them to use other tools such as future regulatory behavior for incentive provision. In this framework, the principal would be able to induce the first-best outcome under the presence of monetary transfers, by using a stationary payment scheme. Such a stationary scheme, however, would be insufficient to sort out the extent to which the principal utilizes the continuation values arising from future regulatory behavior as an incentive device. In order to analyze these dynamics, one should either employ a more general framework incorporating additional aspects, or restrict the existing one. To eliminate technical difficulties and maintain tractability, I follow the latter and rule out the monetary transfers in the analysis.

There is no initial information asymmetry about the type of the project. This does not rule out the possibility of an agent having superior information about other relevant issues. For instance, as the owner of the project, the agent might be better informed about the direction in which to search for “bad news”. This can be considered as the basis for agent’s comparative advantage in terms of monitoring capability. This is effectively an information asymmetry, yet it is not directly related to the type of the project.

Ex-ante monitoring is the only source of information on the type of the project. There is no possibility of ex-post monitoring, and hence the realized harm is never observed. This assumption reflects the fact that, in many circumstances, the harms take place with a significant lag compared to the economic yields of the projects.

A later section of this paper analyzes a model where the principal can also monitor the project with a higher cost. In that setting, in each period prior to making the approval decision, the principal makes a decision regarding monitoring. She either monitors the project on her own, or delegates monitoring to the agent, or completely avoids monitoring. In the current model, requesting self-

---

9 The first-best scheme involves delegating monitoring to the agent and employing the efficient approval decisions in every period.
monitoring from the agent is not delegation since the principal does not have an option to monitor the projects on her own. Nonetheless, for notational ease, I will use the word delegation to refer to a self-monitoring request in this section as well.

**Actions and Preferences**

For each period $t$, after a new project arrives, the principal first decides whether or not to delegate monitoring to the agent. If delegation occurs, the agent moves and decides either to shirk or to exert effort and monitor himself. Then, finally, the principal moves again and decides whether or not to approve the project. This stage game is repeated infinitely many times.

The scope of the conflict between the principal and the agent is not limited to the social costs that the projects may generate. The monitoring costs that the agent assumes is another factor contributing to the extent of the conflict. This paper studies two different specifications. Under the first specification, the principal internalizes the cost of self-monitoring; in the second one, she does not internalize it. The conflict becomes more intense under the second specification. The corresponding stage game, together with the payoffs corresponding to the first specification, is illustrated in the following figure. Note that the principal’s payoffs are in expectation terms except for those terminal nodes resulting from project rejections, and signal realization. The expectation is based on the belief about the type of the project, which is either $\mu$, if no information is acquired, or $\mu_{ns}$ if information is acquired but no signal is realized.
Figure 1: The stage game between the principal and the agent. The principal internalizes the cost of self-monitoring. All the terminal nodes, except those are reached by nature's move $\theta_b$s, and project rejections, are reflecting principal’s expected payoffs over the determination of the type of the project. $\mu$ is the unconditional probability of project being bad, and $\mu_{ns}$ is the probability of project being bad conditional on no signal is realized after information acquisition.

**An Alternative Interpretation**

There is an alternative interpretation of the model. Suppose that there are some costly precautionary measures that can be initiated by the agent if the principal forces him to do so. From the principal’s point of view, these precautions are necessary in case the project is bad since they completely eliminate the harms; otherwise, they are wasteful. The agent wants to avoid these measures regardless of the type of the project, due to the costs. Let $z$ be the cost of these measures; then the corresponding ex-post payoffs are described in the following table:

<table>
<thead>
<tr>
<th>Precautions</th>
<th>No Precautions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>$v-z, v-z$</td>
</tr>
<tr>
<td>Bad</td>
<td>$v-z, v-z$</td>
</tr>
</tbody>
</table>

Table 1: Description of the corresponding payoffs at the ex-post stage for the principal and the agent respectively.

When $z = v$, the above description is equivalent to the baseline model, where forcing the precautions
corresponds to the rejection of the project, and vice versa. However, the results of the paper also follow for a larger set of \( z \) values. One just needs to make sure that \( z \) is not very different from \( v \).\(^{10}\)

**The Policy**

The principal is endowed with the full commitment power, and, at \( t = 0 \), commits to a dynamic policy that specifies delegation and approval decisions over time as a function of the public history. The public history consists of information about the realized decisions in the earlier periods as well as the self-monitoring outcomes for those periods in which the monitoring is delegated to the agent.

For a given time period \( t \), the corresponding delegation decision and the realized outcome of the self-monitoring, if performed, are denoted together by \( r_t \in \{s, ns, n\} \). When monitoring is delegated to agent, \( r_t \) will be either “\( s \)”, if self-disclosure takes place; or will be “\( ns \)”, if no self-disclosure takes place. If there is no delegation, then \( r_t = n \). Moreover, the approval decision at time \( t \) is denoted by \( d_t \in \{0, 1\} \), where 0 and 1 indicates rejection and approval respectively. Consequently, a within-period public history, at the end of the period, which is denoted by \( h_t \), is of the following form:

\[
h_t = (r_t, d_t) \in \{s, ns, n\} \times \{0, 1\},
\]

for each \( t \). At the beginning of a period \( t \), a public history is defined as

\[
h^t = (h_1, \ldots, h_{t-1}).
\]

The initial history is \( h^1 = h_0 = \emptyset \), and \( H^t \) is the set of public histories at period \( t \).

A policy \( \Gamma = \{\gamma_t, x_t\}_{t=1}^{\infty} \) is then a sequence of functions which are defined as follows:

\[
\gamma_t : H^t \to [0, 1],
\]

\[
x_t : H^t \times \{s, ns, n\} \to [0, 1].
\]

The function \( \gamma_t \) is the probability of delegation. Because the delegation takes place at the beginning of each period, it is a function defined over the set \( H^t \). On the contrary, the approval decision is also conditioned on the value of \( r_t \), hence the relevant domain for \( x_t \) is \( H^t \times \{s, ns, n\} \). For each possible value of \( r_t \), I use a separate notation for the approval rate, i.e., \( x_t = (x^s_t, x^{ns}_t, x^n_t) \). Note that, \( x^s_t \) and \( x^{ns}_t \) are well-defined as long as \( \gamma > 0 \); similarly \( x^n_t \) is relevant when \( \gamma < 1 \).

For an incentive compatible policy \( \Gamma \), the expected utilities of the principal and the agent are denoted by \( V \) and \( U \) respectively, and are given by:

\[
U = \mathbb{E} \left[ (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} \left[ \gamma_t \{\mu \lambda x^s_t v + (1 - \mu \lambda)x^{ns}_t v - c \} + (1 - \gamma_t)x^n_t v \right] \right]
\]

\(^{10}\)For instance, when \( z \) is sufficiently small, then the principal would always force these measures and avoid costly information acquisition.
The posterior beliefs about the type of the project affect only the principal’s utility. The agent does not care about the project’s type. Another important point is that the self-monitoring cost is included in the principal’s utility. As it was mentioned earlier, the other case will be analyzed later on.

**Stationary Representation**

Following Spear and Srivastava (1987), I express the principal’s problem within a stationary form, in which the ex-ante expected utility of the agent is the state variable. In this form, the interval \([0,v]\) is the corresponding state space as it consists of all of the possible values that the ex-ante expected utility of the agent can take. The agent’s utility cannot be negative, because he can always guarantee a non-negative utility by shirking every time he is asked to monitor. On the other hand, \(v\) is the maximum that the agent can receive in a policy. The principal can grant this maximal utility to agent by approving all of the projects without requesting self-monitoring.

State variable is updated over time depending on the realized public history. Within-period decisions and the promised future continuation utilities of the agent, depend on this state variable as well as the realized outcomes of the current period. An optimal policy specifies a different continuation utility for each possible \(r_t \in \{s, ns, n\}\) as in the case of the approval probabilities. More precisely, the components of the policy are defined as:

\[
\gamma, x_s, x_{ns}, x_n : [0,v] \rightarrow [0,1], \\
U_s, U_{ns}, U_n : [0,v] \rightarrow [0,v].
\]

The delegation and the approval decisions consist, in essence, of probabilities; therefore, the functions \(\gamma, x_s, x_{ns}, x_n\) take their values from the unit interval. The continuation utilities, on the other hand, specify the state variable for the next period; hence, they are defined as functions from the state space to itself. For a given policy, the functions \(U_s, U_{ns}, x_s, x_{ns}\) are relevant only for those values of \(U \in [0,v]\) satisfying \(\gamma(U) > 0\), whereas the functions \(U_n\) and \(x_n\) are relevant only for those satisfying \(\gamma(U) < 1\).

The promised utility of the agent is calculated in ex-ante terms; hence, it will be granted to the agent only in expectation. It aggregates the flow and continuation utilities of the agent. Its transition is governed by the policy and the stochastic realizations. Starting from \(U\), the state variable of the next period becomes \(U_n(U)\), or \(U_s(U)\), or \(U_{ns}(U)\) with the corresponding probabilities \(1 - \gamma(U)\), \(\gamma(U)\mu\lambda\) and \(\gamma(U)(1 - \mu\lambda)\) respectively.

In an optimal policy, the functions \(U_n, U_s, U_{ns}, \gamma, x_n, x_s, x_{ns}\) are chosen to maximize the principal’s utility.
objective function. There are two constraints that the principal needs to take into account in this problem: incentive constraint and promise keeping constraint. More precisely:

**Principal’s Problem** $\mathcal{P}$

$$V(U) = \max_{x_n, x_s, x_{ns}, U_n, U_s, U_{ns}} \gamma \left[ \mu \lambda \left( (1 - \delta)x_s(v - 1) + \delta V(U_s) \right) + (1 - \mu \lambda) \left( (1 - \delta)x_{ns}(v - \mu_{ns}) + \delta V(U_{ns}) \right) - c \right] + (1 - \gamma) \left[ \mu \lambda \left( (1 - \delta)x_n(v - \mu) + \delta V(U_n) \right) \right]$$

subject to the (PK) and (IC) respectively:

$$U = \gamma \left[ \mu \lambda \left( (1 - \delta)x_s v + \delta U_s \right) + (1 - \mu \lambda) \left( (1 - \delta)x_{ns}v + \delta U_{ns} \right) - c \right] + (1 - \gamma) \left[ \mu \lambda \left( (1 - \delta)x_n v + \delta U_n \right) \right],$$

$$\gamma \left[ \mu \lambda \left( (1 - \delta)x_s v + \delta U_s \right) + (1 - \mu \lambda) \left( (1 - \delta)x_{ns}v + \delta U_{ns} \right) - c \right] \geq \gamma \left[ \mu \lambda \left( (1 - \delta)x_{ns}v + \delta U_{ns} \right) \right].$$

The first line of the principal’s objective function includes her utility contingent upon self-monitoring request; hence, it is multiplied by the probability $\gamma$. The second line, on the other hand, corresponds to the contingency of no self-monitoring; hence, it is multiplied by $1 - \gamma$.

The first constraint is the promise keeping constraint. It makes sure that the agent’s expected utility is equal to $U$, the utility level promised to him. Likewise the principal’s utility, the agent’s utility also has two components depending on the principal’s delegation decision.

The second constraint is the incentive constraint, which is defined to make sure that acquiring information is an optimal choice for the agent when he is asked to do so; hence, it is relevant only if $\gamma > 0$. It guarantees that the utility the agent achieves from shirking is no better than the promised utility. If he shirks, there is no self-disclosure; hence, the current approval rate and the continuation utility will be equal to $x_{ns}$ and $U_{ns}$ respectively.

Note that Blackwell’s sufficiency conditions, i.e., monotonicity and discounting, are fulfilled. Therefore, the existence of a solution for the problem $\mathcal{P}$ is guaranteed.

3 Case 1: Principal Internalizes Cost

In this section, I study the optimal policy and its properties under the first preference specification, i.e., when the principal internalizes the self-monitoring costs. First, I start with a benchmark analysis.
3.1 Observable Information Acquisition

In this benchmark, I consider the case in which the agent’s self-monitoring costs are publicly observed; hence, there is no agency problem. The natural question is whether the principal can induce the first best outcome or not? The first best outcome involves information acquisition and utilization of the optimal approval decision, i.e., approving the project if no-signal is generated and rejecting it otherwise, in every period. The principal can simply induce this first best outcome by punishing any deviation with permanent rejection of the future projects. Such an incentive scheme is sufficient to induce the proper behavior; because the agent has a positive utility from the first best outcome, whereas permanent rejection leaves him a 0 payoff. The expected utilities of the agent and the principal are denoted by \( w \) and \( \pi \) respectively, and are given by:

\[
\begin{align*}
 w &= (1 - \mu \lambda)v - c, \\
 \pi &= (1 - \mu \lambda)(v - \mu_{ns}) - c.
\end{align*}
\]

The optimal policy when the self-monitoring effort is observable induces the first best outcome, which gives the maximum possible utility to the principal. What about the optimal policy conditional on the agent receiving a certain utility \( U \in [0, v] \) when the effort is observable? To answer this question, one needs to solve the same problem \( P \) without including the incentive constraint. This problem has a solution, and the value function arising from its solution, which I denote by \( V^* \), is an upper-bound for the value function \( V \).

To describe \( V^* \) and the corresponding “benchmark policy”, I will point out some initial observations. First of all, the benchmark policy is stationary without loss of generality. The expected utility of the agent stays constant throughout time; and hence, one just needs to characterize the delegation and the approval decisions, which I denote by \( \gamma^*, x^*_n, x^*_s, x^*_{ns} \), as functions of \( U \in [0, v] \).

Second, the information must be used efficiently. In other words, whenever the agent is asked to monitor himself, the following approval decision must be efficient conditional on the content of the resulting information. This is a direct implication of the fact that the principal internalizes the self-monitoring costs. Rather than having \( x^*_s > 0 \) or \( x^*_{ns} < 1 \), the principal could adjust the probability of delegation, \( \gamma^* \), without hurting the promise keeping constraint, and get strictly better off. To see this, first note that:

\[
U = \gamma^*[\mu \lambda x^*_s v + (1 - \mu \lambda)x^*_{ns} v - c] + (1 - \gamma^*)x^*_n v.
\]

Suppose, \( x_s > 0 \) to get a contradiction, then it must be true that \( x_{ns} = 1 \); because, otherwise, there is an immediate deviation that the principal can perform by decreasing \( x^*_s \) and increasing \( x^*_{ns} \). Then, consider decreasing \( x_s \) ↓ by some \( \epsilon > 0 \); and decreasing delegation probability to \( \gamma - \zeta \). Moreover assume that the principal employs a direct approval with the remaining \( \zeta \) probability.
Then $\zeta$ satisfies:

$$\gamma^*[\mu\lambda x_s^* v + (1 - \mu\lambda)v - c] = (\gamma^* - \zeta)[\mu\lambda(x_s^* - \epsilon)v + (1 - \mu\lambda)v - c] + \zeta v.$$ 

Therefore:

$$\zeta = \frac{\gamma \mu \lambda \nu \epsilon}{\mu \lambda v(1 - x_n + \epsilon) + c}.$$ 

Such a deviation is strictly better for the principal, because:

$$\left[(\gamma^* - \zeta)[\mu\lambda(x_s^* - \epsilon)(v-1) + (1 - \mu\lambda)(v - \mu_{ns}) - c] + \zeta(v - \mu)\right] - \left[\gamma^*[\mu\lambda x_s^*(v-1) + (1 - \mu\lambda)(v - \mu_{ns}) - c]\right] > 0.$$ 

A similar contradiction follows for the case $x_{ns}^* < 1$.

