Exploitation Payoffs and Incentives for Exploration

Jangwoo Lee[*] William Fuchs[†] Martin Dumav[‡]

July 10, 2024

Abstract

We study a dynamic moral hazard problem involving initial exploration followed by exploitation, merging experimentation with dynamic corporate finance. We show how the methods and conclusions of the experimentation literature change when considering the exploitation phase's non-monotonic payoff structure that arises naturally in the presence of moral hazard and limited liability. In particular, the agent's incentive constraints may be slack during the exploration phase, which affects compensation dynamics and can reduce inefficiencies from under-experimentation.

1 Introduction

Many economic situations are characterized by an initial phase of exploration, a breakthrough that establishes profitability, and a subsequent phase of exploitation. For example, startup firms often first come up with a working prototype of their product and, after successful development, move on to mass production and sales. The goal of this paper is to study the optimal contract design for such a multistage project. We highlight how the properties of the optimal contract for experimentation hinge on the nature of the payoff structure in the exploitation stage. Importantly, this allows us to connect the dynamic corporate finance literature to the experimentation literature and show how both the methods and conclusions of the latter may be modified.

Concretely, we consider a continuous-time contracting model of a multistage project that involves an initial exploration phase, a breakthrough, and a subsequent exploitation phase. The exploration phase is modeled as an exponential bandit problem. A deep-pocketed principal hires a financially constrained agent protected by limited liability. Both players are ex-ante

[*]Chinese University of Hong Kong
[†]Universidad Carlos III de Madrid and University of Texas at Austin (2110 Speedway CBA 6.276, Austin, TX), William.Fuchs@mccombs.utexas.edu
[‡]Universidad Carlos III de Madrid

symmetrically uncertain about the profitability of the project, which may be good or bad. To maintain tractability, we impose the standard assumption that a bad project can never succeed so that success perfectly reveals that the project is good (Keller et al., 2005). The principal incurs an operational flow cost to keep the project running. The agent can exert effort, which increases the instantaneous success rate of the project. In most of our analysis, we assume that the project can only succeed if the agent exerts effort. In the later part of the paper, we analyze the implications of the possibility of success without effort. Effort is costly to the agent and is unobservable to the principal. Thus, off the equilibrium path, deviations by the agent can lead to differences in beliefs between the two players. Success is publicly observed. If the project succeeds, it achieves a breakthrough, and the relationship transitions into the exploitation stage.

The key feature of our model is that the shape of the exploitation value function (i.e., the principal's maximum exploitation profit that delivers a certain level of exploitation utility to the agent) has important implications for the division of surplus between the players in the exploration stage, which in turn determines the optimal contract design. To capture both the prior literature on exponential bandits and a large literature in dynamic corporate finance, we consider two cases regarding the shape of the exploitation value function:

Case 1) Downward-Sloping Exploitation Value Function

Suppose the breakthrough generates a fixed reward, and the production technology yields no additional value afterward. Then, the two players split the fixed reward upon success, so the principal's exploitation value function must be affine and strictly decreasing in the agent's exploitation utility, as illustrated in Figure 1-(a). This specification of the exploitation value function is in line with the one considered in the prior literature on contracting with exponential bandits (Bergemann and Hege, 1998, 2005; Hörner and Samuelson, 2013; Halac et al., 2016).

Case 2) Inverted-U Shaped Exploitation Value Function

We will mostly focus on the case when the principal's exploitation value function is inverted U-shaped as in Figure 1-(b). This specification for the shape of the principal's exploitation value function arises in several canonical models in the dynamic contracting literature (Quadrini, 2004; Clementi and Hopenhayn, 2006; DeMarzo and Sannikov, 2006; DeMarzo and Fishman, 2007a,b; Vereshchagina and Hopenhayn, 2009; DeMarzo et al., 2012; Fuchs et al., 2022).[1]

Working with reduced-form payoffs for the exploitation stage allows us to highlight the broad

[1]See Hörner (2013) for a clear exposition on why the principal's value function takes this form in many dynamic contracting environments.

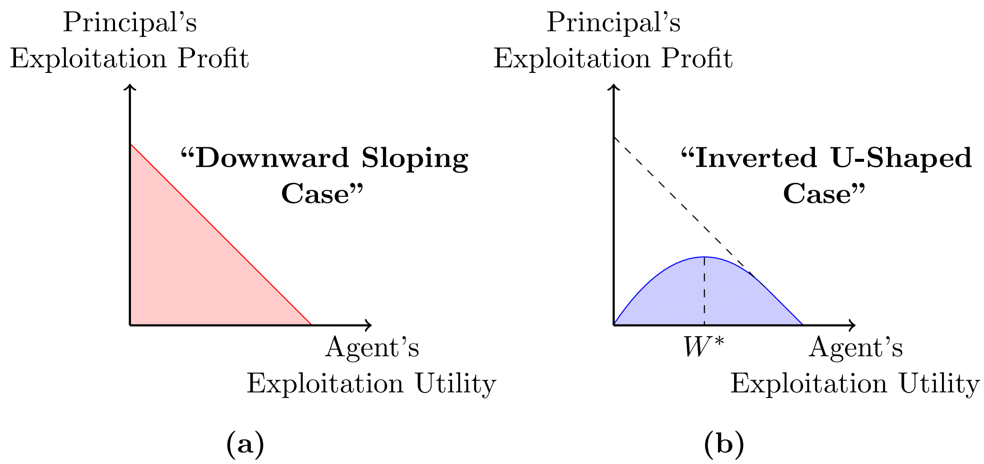


Figure 1: Exploitation Payoff Sets

applicability of our methods and findings.