Finally, upon not requesting self-monitoring, the principal either directly approves or directly rejects the projects and not chooses a randomized approval decision. Instead of choosing an $x_n^* \in (0, 1)$, she could increase the probability of delegation, $\gamma^*$, and adjust the value of $x_n^*$ by respecting the promise keeping constraint. This increases the probability of informed decision making and improves the principal’s objective.

By putting these observations in an order, one can easily specify the details of the benchmark policy. When $U \in [0, w)$, the agent is supposed to get a utility that is less than the utility he would get in the first best outcome. Therefore, the principal randomizes between the first best outcome and the direct rejection without information acquisition. On the other hand, when $U \in (w, v]$, the randomization occurs between the first best outcome and the direct approval without information acquisition. In both cases, the probability of delegation $\gamma$ is chosen such that the agent gets the exact utility promised to him. More precisely, an optimal policy in this benchmark is given by:

$$\gamma^*(U) = \begin{cases} 
\frac{U}{w} & \text{if } U \in [0, w] \\
\frac{v-U}{v-w} & \text{if } U \in (w, v]
\end{cases}$$

$$x_s^* = 0, \ x_{ns}^* = 1, \ \forall U \in (0, v)$$

$$x_n^*(U) = \begin{cases} 
0 & \text{if } U \in [0, w] \\
1 & \text{if } U \in (w, v]
\end{cases}$$

And the resulting value function is:

$$V^*(U) = \begin{cases} 
\frac{U}{w}\pi & \text{if } U \in [0, w] \\
\frac{v-U}{v-w}\pi + \frac{U-w}{v-w}(v - \mu) & \text{if } U \in (w, v]
\end{cases}$$
The following figure illustrates the value function:

![Figure 2: The benchmark value function $V^*$, which consists of an upper bound for $V$. The parameters used in the illustration assumes that $v < \mu$.](image)

3.2 Moral Hazard and The Optimal policy

The focus is now on the agency problem where the agent’s efforts spent on self-monitoring are not observed by the principal. The interest is particularly on the characterization of the value function $V$, within-period decisions $\gamma, x_n, x_s, x_{ns}$, and the promised continuation utilities $U_n, U_s, U_{ns}$, which are defined as functions of the ex-ante expected promised utility $U$. The following lemma, which is proved in appendix, is a first step towards this goal.

**Lemma 1.** The value function, $V$ is concave and hence differentiable almost everywhere. Moreover its derivative is bounded and satisfies:

$$1 - \frac{\mu \lambda}{\mu \lambda v + c} \leq V'(U) \leq 1 - \frac{\mu(1 - \lambda)}{(1 - \mu \lambda)v - c}, \quad \forall U \in [0, v]. \quad (1)$$

**Proof.** See appendix. □

The concavity of the value function is a direct implication of the fact that the principal can randomize between different utility levels while granting a specific promised utility to agent. Since $V$ is concave, it is almost everywhere differentiable, which can also be proved by applying the result of Benveniste and Scheinkman (1979). The following observations are sufficient to show the bounds of the derivative of $V$. First, for any promised utility $U$, $V(U)$ cannot be larger than $V^*(U)$. Second, the values of $V$ and $V^*$ are equal to each other at each other at the boundaries of the state space, i.e., at 0 and $v$. These observations together with the concavity require that the constant slopes of $V^*$ over the intervals $[0, w]$ and $[w, v]$ are the upper and the lower bounds of $V'$ respectively.
In each period, the principal first makes her delegation decision. In case she randomizes with some \( \gamma \in (0, 1) \), each possible outcome of this randomization may result with different utility levels for the principal and the agent. This stems from the fact that the promised utility of the agent is given only in expectation. The principal’s commitment power plays a crucial role here. She can fulfill the exact randomization that the policy specifies, and keep her promises in expectation. Otherwise, the principal, in each period that she is supposed to randomize, could pick the outcome that she prefers most instead of following the specified randomization.

In this regard, one can represent the utility of the agent as a weighted average of two components, one for each possible outcome of the delegation decision. Precisely, let \( U_D \) and \( U_N \) be the resulting utilities of the agent after delegation and no delegation respectively. More precisely:

\[
U_N = (1 - \delta)x_nv + \delta U_n \\
U_D = \mu \lambda [(1 - \delta)x_sv + \delta U_s] + (1 - \mu \lambda) [(1 - \delta)x_ns v + \delta U_ns] - c.
\]

The promise keeping constraint imposes a restriction on the choices of \( U_D, U_N, \) and \( \gamma \), so that the equality \( U = \gamma U_D + (1 - \gamma) U_N \) must hold. When the agent is delegated with certainty, i.e when \( \gamma = 1 \), it must be \( U = U_D \); similarly, \( U = U_N \) must hold when \( \gamma = 0 \).

**Conditional Representation**

In what follows, I will further exploit the above-mentioned observation, and rewrite the principal’s problem \( \mathcal{P} \) as a decomposition of two conditional sub-programs. These sub-programs are defined conditional on the current delegation decision, and their task is to characterize the approval decision in the current period as well as the continuation utilities for the next period.

The first problem is defined conditional on principal delegating monitoring to the agent in the current period. The solution to this problem characterizes the optimal values for \( x_{ns}, x_s, U_{ns}, \) and \( U_s \) depending on the value of \( U_D \). It incorporates the incentive constraint and a promise keeping constraint. The promise keeping constraint here is defined to make sure that the agent’s conditional expected utility is equal to \( U_D \).

The second program is defined conditional on agent not delegating monitoring to the agent in the current period. Its solution characterizes the optimal values of \( x_n \) and \( U_n \) depending on the value of \( U_N \). There is no incentive constraint in this problem, since there is no request of self-monitoring. There is only a promise keeping constraint defined to make sure that \( x_n \) and \( U_n \) are arranged so that the agent’s utility is equal to \( U_N \).

Conditional on no delegation, the agent’s utility can take all the values in the entire state space, hence the range of \( U_N \) is equal to \([0, v]\). However, this is not the case for \( U_D \). It is defined conditional on an incentive compatible self-monitoring in the current period; hence, the agent already assumes the cost \( c \). This means that \( U_D \) cannot be equal to \( v \) or anything sufficiently close to \( v \); therefore, the range of \( U_D \) can only be a proper subset of the state space. The exact range of \( U_D \) will be
discussed later on.

The corresponding value functions arising from these conditional programs are denoted by $V_D$ and $V_N$ respectively. The unconditional value function, $V$, is then given by:

$$V(U) = \gamma(U)V_D(U_D) + (1 - \gamma(U))V_N(U_N),$$

where $U = \gamma(U)U_D + (1 - \gamma(U))U_N$. The following diagram illustrates this conditional representation of the principal’s problem.

![Diagram illustrating the principal’s problem](image)

**Figure 3:** The principal’s problem is equivalent to solving for the optimal values of $U_D$, $U_N$, and $\gamma$.

This new form of the principal’s problem is denoted by $\mathcal{P}'$ and it is defined as follows:

$$V(U) = \max_{\gamma, U_D, U_N} \gamma V_D(U_D) + (1 - \gamma)V_N(U_N),$$

s.t. $U = \gamma U_D + (1 - \gamma)U_N$.

The sub-program that is conditional on no delegation in the current period is denoted by $\mathcal{P}_N$, and is defined as follows:

$$V_N(U_N) = \max_{x_n, U_n} \left[ (1 - \delta)x_n(v - \mu) + \delta V(U_n) \right],$$

s.t. $U_N = (1 - \delta)x_n v + \delta U_n$.

Finally, the sub-program that is conditional on delegation in the current period is denoted as $\mathcal{P}_D$, and is defined as:

$$V_D(U_D) = \max_{x_s, x_s, x_{ns}} \left[ \mu \lambda \left[ (1 - \delta)x_s(v - 1) + \delta V(U_s) \right] + (1 - \mu \lambda) \left[ (1 - \delta)x_{ns}(v - \mu_{ns}) + \delta V(U_{ns}) \right] - c \right]$$
subject to (PK$_D$) and (IC$_D$) respectively:

\[ U_D = \mu \lambda [(1 - \delta)x_s v + \delta U_s] + (1 - \mu \lambda) [(1 - \delta)x_{ns} v + \delta U_{ns}] - c, \]
\[ U_D \geq (1 - \delta)x_{ns} v + \delta U_{ns}. \]

**Road-map**

Thanks to this conditional formulation, it is possible to analyze the principal’s delegation and approval decisions separately. First, I will consider the conditional problems in isolation, and solve for the corresponding optimal approval decisions. Then, I will focus on the unconditional problem $P'$, and characterize the optimal delegation decision over the state space.

There is a crucial observation that I make use of during the above-mentioned process. If, for a given $U$, a randomized delegation decision, $\gamma \in (0, 1)$, is optimal and randomization takes place between $V_D(U_D)$ and $V_N(U_N)$, then $V(U_D) = V_D(U_D)$ and $V(U_N) = V_N(U_N)$. In other words, $\gamma = 1$ is optimal at $U_D$, and $\gamma = 0$ is optimal at $U_N$. This observation is based on the fact that the principal, in order to grant the agent his promised utility $U$, can always randomize between $V(U_D)$ and $V_N(U_N)$. Therefore, $V(U_D)$ cannot be strictly larger than $V_D(U_D)$, as it would contradict with the optimality.

On account of this, characterizing the state variables that are featuring $\gamma = 0$ or $\gamma = 1$ would be sufficient to pin down the optimal delegation decision. In other words, the focus should be on the subsets of the state space over which either $V(U) = V_D(U)$ or $V(U) = V_N(U)$ is satisfied. For the rest of the state space, there will be a randomized delegation decision, and the corresponding values $U_D$ and $U_N$ will always be a part of the subsets of $[0,v]$ satisfying $V = V_D$ and $V = V_N$ respectively.

The discussion in the sequel will follow the plan described above. In order to carry through the first step, I will first focus on the conditional problems in isolation.

**The problem $P_D$**

An initial observation is that the incentive constraint is always binding. First of all, note that the IC can be written as a restriction on the difference between the continuation utilities $U_s$ and $U_{ns}$:

\[ U_s - U_{ns} \geq \frac{1 - \delta}{\delta \mu \lambda} c + \frac{1 - \delta}{\delta} (x_{ns} - x_s) v. \]

In case the difference between $U_s$ and $U_{ns}$ is larger than the value that is necessary to maintain incentive compatibility, the principal can move them closer to each other without violating the promise keeping constraint. More precisely, she can decrease $U_s$ by $\epsilon$ and increase $U_{ns}$ by $\frac{\mu \lambda}{1 - \mu \lambda} \epsilon$. Such a modification is always feasible as long as $\epsilon > 0$ is chosen sufficiently small, since both of the constraints are respected. Moreover, it is preferred by the principal due to the concavity of
the value function $V$. This is because the suggested modification consists of a mean preserving contraction of the continuation utilities, hence the expectation of $V$ for the next period becomes larger. Therefore incentive constraint is always binding, without loss of generality. Solving binding incentive constraint together with the promise keeping constraint gives:

$$U_s = \frac{U_D}{\delta} + \frac{1-\delta}{\delta \mu \lambda} c - \frac{1-\delta}{\delta} x_s v,$$

(2)

$$U_{ns} = \frac{U_D}{\delta} - \frac{1-\delta}{\delta} x_{ns} v.$$

(3)

These expressions suggest that the agent is compensated for the cost of information acquisition only after the signal is realized. To see this more clearly, one can rewrite them as follows:

$$(1-\delta)x_{ns}v + \delta U_{ns} = U_D,$$

$$(1-\delta)x_s v + \delta U_s = U_D + \frac{1-\delta}{\mu \lambda} c.$$

Self-reporting increases the agent’s utility by a constant. Since the signal is verifiable, it also serves as a proof of the effort spent on self-monitoring. Therefore, the most efficient incentive provision scheme involves compensating the agent for the costs of monitoring only after the realization of the signal. Another important aspect is that the continuation utility in one contingency is independent of the approval rate in the other contingency. In other words, $U_s$ is independent of the choice of $x_{ns}$, and $U_{ns}$ is independent of $x_s$.

After figuring out the relation between approval probabilities and the continuation utilities, it is now possible to discuss the domain of value function $V_D$. By using the equations (2), and (3), one can see that the maximum value that $U_D$ can take is equal to $\delta v + \frac{(1-\delta)c}{\mu \lambda}$, which can be achieved by setting $x_s$ and $U_s$ equal to their maximum values, 1 and $v$ respectively. Therefore, the domain of $V_D$ is the interval $[0, \delta v + \frac{(1-\delta)c}{\mu \lambda}]$.

To characterize the solution of the problem $P_D$, one needs to use the equations (2), and (3) that govern the trade-off between the continuation utilities and the current approval rates for both contingencies, i.e., self-disclosure and no self-disclosure. The question is, to what extent the principal would like to use efficient approval decisions, i.e. $x_{ns} = 1$ and $x_s = 0$? It turns out that the approval decisions will be set as close as possible to the efficient ones. More precisely, $x_s = 0$ and $x_{ns} = 1$ as long as the resulting continuation utilities, i.e., $U_s$ and $U_{ns}$, stays inside the state space $[0, v]$. This requires $U_D$ to be in an intermediate range. When $U_D$ is sufficiently small, setting $x_{ns} = 1$ is not feasible, since the resulting $U_{ns}$ would be negative. For these values, the approval rate $x_{ns}$ will be chosen such that the continuation utility $U_{ns}$ becomes 0. By the same logic, for those values of $U_D$ that are sufficiently large, the approval rate $x_s$ will be chosen so that the continuation utility $U_s$ takes its largest possible value $v$. The formal statement of the lemma is given by the following lemma.
Lemma 2. There exists critical values $U = (1 - \delta)v$, and $\bar{U} = \delta v - \frac{(1 - \delta)c}{\mu_x}$, such that the solution to the problem $\mathcal{P}_D$ satisfies:

$$(x_s, x_{ns}) = \begin{cases} 
(0, \frac{U_D}{1 - \delta}v) & \text{if } U_D \leq U, \\
(0, 1) & \text{if } U_D \in (U, \bar{U}), \\
\left(\frac{U_D + \frac{(1 - \delta)c}{\mu_x} - \delta v}{v}, 1\right) & \text{if } U_D \geq \bar{U}.
\end{cases}$$

$$(U_s, U_{ns}) = \begin{cases} 
\left(\frac{U_D}{\delta} + \frac{1 - \delta}{\delta \mu_x}, 0\right) & \text{if } U_D \leq U, \\
\left(\frac{U_D}{\delta} + \frac{1 - \delta}{\delta \mu_x}, \frac{U_D - (1 - \delta)v}{\delta}\right) & \text{if } U_D \in (U, \bar{U}), \\
\left(v, \frac{U_D - (1 - \delta)v}{\delta}\right) & \text{if } U_D \geq \bar{U}.
\end{cases}$$

Proof. See appendix

Focusing on the approval rates in isolation, the principal prefers to increase $x_{ns}$ and decrease $x_s$ as much as possible due to the efficiency concerns. However, these approval rates also alter the continuation utilities, therefore there is a non-trivial tradeoff that the principal needs to take into account. As can be seen from the equations (2), and (3), a higher $x_{ns}$ requires a lower $U_{ns}$, and a smaller $x_s$ requires a higher $U_{ns}$. Lemma 2 proves that, even if there is a loss resulting from a lower $U_{ns}$, the tradeoff always favors a higher approval rate $x_{ns}$. Similarly, even if there is a loss resulting from higher $U_s$, the tradeoff always favors lower $x_s$. The lower and upper bounds of the derivative of the value function $V$, which are defined in lemma 1, are the main driving force behind this result. Precisely, the upper and lower bounds of $V'$ puts a limit on the maximum loss that can arise, and this limit is always less than the gain from employing more efficient approval decisions. Note that the $U < \bar{U}$ is always satisfied due to the restriction imposed on the discount factor $\delta$.

The problem $\mathcal{P}_N$

This problem is defined conditional on no delegation in the current period. Its solution follows from straightforward arguments. The decision is mainly about how much of the promised utility, $U_N$, to provide the agent in the current period in the form of project approval, and how much of it to leave as a continuation utility. The amount that is left as a continuation utility will be granted to the agent starting from the next period without any restriction on the delegation decision. The optimal choice of the approval rate follows from the following maximization problem.

$$\max_{x_n} (1 - \delta)x_n(v - \mu) + \delta V\left(\frac{U_N - (1 - \delta)x_nv}{\delta}\right).$$

The shape of the value function $V$ plays a crucial on the optimal choice of $x_n$ and hence on $U_n$. 

22
Let $I = [a, \bar{a}]$ be an interval where the boundaries $a$ and $\bar{a}$ satisfy:

$$a = \inf\{U \in [0, v] \mid V'(U) \leq \frac{v - \mu}{v}\},$$

$$\bar{a} = \sup\{U \in [0, v] \mid V'(U) \geq \frac{v - \mu}{v}\}.$$ 

In words, $I$ is the interval over which the derivative of the value function $V$ is equal to $\frac{v - \mu}{v}$. This interval resides in the interior of the state space $[0, v]$. This stems from the fact that, the line that connects the points $(0, V(0))$ and $(v, V(v))$ has the slope $\frac{v - \mu}{v}$, and the graph of the value function $V$ locates over this line.\(^{11}\) Then due to the concavity, $V'(0) > \frac{v - \mu}{v} > V'(v)$, hence the interval $I$ is in the interior of the state space $[0, v]$.\(^{12}\) Then, one can conclude that the optimal choice of $x_n$ satisfies the following:

$$x_n(U_N) = \begin{cases} 
0 & \text{if } U_N \leq \delta a, \\
\in (0, 1) & \text{if } U_N \in (\delta a, \delta \bar{a} + (1 - \delta)v), \\
1 & \text{if } U_N \geq \delta \bar{a} + (1 - \delta)v.
\end{cases}$$ \hspace{1cm} (4)

Intuitively, for smaller values of $U_N$, the continuation value $U_n$ will be in a range where the derivative of $V$ is sufficiently large. In this region, it is better for principal to keep the continuation utility as high as possible by setting $x_n = 0$. On the contrary, for larger values of $U_N$, the value of $V'$ becomes small, and increasing the continuation utility does not benefit the principal. Therefore, she keeps the continuation utility of the agent as small as possible by setting $x_n = 1$.

In the intermediate range of $U_N$, however, the value of $x_n$ is set so that the continuation value lies in the interval $I$, and hence the value function has a derivative equal to $\frac{v - \mu}{v}$. The principal is indifferent between marginally increasing $x_n$ and $U_n$. If the interval $I$ consists of a single point, then the optimal value of $x_n$ is also singleton. Otherwise, there is a continuum of optimal values for $x_n$ in this intermediate range.

**The problem $\mathcal{P}'$**

From now on, the conditional problems $\mathcal{P}_D$ and $\mathcal{P}_N$ will be considered together in order to characterize the solution of the unconditional problem. The task is to figure out the optimal way to decompose $U$ into $U_D$ and $U_N$ together with the the optimal choice of $\gamma$. Completing these tasks will lead to the description of the value function $V$, which is equal to the concavification of the functions $V_D$ and $V_N$.\(^{13}\)

---

\(^{11}\)The linear line can be achieved by an always feasible policy: randomizing between its extreme values. Hence it is strictly dominated, and $V$ stays on top of this line.

\(^{12}\)The interval $I$ can be a singleton.

\(^{13}\)The value function $V$ depends on $V_D$ and $V_N$, which in turn depend on $V$. In this regard, $V$, $V_D$, and $V_N$ are the solutions of a complicated fixed point problem. However, a solution for this problem exists since the existence of the solution for the problem $\mathcal{P}$ is guaranteed.
The following lemma describes the set of state variables over which the equality \( V = V_N \) is satisfied. It further indicates that the corresponding approval rates at these state variables must be either 0 or 1. In other words, having an interior probability of approval, i.e. \( x_n \in (0,1) \), after avoiding self-monitoring never happens in an optimal policy.

**Lemma 3.** There are two critical values \( 0 < U_N < \bar{U}_N < v \), such that:

i) \( V(U) = V_N(U) \) if and only if \( U \in [0,U_N] \cup [\bar{U}_N,v] \)

ii) \( V \) is linear over \([0,U_N]\) and \([\bar{U}_N,v]\).

iii) The optimal approval decision \( x_n \) satisfies:

\[
x_n(U) = \begin{cases} 
0 & \text{if } U \in [0,U_N] \\
1 & \text{if } U \in [\bar{U}_N,v] 
\end{cases}
\]

**Proof.** See appendix.

Lemma 3 points out that making an uninformed decision without monitoring can be optimal only if the promised utility is close to the boundaries of the state space. It is already known that the equality \( V = V_N \) holds at the extreme values of the state space, i.e., at 0 and \( v \). By using the comparison between the \( V_N' \) and \( V \), which can be achieved from the equation (4), one can show that the equality \( V = V_N \) can hold only over the union \([0,\delta a] \cup [\delta a + (1 - \delta)v,v]\). Then by using the expression (4), one can conclude that the approval probability \( x_n \) must be either 0 or 1 for all the promised utilities at which no-delegation is optimal. The second step of the proof shows the existence of the cutoffs \( U_N \), and \( \bar{U}_N \), and also the fact that they reside in the interior of the state space.

This result is rather intuitive. When \( U \) is sufficiently low, it is not feasible to utilize a large approval rate; similarly, when \( U \) is sufficiently large, it is not feasible to employ a large rejection rate. Therefore, the principal can get only limited benefit from the information in these state variables, and the extent of this benefit is not sufficient to compensate the cost of acquisition. For this reason, she does not ask the agent to self-monitor at these promised utilities as she also cares about the costs. Therefore, no-delegation is an optimal solution, and hence \( V = V_N \) holds.

The result is also informant about the shape of the value function over the lower and upper ends of the state space, as it points out that the value function \( V \) is linear over the intervals \([0,U_N]\), and \([\bar{U}_N,v]\). Following some simple logic, one can see that the linearity is not limited to these intervals. More precisely:

\[
V(U) = V_N(U) = \begin{cases} 
\delta V\left(\frac{U}{\bar{a}}\right) & \text{if } U \in [0,U_N] \\
(1 - \delta)(v - \mu) + \delta V\left(\frac{U-(1-\delta)v}{\bar{a}}\right) & \text{if } U \in [\bar{U}_N,v] 
\end{cases}
\]
Then, by taking the derivative of both sides, one can get the following:

\[ V'(U) = V'_N(U) = \begin{cases} V'(U) & \text{if } U \in [0, U_N] \\ V'(\frac{U - (1-\delta)v}{\delta}) & \text{if } U \in [\bar{U}_N, v] \end{cases} \]

Since \( V'(U) = V'(\frac{U}{\delta}) \) for every \( U \in [0, U_N] \), and \( V \) is concave, the slope is constant over \([0, U_N] \cup [U_N, \frac{U_N}{\delta}]\). Similarly, the slope of \( V \) is constant over \([\frac{U_N}{\delta}, (1-\delta)v, U_N] \cup [\bar{U}_N, v] \). Due to the linearity of the value function, it is without loss of generality to assume that there is a randomized delegation decision over the intervals \((U_N, U_D)\) and \((\bar{U}_D, \bar{U}_N)\).

The constant slopes, however, does not carry beyond the values \( \frac{U_N}{\delta} \), and \( \frac{U_N}{\delta} - (1-\delta)v \). As a result, no randomized delegation decision can be optimal at these values, since they do not have any neighborhood over which the slope of \( V \) stays constant. At these values, \( \gamma = 0 \) can not be optimal either, because it is already known that \( V_N < V \) for all of the state variables that are outside of \([0, U_N] \cup [\bar{U}_N, v] \) form the previous lemma. As the only remaining option, the equality \( V = V_D \) must hold at these values. To this respect, one shall define:

\[ U_D = \frac{U_N}{\delta} \]
\[ \bar{U}_D = \frac{U_N}{\delta} - (1-\delta)v \]

The values \( U_D \), and \( \bar{U}_D \) are the smallest and largest values of \( U \) satisfying \( V(U) = V_D(U) \) without loss of any generality. The interval \((U_D, \bar{U}_D)\) is the only remaining region where the delegation decision is yet to be described. However, it is natural to expect that \( V = V_D \) and hence \( \gamma = 1 \) over \((U_D, \bar{U}_D)\).

Let \( U \in (U_D, \bar{U}_D) \). From lemma 3, \( V_N < V \), and hence \( \gamma > 0 \) at this \( U \). Moreover, a randomized delegation decision cannot be optimal either. Suppose otherwise to get a contradiction, and assume that a randomization takes place between \( V_D(U_D) \) and \( V_N(U_N) \). The corresponding value of \( U_N \) must belong to either \([0, U_N]\) or \([\bar{U}_N, v]\); without loss of generality assume the precedent. Then, the principal could rather randomize between \( V_D(U_D) \) and \( V_D(U_D) \) and get strictly better off. The reason for this stems from the fact that slope of \( V \) alters at \( U_D \), hence the line connecting the values \( V_D(U_D) \) and \( V_D(U_D) \) locates on top of the line connecting \( V_D(U_D) \) and \( V_N(U_N) \). Therefore, it is optimal to set \( \gamma = 1 \) and hence \( V = V_D \) in this region.

The optimal delegation decision for each possible promised utility \( U \in [0, v] \) is now known, and summarized in the following diagram:

---

\(^{14}\)If \( V' \) were to be constant on any neighborhood of \( \frac{U}{\delta} \) and \( \frac{U_N}{\delta} - (1-\delta)v \), then the equality \( V = V_N \) would hold for a larger subset of \([0, v]\), which contradicts with the definition of \( U_N \) and \( \bar{U}_N \).
To complete the characterization of the optimal policy, one should also describe the approval decisions. Particularly, the description of $x_s$ and $x_{ns}$ over the interval $[\bar{U}_D, \bar{U}_D]$, where delegation is the optimal choice, is incomplete.

The problem $P_D$ is analyzed in isolation, and its solution is provided in lemma 2. The result asserts that the efficient decision making, i.e. $x_s = 0$, and $x_{ns} = 1$ will take place as long as the promised utility $U$ is in between $[\bar{U}, \bar{U}]$. For the rest of the state space, making the efficient decisions is not feasible. Instead, the principal employs the decisions that are closest to the efficient among the feasible ones. However, the problem $P_D$ is defined conditional on delegation in the current period regardless of its optimality. Now, the focus is on the interval $[\bar{U}_D, \bar{U}_D]$, where delegating monitoring is optimal.

The question one shall inquire at this point is, how do the values of $\bar{U}_D$ and $\bar{U}_D$ compare to the values of $\bar{U}_S$ and $\bar{U}_S$ respectively? The answer to this question is important as it indicates whether the approval decisions taking place in an optimal policy are always efficient or not. This comparison will also be informative about the long-run outcome of the optimal policy. It turns out that $U \geq \bar{U}_D$ and $\bar{U}_D = \bar{U}_D$. Therefore, there are some state variables over which the optimal policy employs an inefficient approval decision, i.e. $x_{ns} < 1$ or $x_s > 0$. The next lemma summarizes these findings:

**Lemma 4.** In an optimal policy, $\forall U \in [\bar{U}_D, \bar{U}_D]$, $V(U) = V_D(U)$, and $\gamma(U) = 1$. Moreover, $\bar{U}_D \leq U < \bar{U} \leq \bar{U}_D$.

**Proof.** See appendix.

Now, the optimal delegation and approval decisions, and the transition rule over the entire state space is known. Once the policy is initiated with an initial promised utility for the agent, the rest follows immediately. The question is how to choose this initial state variable. In other words, what is the utility level that the principal promises to the agent in the beginning? To answer this question one shall look for the utility level that is maximizing the value function $V$.

The value function $V$ increases over $[0, \bar{U}_D)$, as it has a positive constant slope; and decreases over $(\bar{U}_D, v]$. The initial utility, which is denoted by $U^*$, must be an element of the interval $[\bar{U}_D, \bar{U}_D]$. Due to the continuity, the existence of $U^*$ is guaranteed. Moreover, I show that the value function
$V$ is strictly concave over $[\bar{U}_D, \bar{U}_D]$, hence $U^*$ is unique. This result, and the earlier findings are summarized in the following theorem; which, at the same time, describes an optimal policy.

**Theorem 1.**  
The value function $V$ is strictly concave over $[\bar{U}_D, \bar{U}_D]$, and attains its maximum at $U^* \in (\bar{U}_D, \bar{U}_D)$. At $t = 0$, the agent is promised an expected utility of $U^*$, and delegated for self-monitoring.

As long as the promised utility stays in the interval $(\bar{U}_D, \bar{U}_D)$, the monitoring is delegated to agent, i.e. $\gamma = 1$. Moreover, the approval decisions and the transition rule for the state variable satisfies:

- $x_s(U) = \max\{0, \frac{U + (1-\delta)c \mu - \delta v}{(1-\delta)v}\}$, and $x_n s(U) = \min\{1, \frac{U}{(1-\delta)v}\}$, $\forall U \in [\bar{U}_D, \bar{U}_D]$.
- $U_s(U) = \min\{v, \frac{U \delta}{\delta v} \}$, and $U_n s(U) = \max\{0, \frac{U}{\delta v} - \frac{(1-\delta)v}{\delta}\}$, $\forall U \in [\bar{U}_D, \bar{U}_D]$.

When the state variable reaches to $[0, \bar{U}_D)$, the monitoring is delegated to agent with probability $\gamma = \frac{\bar{U}_D}{\bar{U}_D}$. If the realized delegation decision is:

- Delegation: $(x_s, x_n) = (0, \frac{U_D}{(1-\delta)v})$, and $(U_s, U_n) = \left(\frac{U}{\delta v} + \frac{(1-\delta)c \mu}{\delta v}, 0\right)$
- No delegation: $x_n = 0$, and $U_n = 0$

When the state variable reaches to $U \in (\bar{U}_D, v]$, the monitoring is delegated to agent with probability $\gamma = \frac{v - U}{v - \bar{U}_D}$. If the realized delegation decision is:

- Delegation: $(x_s, x_n) = \left(\frac{U + (1-\delta)c \mu - \delta v}{(1-\delta)v}, 1\right)$, and $(U_s, U_n) = \left(v, \frac{U}{\delta v} - \frac{(1-\delta)v}{\delta}\right)$
- No delegation: $x_n = 1$, and $U_n = v$

**Proof.** See appendix. □

All of the components of the above theorem, except the strict concavity, are already discussed. Strict concavity of $V$ over the interval $[U_D, \bar{U}_D]$ follows from technical arguments that are depicted in the appendix. The following figure provides an illustration of the value functions $V_D$ and $V_N$. The unconditional value function, $V$, is the concavification of them.
Figure 5: Illustration of conditional value functions $V_D$ and $V_N$. The unconditional value function $V$ is the upper envelope of $V_D$ and $V_N$. The equality $V = V_N$ holds over the intervals $[0, \bar{U}_N]$ and $[\bar{U}_N, v]$. The equality $V = V_D$ holds over the interval $[\underline{U}_D, \bar{U}_D]$. For the rest of the state space $V$ is strictly larger than $V_N$ and $V_D$.

The policy described in the theorem is an optimal policy, and it is not the only one. There is a multiplicity originating from the linearity of the value function over the lower and upper ends of the state space. One can come up with another optimal policy that employs different policy specifications over these regions, where $V$ is linear. For example, when $U \in [0, \bar{U}_N]$, it is also optimal to set $\gamma = 0$, $x_n = 0$, and $\bar{U}_n = \frac{U}{5}$. Over the intermediate region, however, the value function is strictly concave, and there is no such multiplicity.

The specific policy given in the theorem is easy to describe and has some substantiation properties. It features a two-phase structure, in which all of the periods in which self-monitoring takes place are front-loaded. The following corollary describes this in detail.

**Corollary 1 (Two-Phase Structure).**

The optimal policy depicted above consists of two consecutive phases:

i) **The initial phase**, starting at $t = 0$, is where all the information acquisition takes place. The principal delegates monitoring to the agent as long as the policy stays in this stage. Efficient decision making will be employed as long as it is available. The duration of this stage is stochastic and depends on the realized information outcomes.

ii) **The terminal phase** is no-delegation phase. Once it is reached there is no information acquisition anymore, and all the remaining approval decisions are uninformed. The projects are either always approved or always rejected depending on the realized outcomes during the initial phase.
As long as the delegation takes place, the agent is promised a higher continuation utility (in the form of future regulatory approval) each time he discloses bad news. His current project is less likely to get approved but the regulator promises more frequent approval in the future. If he does not disclose any signal, he is downgraded to a lower continuation utility. His current project has higher chances of approval, yet he will be given less frequent approvals in the future. The duration of this phase is stochastic; when it ends, the policy reaches a second phase in which there is no more self-monitoring.

If the policy reaches to the minimum promised utility the agent is black-listed and all of the projects will be directly rejected afterwards. On the contrary, if it reaches to its maximum, the agent is whitelisted and all the projects will be directly approved afterwards.

Such a structure is also observed in the papers Horner and Guo (2015), and (Li et al., 2015). The long run outcome consists of either a permanent rewarding or a permanent punishment depending on the earlier course of the relation. As it will be apparent shortly, this structure is not prevailing when the principal does not internalize the cost of information acquisition.

Efficiency aspects of the optimal policy are also significant. Likewise the papers referred in the previous paragraph, all the inefficiencies are back-loaded in the optimal policy that is defined in theorem 1. In this framework, an inefficiency arises in two consequences: first one arises if information acquisition is not requested, and the second one occurs due to inefficient use of the acquired information. Clearly, a first type inefficiency appears only in the terminal phase of the policy. On the other hand, a second type inefficiency can occur only the last period of the initial phase. When the promised utility is in the interval $[\bar{U}_D, \bar{U}]$, which is sometimes an empty set depending on the parameter values, self monitoring takes place, and the approval decision following a no self-disclosure is inefficient since $x_{ns} < 1$. Right after this event, the state variable is downgraded to its minimum, hence the ultimate stage with a permanent rejection starts. Analogous reasoning works for $(\bar{U}, \bar{U}_D]$ as well. Therefore, all the second type of inefficiencies takes place right before the ultimate stage starts. To this respect, all the inefficiencies are back-loaded.

Some of the properties mentioned above are specific to the policy depicted in the theorem. The long run outcome, on the other hand, is independent of the policy choice. It always features whitelisting or blacklisting. This stems from the fact that $U_D \leq U < \bar{U} \leq \bar{U}_D$.\textsuperscript{15} The following result indicates this observation.

**Corollary 2.** If the principal cares about the cost of monitoring, the eventual outcome features either permanent approval or permanent rejection in any optimal policy. Therefore, in the long run, the agent is either blacklisted or whitelisted with probability 1.

\textsuperscript{15}If it was $U_D > \bar{U}$ or $\bar{U} > \bar{U}_D$, then it would be possible to specify an optimal policy that does not feature blacklisting or whitelisting respectively.
Now, the principal does not internalize the agent’s cost of information acquisition. In this case, the optimal policy turns out to be remarkably different in terms of its long run properties and the efficiency aspects. These discrepancies have their origins on the principal’s richer capacity of providing incentives. From a technical point of view, principal can use monitoring requests purely on the grounds of punishing the agent, as she does not care about its cost. From time to time, the agent is asked to monitor himself, even though the resulting information will be completely ignored in the approval decision. Such a punishment scheme was feasible for the the previous case as well; however, as it also punishes the principal, it is not used in the optimal policy. The verifiability of the “bad news” plays a crucial role on this additional channel of incentive provision. Precisely, the signal provides a hard evidence for the efforts the agent spent on self-monitoring; therefore, the principal can make sure that the necessary punishment is executed.

The formal description of the principal’s problem is barely changed, with a minor modification due to principal’s shifted preferences towards monitoring costs. The objective function does not include the cost parameter, now. To recognize the differences, all of the variables are denoted with a tilde now. More precisely, the policy consists of the functions \( \gamma, \tilde{x}_n, \tilde{x}_s, \tilde{x}_{ns}, \tilde{U}_n, \tilde{U}_s, \tilde{U}_{ns} \).

### 4.1 Benchmark: Observable Effort

The initial focus is on the benchmark where the agent’s effort is observable. This will be helpful to reveal how this case is different than the previous one. The first best policy is as before, inducing the efficient outcome, i.e. monitoring and efficient decision making in each period with a stationary payment scheme. The agent’s utility form this first best outcome is the same as before, i.e. equal to \( w \); yet, the principal has a different utility since she does not care about the cost of monitoring. Her utility is denoted by \( \tilde{\pi} \), and is equal to:

\[
\tilde{\pi} = (1 - \mu \lambda) v - \mu (1 - \lambda).
\]

The benchmark policy, which is defined conditional on the agent’s ex-ante expected utility when effort is observable, is now different then the benchmark policy of the previous case. The approval decisions employed after information acquisition are not always efficient. The principal requests self-monitoring more often since she does not care about its costs.

The benchmark policy is stationary as before, hence there is no need to define the continuation utilities. The benchmark policy only consists of delegation and approval decisions, which are denoted by \( \gamma^*, \tilde{x}_{n}^*, \tilde{x}_{s}^*, \tilde{x}_{ns}^* \).

When \( U < w \), the principal, while keeping her promise, cannot employ the efficient scheme, because the agent is promised a utility that is less then the utility of the first best outcome. She can
decrease the probability of self-monitoring and employ the direct rejection with positive probability instead. This was exactly what had been done in the previous case. Alternatively, she can decrease the probability of approval without decreasing the probability of delegation. Since she does not internalize the costs of self-monitoring, this alternative method turns out to be better for the principal. First, note that the probability of wrong approval, \( \tilde{x}_{s}^{*} \), is always 0. Therefore, the principal only needs to choose \( \tilde{\gamma}^{*} \) and \( \tilde{x}_{ns}^{*} \) properly so that she keeps her promise. In doing so, her only concern is to maximize the probability of true approval, \( \tilde{x}_{ns}^{*} \), since she does not care about the cost of self-monitoring. And, the probability of true approval is maximized when the probability of delegation is at its maximum, conditional on \( U \) being less than \( w \). Therefore, the principal always delegates the monitoring to agent and approves the project with some probability chosen specifically to meet the promised utility \( U \). In other words, \( \tilde{x}_{ns}^{*} \) satisfies:

\[
(1 - \mu \lambda) \tilde{x}_{ns}^{*} v - c = U.
\]

Therefore, for every \( U < w \), the benchmark policy is characterized by: \( \tilde{\gamma}^{*} = 1 \), \( \tilde{x}_{s}^{*} = 0 \), and

\[
\tilde{x}_{ns}^{*} = \frac{U + c}{(1 - \mu \lambda)v}.
\]

When \( U > w \), the first best policy is exactly the same with the one depicted in the previous case. The promised utility of the agent is higher than the utility of the first best outcome. The optimal way to grant the promised utility to the agent requires to employ the direct approval without delegation due to the same arguments provided in the previous section. Consequently, there is a randomization between first best outcome and the direct approval without monitoring. The value function \( \tilde{V}^{*} \) over \([w, v]\), is the linear line combining the points \((\tilde{\pi}, w)\) and \((v, v - \mu)\). Therefore, the benchmark policy can be summarized as follows:

\[
\tilde{\gamma}^{*}(U) = \begin{cases} 
1 & \text{if } U \in [0, w) \\
\frac{U - w}{\mu \lambda v + c} & \text{if } U \in (w, v]
\end{cases}
\]

\[
(\tilde{x}_{s}^{*}(U), \tilde{x}_{ns}(U)) = \begin{cases} 
(0, \frac{U + c}{(1 - \mu \lambda)v}) & \text{if } U \in [0, w) \\
(0, 1) & \text{if } U \in (w, v]
\end{cases}
\]

\[
\tilde{x}_{n}^{*}(U) = \begin{cases} 
\text{Irrelevant} & \text{if } U \in [0, w) \\
1 & \text{if } U \in (w, v]
\end{cases}
\]
Therefore the corresponding value function \( \tilde{V}^* \) is given by:

\[
\tilde{V}^*(U) = \begin{cases} 
\frac{w-U}{w} v + \frac{U}{w} \tilde{\pi} & \text{if } U \in [0, w) \\
\frac{v-U}{v-w} \tilde{\pi} + \frac{U-w}{v-w} (v - \mu) & \text{if } U \in (w, v] 
\end{cases}
\]

The following figure illustrates the value function \( \tilde{V}^* \):

![Figure 6: The value function \( \tilde{V}^* \). The illustration assumes \( v < \mu \).](image)

4.2 Moral Hazard and the Optimal policy

Now, I revert back to the agency problem in which the self-monitoring efforts are not observed by the principal. The observable effort benchmark is already informative about how the optimal policy would be different from the optimal policy given in previous case. As it will be clear shortly, the corresponding optimal delegation decision is different over the lower end of the state space, and this will significantly alter the transitional dynamics as well as the long-run outcome.

The description of the problem is barely changed. The cost is not included in the objective function; however, the incentive constraint and the promise keeping constraint remain the same. Following the exposition of the previous section, the problem is decomposed into two conditional sub-problems. The corresponding value functions are now denoted by \( \tilde{V}, \tilde{V}_N \), and \( \tilde{V}_N \).

For the maximal possible promised utility of the agent, permanent approval without any monitoring is still the only feasible policy, hence the optimal one. Therefore, \( \tilde{V}(v) = \tilde{V}_N(v) \), and \( \tilde{\gamma}(v) = 0 \). On the contrary, for the minimal expected utility of the agent, permanent rejection is not the optimal policy, unlike the previous case. The principal can do better by requesting self-monitoring. To see this, note that the incentive constraint is still binding, and the equations (3) and (2) are still valid. Then by plugging \( U = 0 \) into these equations, one can get:
\[ \hat{U}_{ns}(0) = -\frac{(1 - \delta)v}{\delta} \bar{x}_{ns}(0), \]
\[ \hat{U}_s(0) = \frac{1 - \delta}{\delta \mu \lambda} c - \frac{(1 - \delta)v}{\delta} \bar{x}_s(0). \]

Clearly, \( \bar{x}_{ns} \) is equal to 0, since the continuation utility \( \hat{U}_{ns}(0) \) cannot be negative. On the other hand, after a self-disclosure, one possible policy is to set \( \bar{x}_s(0) = 0 \) and \( \hat{U}_s(0) = \frac{1 - \delta}{\delta \mu \lambda} c \).\(^{16}\) Therefore:

\[ \hat{V}_D(0) \geq \delta \mu \lambda \hat{V} \left( \frac{(1 - \delta)c}{\delta \mu \lambda} \right) > 0 \]

Therefore no delegation cannot be the optimal policy at the minimal promised utility, as it would require \( V(0) = 0 \). This situation alters the structure of the optimal policy over the lower end of the state space.

On the other hand, there is still a neighborhood of the maximal promised utility \( v \) over which no delegation is an optimal policy. In the outside of this neighborhood, \( \hat{V}_N \) is always strictly smaller than \( \hat{V} \), hence the probability of delegation is always strictly positive. In other words, not requesting self-monitoring can be an optimal policy only if the promised utility of the agent is sufficiently large. The following lemma summarizes these findings. And its proof follows the same logic used in the proof of lemma 3.

**Lemma 5.** There is a critical value \( 0 < \hat{U}_N < v \), such that:

i) \( \hat{V}(U) = \hat{V}_N(U) \) if and only if \( U \in [\hat{U}_N, v] \), moreover \( \bar{x}_n = 1 \) in this region.

ii) \( \hat{V} \) is linear over \( [\hat{U}_N, v] \).

**Proof.** See appendix. \( \square \)

Using the same arguments provided in the previous section, one can show that the linearity of \( \hat{V} \) carries over to a larger interval, which is denoted by \( [\hat{U}_D, v] \) where \( \hat{U}_D = \frac{\hat{U}_N - (1 - \delta)v}{\delta} \). The slope of \( V \) changes at \( \hat{U}_D \), and the equality \( \hat{V}(\hat{U}_D) = \hat{V}_D(\hat{U}_D) \) holds.\(^{17}\)

Over the interval \([0, \hat{U}_D]\), delegation is optimal for the principal. It is already known that this statement is valid for the boundaries of the interval, and the rest follows from the same idea used in lemma 4. The following diagram summarizes the findings so far.

\(^{16}\)Setting \( \bar{x}_s = 0 \) is optimal at \( U = 0 \) as it will become clear shortly.

\(^{17}\)The reasons for the latter are twofold. First, it is already known that \( \hat{V}(\hat{U}_D) > \hat{V}_N(\hat{U}_D) \) from the previous result, hence \( \tilde{\gamma} = 0 \) cannot be optimal. Second, there is no neighborhood of \( \hat{U}_D \) over which \( \hat{V}' \) stays constant, hence a randomization cannot be optimal either.
Due to the linearity of the value function over the interval $[\tilde{U}_D, v]$, there are many different possible ways to specify the optimal policy. The policy the paper focuses on assumes that whenever the promised utility reaches to this region, the principal randomizes between $\tilde{V}_N(v)$ and $\tilde{V}_D(\tilde{U}_D)$ with a properly chosen delegation probability. This completes the characterization of the optimal delegation decisions.

When it comes to the approval decisions, it is known that, if the agent is not asked to monitor himself, the project will be approved directly, hence $\tilde{x}_n$ is equal to 1. To achieve a complete characterization of the approval decisions, one also needs to describe the optimal values of $\tilde{x}_s$, and $\tilde{x}_{ns}$. For this, it is sufficient to focus on the interval $[0, \tilde{U}_D]$, where $\tilde{\gamma} = 1$ is the optimal decision.

Recall that the approval decision and the promised continuation utility are substitutes of each other, after both self-disclosure and no self-disclosure. More precisely, $\forall U \in [0, \tilde{U}_D]$:

\[
(1 - \delta)\tilde{x}_sv + \delta\tilde{U}_s = U + \frac{(1 - \delta)c}{\mu\lambda} \\
(1 - \delta)\tilde{x}_{ns}v + \delta\tilde{U}_{ns} = U
\]

In the previous section, it was asserted that the principal employs the efficient decision making as long as the resulting continuation utilities are feasible. Bounds of the derivative of the value function was the main driving force behind this result. It was shown that the gain from increasing $x_{ns}$ always dominates any possible loss due to the reduction on $U_{ns}$. Similarly, the gain from decreasing $x_s$ always dominates any possible loss due to the amplification on $U_s$.

Here in this case, there is an obvious lower bound for $\tilde{V}'$, which is equal to the slope of the benchmark value function $\tilde{V}^*$ over $[w, v]$. And the logic given in the previous paragraph leads to the same conclusion for $\tilde{x}_s$. On the contrary, the slope of $\tilde{V}^*$ over the region $[0, w]$ does not necessarily constitute an upper bound for $\tilde{V}'$, since the equality $\tilde{V}^*(0) = \tilde{V}(0)$ does not hold. As a result, it is not clear whether or not the principal always prefers to set $\tilde{x}_{ns}$ as high as possible.

In principle, $\tilde{V}'$ can be sufficiently large, when $U$ is sufficiently close to 0; and, hence, the gain from a higher continuation utility might dominate the gain from higher $\tilde{x}_{ns}$. Yet, it turns out that the
trade-off is still in favor of $\tilde{x}_{ns}$. In other words, as long as it is feasible, setting $\tilde{x}_{ns} = 1$ is optimal. Otherwise $\tilde{x}_{ns}$ will be set such that the continuation utility hits its minimum.

Finally, the value function $\tilde{V}$ is strictly concave over the delegation region $[0, \tilde{U}_D]$. Therefore there is a unique state variable $\tilde{U}^* \in [0, \tilde{U}_D]$ maximizing $\tilde{V}$. And the value of $\tilde{U}_D$ is larger than $\tilde{U}$, hence incorrect approval following a self-disclosure takes place with positive probability in the optimal policy. The following theorem, which constitutes the main result of this section, summarizes these findings:

**Theorem 2.**
The value function $\tilde{V}$ is strictly concave over $[0, \tilde{U}_D]$, and attains its maximum at some $\tilde{U}^* \in (0, \tilde{U}_D)$. At $t = 0$, the agent is promised the expected utility of $\tilde{U}^*$, and the monitoring is delegated to him.

As long as the state variable stays in the interval $[0, \tilde{U}_D]$, the principal requests self-monitoring with probability 1. The approval decisions and the transition rule satisfies:

- $\tilde{x}_s(U) = \max\{0, \frac{U + \frac{(1-\delta)c}{\delta}}{(1-\delta)v} \}$, and $\tilde{x}_{ns}(U) = \min\{1, \frac{U}{(1-\delta)v} \}$, $\forall U \in [0, \tilde{U}_D]$.
- $\tilde{U}_s(U) = \min\{v, \frac{U}{\delta} + \frac{(1-\delta)c}{\delta^2} \}$, and $\tilde{U}_{ns}(U) = \max\{0, \frac{U}{\delta} - \frac{(1-\delta)v}{\delta} \}$, $\forall U \in [0, \tilde{U}_D]$.

When the state variable reaches to $(\tilde{U}_D, v]$, there will be a randomized delegation decision; the monitoring is delegated to agent with probability $\tilde{\gamma} = \frac{v - U}{v - \tilde{U}_D}$. If the realized randomized decision is:

- Delegation: $(\tilde{x}_s, \tilde{x}_{ns}) = (\frac{U + \frac{(1-\delta)c}{\delta}}{(1-\delta)v} - \delta v, 1)$, and $(\tilde{U}_s, \tilde{U}_{ns}) = (v, \frac{U}{\delta} - \frac{(1-\delta)v}{\delta})$.
- No delegation: $\tilde{x}_n = 1$, and $\tilde{U}_n = v$

**Proof.** See appendix.

The dynamic properties of this policy are remarkably different now. Particularly, there is no permanent rejection state, and hence the agent is never blacklisted unlike the previous case. Instead, there is a state of probation, which occurs when the agent’s promised utility reaches to its minimum. At this state, the project will be rejected for sure, yet the agent is still asked to monitor himself. This probation state keeps repeating itself until a self-disclosure takes place. After a self-disclosure the agent is promoted to a higher promised utility, and get out of probation for one period. Leaving the probation state today does not rule out the possibility of facing it again in the future.

Eventually, a terminal phase will be reached where there is no information acquisition anymore and all the projects are directly approved. This happens when the agent’s promised utility reaches its maximum. The following corollary summarizes these implications of the optimal policy.

**Corollary 3.**
The optimal policy has a two-phase structure.
i) The initial phase is where the agent is always asked to monitor himself. Acquired information is not always used for the current approval decision. The policy reaches to a probationary state with positive probability during this phase. In this probationary state, the content of the information will be completely ignored for the current approval decision. This probation state keeps repeating itself until a self-disclosure takes place. The duration of this stage is stochastic and depends on the outcomes of the self-monitoring.

ii) The terminal phase is no-delegation phase. Once it is reached there is no information acquisition anymore, and all the remaining approval decisions are uninformed. Projects are directly approved in this phase. This phases starts when the promised utility of the agent reaches its maximum.

The distribution of inefficiencies over time constitutes an important divergence from the previous case, and from the existing papers. The inefficiencies are not entirely back-loaded now. Especially, the second type of inefficiencies arise in a stochastic nonconsecutive order until the second phase of the policy is reached. To this respect, the efficiency will be gained and lost throughout time.

The eventual outcome described above is not specific to the policy that I focus. Since the inequality $\bar{U} \geq \tilde{U}_D$ is satisfied, the permanent approval is the eventual outcome of all optimal policies. The following corollary points out this issue.

**Corollary 4.** In any optimal policy, there is a unique long-run outcome featuring permanent approval of the projects without any request of self-monitoring. Therefore, in the long-run, the agent is whitelisted with probability 1.

5 **Limited Commitment**

So far, I assumed that the principal has full commitment power in that she can commit to any incentive compatible policy. In this section I relax this assumption and consider a case where the principal’s commitment power is limited. More precisely, I assume that she cannot commit to any policy with a negative continuation value. Precisely, at the beginning of each period, she needs to have a non-negative value in expectation. She still has within period commitment power and can fulfill any specified within period randomization. In other words, the limited commitment structure I impose does not rule out the possibility of having a negative realized value for some periods. For expositional convenience, I will call the model with full commitment power as the baseline model and denote the corresponding value function by $V_B$ afterwards. Clearly, when the expected social cost of the projects is less then their economic benefits, i.e. $\mu \leq v$, an optimal policy for the baseline model is also optimal in this limited commitment environment. This stems from the fact that the continuation valuation of the principal never becomes negative in this case, i.e. $V_B(U) \geq 0$ for every $U \in [0, v]$.

On the contrary, when the expected cost is higher then the economic benefits, i.e. $\mu > v$, baseline
policy can not be an optimal policy. Principal’s continuation value becomes negative with positive probability, as in the outcome of whitelisting. She cannot promise the maximal utility to agent in this limited commitment environment. As a result, the optimal policy differs from the one of baseline model. From now on, the analysis will be based on the case $\mu > v$.

For a given history, the continuation of an optimal policy must also be optimal condition on the agent’s expected utility. Moreover it maintains a non-negative value for the principal by definition. Therefore, it is still possible to represent principal’s problem within a stationary form. The distinction is that the state space is endogenous and will be a proper subset of $[0, v]$. In describing this endogenous state space, the most prominent property one shall look for is the non-negativity of the corresponding value function for each value inside the state space. Indeed, the state space will be the maximum subset of $[0, v]$ that is satisfying this property. This is due to the fact that an optimal policy conditional on a specific state space is also a feasible policy under any larger state space. Hence enlarging state space without hurting the non-negativity can only improve the principal’s policy.

In this regard, it is not hard to conclude that the state space is an interval and contains the utility level 0. Then, it is sufficient to find out the maximal utility $U_{\text{max}}$ that the principal can promise to the agent without violating the limited commitment constraint. This would complete the characterization of the endogenous state space $[0, U_{\text{max}}]$. However, one needs to be sure about the existence of such a value. To this end, I define some auxiliary objects. First of all, for a given value $W < v$, it is known that the Bellman equation that is defined over $[0, W]$ is guaranteed to have a solution since Blackwell sufficiency conditions are satisfied. Let $V_W$ be the corresponding value function arising from the solution of Bellman equation.

**Lemma 6.** When principal has a limited commitment power, there exists a utility level $U_{\text{max}}$ which is equal to the maximum utility that the principal can promise the agent in an optimal policy. Its value is given by:

$$U_{\text{max}} = \sup \{ W \mid V_W(W) \geq 0 \}.$$  

Moreover, the value function arising from an optimal policy satisfies:

$$V = V_{U_{\text{max}}}, \quad \text{and} \quad V(U_{\text{max}}) = 0.$$  

The value of $U_{\text{max}}$ strictly decreases with $\mu$, conditional on having an optimal policy inducing self-monitoring.

**Proof.** See appendix. 

The proof of this result, first indicates that the existence of $U_{\text{max}}$ is guaranteed. Then it is argued that $V_W(W)$ is continuous, and hence the resulting value function always takes non-negative values over the interval $[0, U_{\text{max}}]$. The fact that the value of the principal at the maximal promised utility
is equal to 0 follows from the continuity of $V(W)$. Finally, the proof points out an important observation in order to conclude the monotonicity of $U_{\max}$ with respect to $\mu$. Incentive compatibility of an optimal policy is independent of the prior belief. Therefore an optimal policy at the maximal utility for some prior is also incentive compatible and brings a higher value to principal for any other smaller prior.

Then by using the arguments provided in the earlier sections one can conclude the following result.

**Theorem 3.**

When principal has a limited commitment power and the expected social cost of a project is larger than its value, whitelisting never appears in an equilibrium. Conditional on having an informative optimal policy,

- If principal internalizes the cost of self-monitoring, then the agent gets blacklisted eventually in the optimal contract.
- If principal does not internalize the cost of self-monitoring, then the optimal policy never reaches to a stable outcome and fluctuates over time.

6 Principal’s Monitoring

This section provides a more general framework than the baseline model by introducing monitoring opportunity to the principal. Instead of requesting self-monitoring from the agent, the principal can monitor the projects on her own. In particular, in each period, before making the approval decision, the principal makes a decision regarding monitoring. She either monitors the project on her own, or delegates monitoring to the agent, or completely avoids monitoring. Principal’s own monitoring will be referred as non-delegated monitoring, and self-monitoring will be referred as delegated monitoring as before.

The underlying signal structure for non-delegated monitoring is same with the signal structure of self-monitoring, in that there is still a unique verifiable signal that is revealing bad news. However, the principal assumes a higher cost $\kappa$ compared to the self-monitoring cost $c$, which still puts the agent into a superior position in information acquisition. Therefore, the principal still would like to use agent’s information in an efficient regulatory regime. Throughout this section, the principal is endowed with full commitment power, and hence the relevant state space is $[0, v]$. The following figure illustrates the corresponding stage game.
The game tree depicts the payoffs in the same way as before. As can be seen, the principal internalizes the cost of self-monitoring. This specification will be maintained throughout the entire section, and the analysis will only focus on this case. In this regard, the baseline model in this section refers to the model without the principal’s monitoring when she internalizes the cost of acquisition.

The delegation and monitoring decision of the principal is now denoted by a pair \((\gamma, \gamma_m)\). Where \(\gamma\) is the probability of delegated monitoring, and \(\gamma_m\) is the probability of non-delegated monitoring, and hence \(1 - \gamma - \gamma_m\) is the probability of no monitoring.

Clearly, optimality requires the principal to utilize the efficient approval decisions after acquiring information via non-delegated monitoring. Instead of making an inefficient decision, she can decrease the probability of non-delegated monitoring and hence the cost. This trivial observation will be further exploited to simplify the exposition in this section. The corresponding flow utilities after non-delegated monitoring are denoted by \(\hat{w}\) and \(\hat{\pi}\) for the agent and the principal respectively, and are equal to:

\[
\hat{w} = (1 - \mu \lambda)v \\
\hat{\pi} = (1 - \mu \lambda)(v - \mu_{ns}) - \kappa
\]

Since \(\kappa > c\), the first best outcome has not changed, and still involves self monitoring together with the efficient approval decisions as it was in the baseline model without principal’s monitoring.
When agent’s monitoring is observable, the principal can induce this first best outcome. Therefore, introducing monitoring opportunity to the principal does not affect the first best policy. However, the benchmark policy conditional on the agent’s utility $U \in [0, v]$ may change, despite the facts that $\kappa > c$ and self-monitoring efforts are observable. The following lemma points out to this issue.

**Lemma 7 (Benchmark policy).**

There exists a cutoff $\kappa = \frac{\mu_c}{\mu_v + c}$ such that when self-monitoring efforts are observable:

- If $\kappa > \bar{\kappa}$, the benchmark policy of the baseline model is also the optimal benchmark policy here, hence non-delegated monitoring is never utilized.

- If $\kappa \leq \bar{\kappa}$, in the benchmark policy, the principal monitors for some values of $U$. The approval decisions taking place after principal’s monitoring and agent’s self-monitoring are efficient. The delegation and monitoring decisions $(\gamma^*, \gamma_m^*)$ satisfy:

$$
(\gamma^*(U), \gamma_m^*(U)) = \begin{cases} 
(U,w), 0 & \text{if } U \leq w \\
\left( \frac{w-U}{c}, \frac{U-v}{c} \right) & \text{if } U \in (w, \hat{w}) \\
(0, \frac{v-U}{w-v}) & \text{if } U \geq \hat{w}
\end{cases}
$$

**Proof.** See appendix.

When $\kappa$ is lower than a certain threshold, for some values of the agent’s utility, the principal monitors the project on her own instead of delegating it to agent. This happens despite the fact that self-monitoring has lower costs. This result is quite intuitive. When promised utility $U$ is larger than the utility of the first best outcome, requesting self-monitoring necessitates a positive probability of direct approval. And this constitutes the intuitive reason behind eliminating self-monitoring for sufficiently large values of promised utility. The principal rather than delegating monitoring, can monitor the project on her own and increase the total probability of informed decision making. This is favorable for her if and only if $\kappa$ is not very large.

The following figure illustrates the resulting value function $V^*$ for both possible cases. As can be seen from the figure, when $\kappa < \bar{\kappa}$, the point $(\hat{\omega}, \hat{\pi})$ is not inside the convex hull of the utilities resulting from direct approval, direct rejection and first best outcome. This alters the benchmark policy.
Now, I revert the focus back to the agency problem in which the efforts spent on self monitoring are unobservable. By following the earlier exposition strategy, I will present principal’s problem within a form incorporating conditional sub-problems. There is an additional component that arises from the additional monitoring opportunity. The corresponding value function of the problem conditional on this non-delegated monitoring is denoted by $V_M$. The value function $V$ is now equal to the concavification of three conditional value functions $V_D$, $V_N$, and $V_M$ over the entire state space $[0,v]$. More precisely:

$$V(U) = \max_{U_{D,M,N}} \gamma V_D(U_D) + \gamma_m V_M(U_M) + (1 - \gamma - \gamma_m)V_N(U_N)$$

s.t. $U = \gamma U_D + \gamma_m U_M + (1 - \gamma - \gamma_m)U_N$.

The conditional value function $V_M$ can be written as:

$$V_M(U_M) = (1 - \delta)\hat{\pi} + \delta V(U_M - (1 - \delta)\hat{\mu}), \quad (5)$$

The flow payoffs for the initial period are $\hat{\pi}$ and $\hat{\mu}$ for the principal and the agent respectively. This is due to the remark we made earlier. The principal always utilizes efficient approval decisions after initiating non-delegated monitoring.

The following theorem indicates that if the principal’s cost of monitoring is sufficiently large, then she never utilizes her own monitoring in an optimal policy. In such a situation, any optimal policy of the baseline policy is an optimal one in this more general model.
Theorem 4 (Large $\kappa$).
There exists a $\bar{\kappa}$ satisfying $\bar{\kappa} > \bar{\kappa}$, such that in an optimal policy:

$$\gamma_m(U) = 0, \forall U \in [0, v] \iff \kappa > \bar{\kappa}$$

Moreover, any optimal policy of the baseline model is still an optimal policy in this case.

Proof. See appendix.

The most crucial step of the proof is to show that if non-delegated monitoring is ever used in an optimal policy, then the resulting value function must satisfy $V(\hat{w}) = \hat{\pi}$. This point can be achieved by always utilizing non-delegated monitoring together with the optimal approval decisions, and always feasible for the principal.

If the point $(\hat{w}, \hat{\pi})$ locates under the graph of the value function $V_B$, the corresponding value function of the baseline model, then non-delegated monitoring will never occur. Any optimal policy of the baseline model is still an optimal policy, and $V = V_B$. On the contrary, if it locates above the graph of $V_B$, then non-delegated monitoring takes place in the optimal policy. Because at $U = \hat{w}$, the principal can always do better than the baseline optimal policy by utilizing non-delegated monitoring. On account of this, the cutoff $\bar{\kappa}$ is the value of $\kappa$ for which the corresponding pair $(\hat{w}, \hat{\pi})$ locates exactly on the graph of $V_B$. In other words:

$$\bar{\kappa} = (1 - \mu \lambda) (v - \mu_{ns}) - V_B(\hat{w})$$

From now on, rather than analyzing the entire range of $\kappa$ that generates an optimal policy incorporating non-delegated monitoring, I focus on a particular subset of it, i.e $\kappa < \bar{\kappa}$. When $\kappa$ is in this range, the value function satisfies $V = V^*$ over $U \in [\hat{w}, v]$. This stems form the fact that the benchmark value function $V^*$ does not employ delegated monitoring when $u \in [\hat{w}, v]$, and hence the same value can be induced by the principal in the agency problem. For each $U \in [\hat{w}, v]$, the optimal policy is stable without loss of generality, and features a randomization between non-delegated monitoring and direct approval every period with the same probability. The next lemma focuses on the rest of the state space, and indicates that self-monitoring is optimal over an intermediate region inside the interval $[0, \hat{w}]$.

Lemma 8. When $\kappa \leq \bar{\kappa}$, and $\delta$ is large enough, there exists $0 < U_D < \bar{U}_D < \hat{w}$, such that:

$$V(U) = V_D(U) \iff U \in [U_D, \bar{U}_D]$$

Moreover, the value function $V$ is linear over $[0, U_D]$ and $[\bar{U}_D, \hat{w}]$.

Proof. See appendix.
Then by using these findings, one can define an optimal policy conditional on promised utility of the agent as follows:

**Definition (An optimal Policy).** When \( \kappa < \bar{\kappa} \), and \( \delta \) is large enough, the following constitutes an optimal policy conditional on the promised utility \( U \in [0, v] \).

- For \( U = 0 \), permanent rejection.
- For \( U \in (0, U_D) \), randomize between \( V(0) \) and \( V(U_D) \).
- For \( U \in [U_D, \bar{U}_D] \), delegate monitoring and set
  \[
  x_s(U) = 0, \quad x_{ns}(U) = \min\{1, \frac{U}{(1-\delta)v}\}, \quad U_s(U) = \frac{U}{\delta} + \frac{(1-\delta)c}{\delta\mu \lambda}, \quad U_{ns}(U) = \max\{0, \frac{U}{\delta} - \frac{(1-\delta)c}{\delta}\}.
  \]
- For \( U \in (\bar{U}_D, \hat{w}) \), randomize between \( V(\bar{U}_D) \) and \( V(\hat{w}) \).
- For \( U \in [\hat{w}, v] \), randomize between non-delegated monitoring and direct approval every period, where the probability of the precedent is \( \gamma_m = \frac{v-U}{v-\hat{w}} \).

As can be seen in this description, it is optimal to set \( x_s \) and \( x_{ns} \) as close as possible their efficient level. This follows from the same arguments provided before and hence will not be repeated here. An important issue to point out is that the probability of wrong approval \( x_s \) is always 0. The value of \( \bar{U}_D \) is not high enough to require wrong approval. On the other hand, \( U_D < \bar{U} \), therefore, the probability of correct approval is sometimes less than 1.

The above description, specifies a set of choices for monitoring and approval decisions conditional on each possible promised utility level. However it does not specify an initial promise and a long run outcome. One possibility is that \( \hat{w} \) is the maximizer of the value function \( V \), and hence the optimal policy requires principal to monitor in every period on her own and utilize efficient approval decisions. For this to be true, the constant slope of \( V \) over the interval \([\bar{U}_D, \hat{w}] \) must be positive. Otherwise, there is another utility level \( U \) that is smaller than \( \hat{w} \) maximizing the value function \( V \). For this to be true, the constant slope of \( V \) over the interval \([\bar{U}_D, \hat{w}] \) must be negative. The next theorem, offers a sufficient condition for the latter. So that the policy starts with delegated monitoring, the transition is governed as before, and eventually one of the absorbing states will be reached.

**Theorem 5.**

If \( \frac{\pi}{\kappa-\bar{\kappa}} < \frac{(\frac{1}{\lambda}(1-\mu\lambda))\mu\lambda v + (1-\delta\mu\lambda)c}{\mu\lambda v + (1-\delta\mu\lambda)c} \), \( \delta \) is large enough, and \( \kappa < \bar{\kappa} \), then there is a utility level \( U^* \in [U_D, \bar{U}_D] \) maximizing \( V \). At \( t = 0 \), the principal promises \( U^* \) to the agent and delegates monitoring to him. The policy proceeds as it is defined above, and eventually reaches to a stable outcome that features either blacklisting the agent, or non-delegated monitoring, or a randomization between non-delegated monitoring and direct approval.

**Proof.** See appendix.

The proof just shows that when the conditions provided in the theorem are satisfied, it is possible
to have a $U \in [0, \hat{w}]$ satisfying $V(U) > \hat{\pi}$. This requires that initial period of the optimal policy starts with the delegated monitoring, and proceeds. The proof also shows that it is possible to reach a promised utility that is higher than $\hat{w}$, because $U_s(\hat{U_D}) \geq \hat{w}$. Therefore en eventual outcome in which the principal always randomizes between non-delegated monitoring and direct approval is possible.

7 Concluding Remarks

This paper studies a regulatory system that incorporates self-monitoring. More precisely, it explores the behavior of regulators in environments where there is a significant uncertainty about the activities that the regulated agent carries on. The regulator, in an efficient regulatory regime, would like to use the information of the agent, who is superior in acquiring information. In order to incentivize the agent to acquire unfavorable information about his own activities, the regulator uses continuation values arising from future regulatory behavior as an incentive device.

I show that, when the regulator has full commitment power, self-monitoring can only be sustained in an initial phase of the optimal policy. During this phase, the agent is promised a higher continuation utility (in the form of future regulatory approval) each time he discloses “bad news”; otherwise, he is downgraded to a lower continuation utility in order to encourage him to acquire information. The eventual outcome crucially depends on the regulator’s preferences over the cost of self-monitoring. If she internalizes this cost, both whitelisting and blacklisting are possible long-run outcomes; otherwise, whitelisting is the only long run outcome.

When the regulator has limited commitment power in that she can only commit to policies with non-negative continuation values, the results are remarkably different. When the expected social costs of the projects are larger than their economic yields, the policy does not feature whitelisting anymore. Furthermore, it is possible to sustain self-monitoring over the long-run in this case.

References


A Appendix

Proof of Lemma 1.
The principal can always randomize over different values of $U$ with a restriction that the expectation resulting from the randomization is exactly equal to the promised utility. This immediately requires the concavity of the value function $V$, and then the almost everywhere differentiability directly follows.\footnote{One can also apply the result of Benveniste and Scheinkman (1979), to show the differentiability of the value function. To see how their result can be applied in this context see for example (Horner and Guo, 2015).}

To see the bounds indicated in expression (1), first note that $V(0) = V^*(0) = 0$ and $V(v) = V^*(v) = v - \mu$. We also know that $V(U) \leq V^*(U)$ for every $U \in [0,v]$, since the value function $V^*$ is defined by an optimization problem with a smaller set of constraints compared to the one of the value function $V$. Moreover, $V'$ is decreasing over the state space as it is a concave function. Therefore its derivative cannot be larger than the slope of $V^*$ at 0 and cannot be smaller than the slope of $V^*$ at $v$. But we know that the value function $V^*$ is piece-wise linear, and its slope is $1 - \frac{\mu\lambda}{(1-\mu\lambda)v+c}$ over the interval $[0, (1-\mu\lambda)v-c)$, and equal to $1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)v-c}$ over the interval $((1-\mu\lambda)v-c, v]$. This concludes our proof. \hfill $\square$

Proof of Lemma 2.
The continuation utilities are given in the equations 2 and 3. By utilizing these expressions, it will be shown that the approval probabilities $x_s$ and $x_{ns}$ will be set equal to their efficient levels, 0 and 1 respectively, as long as this does not violate the promise keeping constraint.

To this end, assume that $x_{ns}(U_D) < 1$ for some value of $U_D$. Then consider the following deviation that is acquired by increasing $x_{ns}$ by $\epsilon$ and decreasing $U_{ns}$ by $\frac{(1-\delta)v}{\delta}$ for a sufficiently small $\epsilon$. Note that this change is respecting the constraint PK. In consequence, the principal’s utility increases by:

$$\Delta = (1 - \mu\lambda) \left[ (1 - \delta)\epsilon(v - \frac{\mu(1-\lambda)}{1-\mu\lambda}) + \delta \left( V(U_{ns}) - V(U_{ns} - \frac{(1-\delta)v}{\delta}) \right) \right]$$

Moreover, from lemma 1, it is known that $V' \leq 1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)v-c}$. Then by using the fundamental theorem of calculus, one can get:\footnote{The concavity of the value function $V$ is sufficient for utilizing the fundamental theorem.}

$$\Delta \geq (1 - \delta)\epsilon(1 - \mu\lambda) \left[ v - \frac{\mu(1-\lambda)}{1-\mu\lambda}v - \frac{\mu(1-\lambda)v}{(1-\mu\lambda)v-c} \right] > 0.$$  

Therefore this deviation, in case if it is feasible, strictly benefits the principal. The suggested deviation would not be feasible if $U_{ns} = 0$, which occurs when $U_D < (1 - \delta)v$. Therefore, when
$U < U = (1 - \delta)v$ the approval rate $x_{ns}$ will be set such that the continuation utility $U_{ns} = 0$, otherwise $x_{ns} = 1$. This proves the lemma for $x_{ns}$ and $U_{ns}$.

Now suppose that $x_s > 0$ for some value of $U$. Then consider decreasing $x_s$ by $\epsilon$ and hence increasing $U_s$ by $\frac{(1-\delta)v}{\delta}$ for a sufficiently small $\epsilon$. This results with a change $\Delta$ that is equal to:

$$\Delta = \mu \lambda \left[ -(1-\delta) \epsilon (v-1) + \delta \left( V(U_s) - V(U_s + \frac{(1-\delta)v}{\delta}) \right) \right]$$

Again from lemma 1, we know that $V'(0) = V_N'(0)$ and $V(v) = V_N(v)$. Moreover, from the optimal choice of $x_n$ which is indicated in (4), it is known that:

$$V_N'(U) \leq V'(U) \text{ if } U \leq \delta a$$
$$V_N'(U) < V'(U) \text{ if } U \in (\delta a, a)$$
$$V_N'(U) = V'(U) \text{ if } U \in [a, \bar{a}]$$
$$V_N'(U) > V'(U) \text{ if } U \in (\bar{a}, \delta a + (1-\delta)v)$$
$$V_N'(U) \geq V'(U) \text{ if } U \in [\delta a + (1-\delta)v, v]$$

Then by the concavity, the comparison of $V'$ and $V_N'$ satisfies:

$$V_N'(U) \leq V'(U) \text{ if } U \in [0, \delta a]$$
$$V_N'(U) < V'(U) \text{ if } U \in (\delta a, a)$$
$$V_N'(U) = V'(U) \text{ if } U \in [a, \bar{a}]$$
$$V_N'(U) > V'(U) \text{ if } U \in (\bar{a}, \delta a + (1-\delta)v)$$
$$V_N'(U) \geq V'(U) \text{ if } U \in [\delta a + (1-\delta)v, v]$$

But this requires that:

$$V_N(U) \leq V(U) \text{ if } U \in [0, \delta a]$$
$$V_N(U) < V(U) \text{ if } U \in (\delta a, \delta a + (1-\delta)v)$$
$$V_N(U) \leq V(U) \text{ if } U \in [\delta a + (1-\delta)v, v]$$
Otherwise it would not be possible to have $V(0) = V_N(0)$ and $V(v) = V_N(v)$. Therefore, the equality $V = V_N$ can only arise in $[0, \delta a] \cup [\delta a + (1 - \delta)v, v]$, and this in turn implies that $x_n \in \{0, 1\}$ in the optimal policy.

Focusing on the interval $[0, \delta a]$ first, from the solution of the problem $P_N$, it is known that $x_n = 0$ and $V'_N \leq V'$. Then due to the fact $V(0) = V_N(0)$, it immediately follows that there must exists a critical value $U_N$, satisfying the following:

$$U_N = \sup\{U \in [0, v] \mid V = V_N \text{ over } [0, U]\}$$

Since $V_N < V$ over the interval $(\delta a, \delta a + (1 - \delta)v)$, we must have $U_N \leq \delta a$. In principle $U_N$ might be 0 so that $[0, U_N] = \{0\}$. However, we will show that $U_N > 0$.

Suppose $U_N = 0$ to get a contradiction, and hence $V_N(U) < V(U)$ for each $U$ sufficiently close to 0. However, when $U$ is sufficiently close to 0, the solution to the problem $P_N$ implies that $x_s = 0$, $x_{ns} = \frac{U}{(1-\delta)v}$, $U_{ns} = 0$, and $U_s = \frac{U}{\delta} + \frac{(1-\delta)c}{\delta \mu \lambda}$, therefore:

$$V_D(U) = (1 - \delta) \left[ (1 - \mu \lambda) \frac{U}{(1-\delta)v} (v - \frac{\mu(1-\lambda)}{1-\lambda}) - c \right] + \delta \mu \lambda V\left(\frac{U}{\delta} + \frac{(1-\delta)c}{\delta \mu \lambda}\right)$$

Then from the upper bound of $V'$, it is known that:

$$V\left(\frac{U}{\delta} + \frac{(1-\delta)c}{\delta \mu \lambda}\right) \leq \left(\frac{U}{\delta} + \frac{(1-\delta)c}{\delta \mu \lambda}\right) \left(1 - \frac{\mu(1-\lambda)}{(1-\lambda)v - c}\right).$$

Thus, $V_D(U) < 0$ for sufficiently small $U$. Which suggests that $V_D(U) < V(U)$, as well as $V_N(U) < V(U)$, when $U$ goes to 0.

Consequently, there must be a randomization, i.e. $\gamma \in (0, 1)$, for the values of $U$ that are close enough to 0. This in turn, requires that $V$ is linear in this region, with a slope $m$. Then in this region, $V_N(U) = \delta V\left(\frac{U}{\delta}\right) = \delta m \frac{U}{\delta} = V(U)$, and this gives us a contradiction as we assumed that $V_N < V$. Therefore $U_N > 0$.

To show the linearity of $V$, and hence $V_N$, over $[0, U_N]$, note that $V'(U) = V'_N(U) = V'\left(\frac{U}{\delta}\right)$ in this region. However, since $V$ is concave, the derivative is weakly decreasing. Hence the derivative must be constant, i.e. the value function is linear.

The proof for the other end of the state space follows from analogous arguments.

**Proof of Lemma 4:**

Refer to the main text for the proof of the first part, i.e. $V(U) = V_D(U)$, and $\gamma(U) = 1$ for every $U \in [U_D, \bar{U}_D]$.

To show the second part of the result, i.e the inequality $U_D \leq U < \bar{U} \leq \bar{U}_D$, I focus on the comparison between $U_D$ and $U$. To get a contradiction, suppose that the inequality $U_D > U$ holds,
and hence:

\[ V(U_D) = (1 - \delta) \left[ (1 - \mu \lambda)v - \mu(1 - \lambda) - c \right] + \delta \left[ \mu \lambda V \left( \frac{U_D}{\delta} \right) + \frac{(1-\delta)c}{\delta \mu \lambda} \right] + (1 - \mu \lambda) V \left( \frac{U_D - (1-\delta)v}{\delta} \right) \]

\[ = m U_D \]

where \( m \) is the constant slope of \( V \) over \([0, U_D]\). Then, consider some \( U = U_D - \epsilon \) for a sufficiently small \( \epsilon > 0 \) satisfying \( U > \bar{U} \). Clearly:

\[ V_D(U) = (1 - \delta) \left[ (1 - \mu \lambda)v - \mu(1 - \lambda) - c \right] + \delta \left[ \mu \lambda V \left( \frac{U}{\delta} \right) + \frac{(1-\delta)c}{\delta \mu \lambda} \right] + (1 - \mu \lambda) V \left( \frac{U - (1-\delta)v}{\delta} \right) \]

\[ > V(U_D) - m \epsilon. \]

The strict inequality stems from the fact that:

\[ V \left( \frac{U_D}{\delta} + \frac{(1-\delta)c}{\delta \mu \lambda} \right) - V \left( \frac{U}{\delta} + \frac{(1-\delta)c}{\delta \mu \lambda} \right) < \frac{m \epsilon}{\delta} \]

\[ V \left( \frac{U_D - (1-\delta)v}{\delta} \right) - V \left( \frac{U - (1-\delta)v}{\delta} \right) = \frac{m \epsilon}{\delta}. \]

But this consists of a contradiction, since

\[ V_D(U) > V(U_D) - m \epsilon = m(U_D - \epsilon) = V(U) \]

Therefore the inequality \( \bar{U}_D \leq \bar{U} \) holds. By analogous arguments one can also show that \( \bar{U}_D \geq \bar{U} \), hence the proof is complete.

\[ \square \]

**Proof of Theorem 1**.

To complete the proof, one just needs to show that the value function \( V \) is strictly concave over the interval \([U_D, \bar{U}_D]\).

It is already shown that \( U_D \leq U < \bar{U} \leq \bar{U}_D \). Moreover, the description of the optimal approval decisions \((x_s, x_{ns})\), and the continuation utilities \((U_s, U_{ns})\) are provided in lemma 2. Then, by using the fact that \( V = V_D \) over \([U_D, \bar{U}_D]\), one can get:

\[ V(U) = \begin{cases} 
(1 - \delta)(1 - \mu \lambda) \frac{U}{1-\delta v}(v - \mu_{ns}) + \delta \mu \lambda V(U_s) & \text{if } U_D \in [U_D, \bar{U}_D) \\
(1 - \delta)\pi + \delta [(1 - \mu \lambda)V(U_{ns}) + \mu \lambda V(U_s)] & \text{if } U_D \in [U, \bar{U}] \\
(1 - \delta)[\pi + \mu \lambda(\frac{U + (1-\delta)c}{\delta \mu \lambda - \delta v})(v - 1)] + \delta(1 - \mu \lambda)V(U_{ns}) & \text{if } U_D \in (\bar{U}, \bar{U}_D] 
\end{cases} \]

Where the continuation utilities are given by:

\[ U_s = \frac{U}{\delta} + \frac{(1 - \delta)c}{\delta \mu \lambda} \quad \text{and} \quad U_{ns} = \frac{U}{\delta} - \frac{(1 - \delta)v}{\delta}. \]
For the first case, due to the concavity of \( V \), it must be the subset of \( [\overline{\lambda}_V] \). 

First, note that \( s \) must be the subset of \( [\overline{s}_V] \). Moreover, either \( I\subset\overline{\lambda}_V \) or \( I\subset\overline{s}_V \). An analogous contradiction carries over the second case.

For the last case, \( V' \) cannot stay constant in any neighborhoods of \( U \) and \( \tilde{U} \). As a result, there are three possible cases: i) \( I\subset[\overline{U},U) \), ii) \( I\subset(\overline{\tilde{U}},\overline{U}) \), iii) \( I\subset[U,\overline{\tilde{U}}] \).

For the first case, due to the concavity of \( V \), \( V' \) must be constant over \( U_s|_I = (\inf(I),\sup(I)) \). Moreover, due to the lower bound on \( \delta \), \( U_s(\inf(I)) < \overline{U} \), hence \( U_s|_I \subset [U_D,\overline{U}_D] \). This gives an immediate contradiction with the definition of \( I \), since the length of \( U_s|_I \) is larger than the length of \( I \). An analogous contradiction carries over the second case.

For the last case, \( V' \) must be constant along the intervals \( U_s|_I \) and \( U_{ns}|_I \), due to the concavity of \( V \). Moreover, either \( U_s(\inf(I)) < \overline{U} \) or \( U_{ns}(\sup(I)) > \overline{\tilde{U}} \) must be correct, because both of them cannot be wrong at the same time due to the lower bound on \( \delta \). Therefore, either \( U_s|_I \) or \( U_{ns}|_I \) must be the subset of \( [U,\overline{\tilde{U}}] \subset [U_D,\overline{U}_D] \), which contradicts with the definition of \( I \), because both of these intervals have a larger length than \( I \).

\( \square \)

**Proof of Lemma 5:**

Define:

\[
b = \inf\{U \in [0,v] \mid \hat{V}'(U) \leq \frac{v-\mu}{v}\},
\]

\[
\bar{b} = \sup\{U \in [0,v] \mid \hat{V}'(U) \geq \frac{v-\mu}{v}\}.
\]

Then the optimal approval decision conditional on no self-monitoring satisfies:

\[
\tilde{x}_n(U_N) = \begin{cases} 
0 & \text{if } U_N \leq \delta \bar{b} \\
\in (0,1) & \text{if } U_N \in (\delta \bar{b}, \delta \bar{b} + (1-\delta)v) \\
1 & \text{if } U_N \geq \delta \bar{b} + (1-\delta)v
\end{cases}.
\]

(6)

Therefore:

\[
\hat{V}'_N(U_N) = \begin{cases} 
\hat{V}'(\frac{U}{\delta}) & \text{if } U_N \leq \delta \bar{b} \\
\frac{v-\mu}{v} & \text{if } U_N \in (\delta \bar{b}, \delta \bar{b} + (1-\delta)v) \\
\hat{V}'(\frac{U-(1-\delta)v}{\delta}) & \text{if } U_N \geq \delta \bar{b} + (1-\delta)v
\end{cases}
\]

\[\text{See lemma 1.}\]
Clearly \( \tilde{V}_N(v) = \tilde{V}(v) \), since there is only one possible way to provide the maximal utility to the agent. However, unlike the previous case, the equality \( \tilde{V}_N(0) = \tilde{V}(0) \) does not hold. To see this, first observe that \( \tilde{V}_N(0) = 0 \). Then, in order to examine the value of \( \tilde{V}_D(0) \), by using the fact that the incentive and the promise keeping constraints are binding, we get:

\[
\tilde{U}_{ns}(0) = -(1 - \delta)v - \frac{1 - \delta}{\delta} \tilde{x}_{ns}(0),
\]

\[
\tilde{U}_s(0) = \frac{1 - \delta}{\delta \mu \lambda} - \frac{(1 - \delta)v}{\delta} \tilde{x}_s(0).
\]

Obviously, \( \tilde{x}_{ns} = 0 \) since the continuation utility \( \tilde{U}_{ns}(0) \) cannot be negative. Moreover, \( \tilde{U}_s(0) = \frac{1 - \delta}{\delta \mu \lambda} \), since the optimal choice of \( \tilde{x}_s(0) \) is 0.\(^{21}\) As a result, \( \tilde{V}_D(0) = \delta \tilde{V}(\frac{1 - \delta \mu \lambda}{\delta}) \), which is strictly positive. Therefore we must have \( \tilde{V}_N < \tilde{V}(0) = \tilde{V}_D(0) \). Then by using this and fundamental theorem of calculus together with the following

\[
\tilde{V}_N'(U) \leq \tilde{V}'(U) \text{ if } U \in [0, \delta b]
\]

\[
\tilde{V}_N'(U) < \tilde{V}'(U) \text{ if } U \in (\delta b, b)
\]

\[
\tilde{V}_N'(U) = \tilde{V}'(U) \text{ if } U \in [b, \bar{b}]
\]

\[
\tilde{V}_N'(U) > \tilde{V}'(U) \text{ if } U \in (\bar{b}, \delta \bar{b} + (1 - \delta)v)
\]

\[
\tilde{V}_N'(U) \geq \tilde{V}'(U) \text{ if } U \in [\delta \bar{b} + (1 - \delta)v, v]
\]

we can conclude that there exists \( \tilde{U}_N \) such that \( \tilde{V}(U) = \tilde{V}_N(U) \) if and only if \( U \in [\tilde{U}_N, v] \). In addition, the facts that \( \tilde{U}_N \) is strictly smaller than \( v \), and \( \tilde{V} \) is linear over \([\tilde{U}_N - \frac{(1 - \delta)\mu \lambda}{\delta}, v]\) follow from exactly the same arguments provided in the proof of lemma 3. On the other hand, since \( \tilde{U}_N \in [\delta \bar{b} + (1 - \delta)v, v] \), it immediately follows that \( \tilde{x}_n = 1 \) over \([\tilde{U}_N, v] \).

Proof of Theorem 2:

Initially note that it is already known that the equality \( \tilde{V} = \tilde{V}_D \) holds at 0 and \( \tilde{U}_D \). Then the same logic that is used in lemma 3, it immediately follows that the equality \( \tilde{V} = \tilde{V}_D \) holds for each \( U \in [0, \tilde{U}_D] \).

When it comes to the approval decisions, the case for \( \tilde{x}_s \) has already been discussed. Here the focus will be on the contingency of no self-reporting, i.e the choice variables \( \tilde{x}_{ns} \), and \( \tilde{U}_{ns} \). Since these variables are isolated from the other contingency, their choice satisfy the following local problem:

\[
\max_{\tilde{x}_{ns}} (1 - \delta)\tilde{x}_{ns}(v - \frac{\mu(1 - \lambda)}{1 - \mu \lambda} + \delta \tilde{V}(\tilde{U}_{ns})
\]

s.t. \((1 - \delta)v \tilde{x}_{ns} + \delta \tilde{U}_{ns} = U
\]

\(^{21}\) Refer to earlier discussions to see why it is optimal to set \( \tilde{x}_s \) equal to 0.
From the constraint, it is possible to substitute between \( \bar{x}_{ns} \) and \( \bar{U}_{ns} \) at a rate 1 to \( \frac{(1-\delta)\nu}{\delta} \); moreover their marginal returns for the principal are \( (1-\delta)(v-\frac{\mu(1-\lambda)}{1-\mu\lambda+v}) \), and \( \delta V'(U_{ns}) \) respectively. Therefore, showing that \( \bar{V}' < 1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda+v)} \), would be sufficient to complete the proof. Suppose not to get a contradiction, and define:

\[
d = \inf\{U \in [0,v] \mid \bar{V}'(U) \leq 1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)}\nu\},
\]

\[
\bar{d} = \sup\{U \in [0,v] \mid \bar{V}'(U) \geq 1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)}\nu\}.
\]

From the hypothesis it is known that \( \bar{d} > 0 \). Moreover, \( \bar{x}_{ns}(U) = 0 \), \( \bar{U}_{ns}(U) = \frac{U}{\nu} \), and \( \bar{x}_s(U) = \frac{U}{\nu} + \frac{(1-\delta)c}{\mu\lambda} \), \( \forall U \in [0,\bar{d}] \). This would require that \( \bar{V}' = (1-\mu\lambda)\bar{V}'(\bar{U}_{ns}) + \mu\lambda\bar{V}'(\bar{U}_s) \), i.e. the derivative of the value function is equal to the the expectation of the derivative over the continuation values. This in turn requires \( \bar{V}' \) to be constant over \( [0,d+(1-\delta)c]\frac{\nu}{\mu\lambda} \), due to the concavity together with the fact that \( \bar{U}_s(U), \bar{U}_{ns}(U) > U \), for every \( U \in [0,\bar{d}] \). Therefore \( \bar{d} \) must be equal to 0.

Now it is known that, \( \bar{V}'(U) = 1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)}\nu \), for every \( U \in [0,\bar{d}] \). Then again from the same logic, this constant slope must carry over to a larger region, and hence constitutes a contradiction with the definition of \( \bar{d} \), as it requires that \( \bar{V}'(\bar{U}_s(U)) = 1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)}\nu \), \( \bar{x}_s = 0 \), \( \forall U \in [0,\bar{d}] \). This stems from the fact that \( \bar{V}' = (1-\mu\lambda)\left(1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)}\nu\right) + \mu\lambda\bar{V}'(\bar{U}_s) \), as the principal is indifferent between the marginal increase on \( \bar{x}_{ns} \) and \( \bar{U}_{ns} \) in this region. As a result we have \( \bar{V}' < 1 - \frac{\mu(1-\lambda)}{(1-\mu\lambda)}\nu \), and the result immediately follows.

Finally the strict concavity, and the inequality \( \bar{U}_D > \bar{U} \) follows from the identical arguments presented in the previous case, hence not repeated here.

\[ \square \]

**Proof of Lemma 6.**

First of all, \( U_{max} = \sup\{W \mid V_W(W) \geq 0\} \) exists since \( W \) takes its values from a bounded interval. Therefore, to complete the proof, one just needs to show that \( V_{U_{max}}(U) \geq 0 \) for each \( U \in [0,U_{max}] \). Suppose not to get a contradiction, and assume that \( V_{U_{max}}(U_{max}) < 0 \). Then the definition of \( U_{max} \) implies that \( \forall \epsilon > 0, \exists U_{\epsilon} \in [U_{max} - \epsilon, U_{max}] \), such that \( V_{U_{\epsilon}}(U_{\epsilon}) \geq 0 \). However, it is known that \( V_W(U) \) is continuous on \( U \) and \( W \), therefore \( V_W(W) \) must be continuous in \( W \). Therefore \( \exists^{*} \epsilon > 0 \) such that \( \forall W \in [U_{max} - \epsilon^{*}, U_{max}], V_W(W) < 0 \). Which is a contradiction. Therefore \( V_{U_{max}}(U_{max}) \geq 0 \).

To show that the last inequality holds with equality, suppose that \( V_{U_{max}}(U_{max}) > 0 \). Then the principal could grant the agent a utility level that is slightly higher than \( U_{max} \) due to the continuity of \( V_W(w) \), and this contradicts with the definition of \( U_{max} \).

Finally, in order to show that increasing \( \mu \) strictly decreases \( U_{max} \), let \( \mu \) and \( \mu' \) are two values with \( \mu > \mu' \). Note that, an incentive compatible policy when the prior is equal to \( \mu \) is also incentive
compatible when prior is $\mu'$ and vice versa. This stems from the fact that the agent does not care about the type of the project. Let $U_{\text{max}}$ be the corresponding maximal state variable when prior is $\mu$. Take an optimal policy when the agent is given $U_{\text{max}}$. It is clear that this policy would bring a strictly positive value to principal when the prior is $\mu'$, therefore the corresponding maximal state variable, $U'_{\text{max}}$ must be strictly larger than $U_{\text{max}}$.

\[\square\]

**Proof of Lemma 7.**

Since there is no incentive constraint, the value function of this benchmark is equal to the convex hull of the utilities arising from degenerate outcomes. When principal monitors on her own and employs efficient approval decisions, the resulting utilities are $\tilde{\pi} = (1 - \mu\lambda)(v - \mu_{\text{ns}}) - \kappa$ and $\tilde{w} = (1 - \mu\lambda)v$.

- If $\kappa > \bar{\kappa}$, the point $(\tilde{w}, \tilde{\pi})$ is inside the convex hull of the points $(0, 0)$, $(v, v - \mu)$, and $(w, \pi)$. Therefore the benchmark policy is the same as before.

- If $\kappa \leq \bar{\kappa}$, then the point $(\tilde{w}, \tilde{\pi})$ is part of the convex hull, and hence principal’s monitoring is part of the benchmark policy. The choices of $\gamma^*$ and $\gamma^*_m$ are coming from the weights used in the convex hull.

\[\square\]

**Proof of Theorem 4.**

Note that one can restrict the domain of the value function $V_M$ to the interval $[(1 - \delta)\tilde{w}, \delta v + (1 - \delta)\tilde{w}]$. This is because of the fact that the approval decisions following the principal’s monitoring are always efficient. Over this domain, it is already known that:

$$V_M(U) = (1 - \delta)\tilde{\pi} + \delta V\left(U - (1 - \delta)\tilde{w}\right).$$

Then by taking the derivative of both sides, one can get:

$$V'_M(U) = V'(U - \frac{(1 - \delta)\tilde{w}}{\delta}).$$

Therefore:

$$V'_M(U) \geq V'(U) \quad \text{if} \quad U \leq \tilde{w}$$

$$V'_M(U) \leq V'(U) \quad \text{if} \quad U \geq \tilde{w}$$

Now, assume that the principal’s monitoring is used in an optimal policy. Then there must exist a $U \in [0, v]$ satisfying $V(U) = V_M(U)$. Then by using the comparison between $V'_M$ and $V'$ depicted above, one can further conclude that:

- If $U \leq \tilde{w}$, then by concavity of $V$ one can conclude that $V = V_M$ over $[U, \tilde{w}]$.
• If $U \geq \hat{w}$, then by concavity of $V$ one can conclude that $V = V_M$ over $[\hat{w}, U]$.

To this respect, if principal’s monitoring is utilized in an optimal policy, then $V(\hat{w}) = V_M(\hat{w})$ and this requires:

$$V(\hat{w}) = (1 - \delta)\hat{\pi} + \delta V(\hat{w})$$

Hence $V(\hat{w}) = \hat{\pi}$, and the rest immediately follows from the arguments provided in the main text.

Proof of Lemma 8.

The focus is on the interval $[0, \hat{w}]$. The first step is to show that the delegated monitoring must be optimal for some utility level $U$ in this interval. If not, then the value function $V$ would be equal to the line that is connecting the points $(0, 0)$ and $(\hat{w}, \hat{\pi})$. But then take some $U$ satisfying $\frac{U}{\delta} - \frac{(1-\delta)v}{\delta} \in [0, \hat{w}]$, and $\frac{U}{\delta} + \frac{(1-\delta)c}{\mu \lambda \delta} \in [0, \hat{w}]$, which exists for sure since $\delta$ is large enough. Then one can easily see that the delegated monitoring together with the efficient approval decisions and the above continuation utilities would bring a higher utility to the principal compared to $V(U)$ at this $U$. This contradicts with the definition of $V$.

The next step is to show that no monitoring is optimal for some interval $[0, \hat{U}_N]$. This follows from the arguments provided in the baseline model. Moreover $V$ is linear over $[0, \hat{U}_N]$. Moreover, similar arguments can be used to show that non-delegated monitoring is optimal for some interval $[\hat{U}_M, \hat{w}]$, and $V$ is linear over $[\hat{U}_M - (1-\delta)\hat{w}, \hat{w}]$. Then denoting $\bar{U}_N = \frac{\hat{U}_N}{\delta}$, and $\bar{U}_D = \frac{\hat{U}_D - (1-\delta)\hat{w}}{\delta}$ completes the proof.

Proof of Theorem 5.

The proof will show that, when the conditions provided in the theorem are satisfied, the value of $V(\delta \hat{w} - \frac{(1-\delta)c}{\mu \lambda})$ is greater than $\hat{\pi}$. If principal delegates monitoring at $U = \delta \hat{w} - \frac{(1-\delta)c}{\mu \lambda}$, then she can utilize the efficient approval decisions. The resulting continuation utilities are

$$U_s = \hat{w}$$

$$U_{ns} = \hat{w} - \frac{(1-\delta)c}{\delta \mu \lambda} - \frac{(1-\delta)v}{c}$$

Note that, since the discount factor $\delta$ is large enough, the resulting $U_{ns}$ is non-negative and feasible. It is already known that $V(\hat{w}) = \hat{\pi}$. In what follows, I construct a lower bound on the value of $V(U_{ns})$, and show that the resulting utility of the principal from delegated monitoring at $U = \delta \hat{w} - \frac{(1-\delta)c}{\mu \lambda}$ is higher than $V(\hat{w})$. More precisely,

$$(1 - \delta)\pi + \delta \mu \lambda \hat{\pi} + \delta (1 - \mu \lambda)V(U_{ns}) > \hat{\pi},$$

where $\pi = (1 - \mu \lambda)(v - \mu_{ns}) - c$ as before. If principal delegates monitoring at $U_{ns}$, the resulting continuation utilities after self-disclosure and no self-disclosure will be inside the interval $[0, \hat{w}]$. 

55
since \( \delta \) is large enough. One can construct a lower bound for the value of \( V(U_{ns}) \) by assuming a randomization at these continuation values between non-delegated monitoring and direct rejection. Then, by plugging the resulting lower bound into the above equation, one can easily show the inequality holds.

Finally, I will show that it is possible to reach a promised utility that is higher than \( \hat{w} \) in the optimal policy, because \( U_s(\bar{U}_D) \geq \hat{w} \). Suppose \( U_s(\bar{U}_D) < \hat{w} \) to get a contradiction. Then, at \( \bar{U}_D + \epsilon \), where \( \epsilon \) is sufficiently small, the principal, by delegating monitoring with probability 1, can do better than \( V(\bar{U}_D + \epsilon) \). First, note that

\[
\begin{align*}
U_s(\bar{U}_D + \epsilon) &= U_s(\bar{U}_D) + \frac{\epsilon}{\delta}, \\
U_{ns}(\bar{U}_D + \epsilon) &= U_{ns}(\bar{U}_D) + \frac{\epsilon}{\delta}.
\end{align*}
\]

Denoting the constant slope of \( V \) along \([\bar{U}_D, \hat{w}]\) by \( s \), it is known that \( V(\bar{U}_D + \epsilon) = V(\bar{U}_D) + s\epsilon \). However, if principal delegates the monitoring to agent at \( U = \bar{U}_D + \epsilon \), she will get:

\[
(1 - \delta)\pi + \delta \mu \lambda V(U_s(\bar{U}_D + \epsilon)) + \delta(1 - \mu \lambda) V(U_{ns}(\bar{U}_D + \epsilon)),
\]

which is greater than:

\[
(1 - \delta)\pi + \delta \mu \lambda V(U_s(\bar{U}_D)) + \delta(1 - \mu \lambda) V(U_{ns}(\bar{U}_D)) + s\epsilon = V(\bar{U}_D) + s\epsilon
\]

because the slope of \( V \) at \( U_s(\bar{U}_D) \) is strictly larger than \( s \). And this concludes the proof as it generates a contradiction.

Therefore en eventual outcome in which the principal always randomizes between non-delegated monitoring and direct approval is possible.