Before discussing the results, we highlight that the contracting problem we study poses two methodological challenges. First, the possibility of persistent private information can complicate the analysis since it may make multistage deviations profitable (see, for instance, Fernandes and Phelan (2000)). [2] Therefore, we modify the standard martingale approach to account for the agent's incentive to deviate from the effort recommendation.

Second, temporary incentive compatibility constraints may not always bind under the optimal contract. Thus, it is restrictive to confine the analysis to contracts with binding temporary incentive compatibility constraints. This precludes us from using some standard methods in the continuous time contracting literature. Instead, we extend Grossman and Hart (1983)'s two-step approach and apply techniques in dynamic control theory to derive the optimal contract.

Our analysis proceeds as follows. First, we solve for the optimal contract that incentivizes the agent to exert effort until a fixed termination time. [3] Second, we look for the termination time that maximizes the principal's expected total profit. Each step involves an optimal control program, whose optimality conditions can be characterized by Pontryagin's maximum principle and dynamic envelope theorems.

[2] In our model, in the exploration stage, if the agent deviates from the principal's effort recommendation (which only occurs off the equilibrium path), the agent can acquire private information about profitability and potentially enjoy an information rent afterward.

[3] As discussed below, when the principal's investment cost per unit time is sufficiently high, it entails no loss of generality to restrict our attention to recommendations in which the agent exerts effort until the termination time.

Our main result is that the optimal contract design crucially depends on the shape of the principal's exploitation value function. In particular, when the exploitation value function is inverted U-shaped, the agent's temporary incentive compatibility constraint may be slack for a positive duration of time in the exploration phase. More strikingly, there exist a range of parameters for which the temporary incentive compatibility constraint in the optimal contract never binds throughout the entire exploration phase. In contrast, when the exploitation value function is downward-sloping, the temporary incentive compatibility constraint must always bind in the exploration phase under the optimal contract.

The intuition is as follows. When the principal's exploitation value function is downward-sloping, she finds it optimal to minimize the agent's rent upon success. Therefore, at any time in the exploration phase, the principal promises the agent the minimal rent upon success required to elicit his effort, which results in binding temporary incentive constraints at all times. In contrast, when the exploitation function is inverted U-shaped, the principal can potentially benefit from promising an additional rent upon success. Therefore, it may be profitable to "overcompensate" the agent with an exploitation utility that exceeds the minimal level needed for incentive provision, which in turn might slacken the temporary incentive constraint.

The shape of the exploitation value function has two additional implications for the optimal contract design. First, we compare two environments: one with an inverted inverted U-shaped value function and another with a downward-sloping value function. Furthermore, assume that the principal's profits attainable in the former are lower than those in the latter for any promised exploitation payoff to the agent. Surprisingly, we show that the former can still induce longer experimentation in the optimal contract.

Intuitively, when the principal's exploitation value function is downward-sloping, she incurs an additional shadow cost from extending the duration of experimentation. In contrast, with an inverted U-shaped value function, the principal finds it less costly to promise additional rent to the agent upon success. Consequently, she has less incentive to minimize the agent's exploitation utility by reducing the duration of experimentation, potentially reducing the loss from under-experimentation. Thus, an inverted U-shaped value function, despite being strictly lower, can induce more experimentation and reduce inefficiency in the optimal experimentation contract. This implies that when considering the duration of experimentation contracts, we must consider not just the magnitude of the attainable ex-post values but also their shape.

Second, when the principal's exploitation value function is inverted U-shaped, the agent's exploitation utility stays constant over a time interval when the temporary incentive constraint is slack. In particular, if the temporary incentive constraint is always slack, the principal implements the exploitation contract that maximizes her exploitation utility (i.e., promises W^* in Figure 1-(b)). Heuristically, absent the temporary incentive constraint, the principal strictly prefers to promise W^* to the agent upon success in order to maximize her exploitation profit. Therefore, until the temporary incentive constraint tightens and prevents her from doing so, she promises the same exploitation utility to the agent for a positive duration of time.

Our paper highlights the importance of considering the structure of exploitation payoffs when designing incentives during the exploration phase and conducting empirical tests in environments that involve learning. For example, if the exploitation value function is downward-sloping, the agent's compensation for achieving a breakthrough increases over time, which contradicts commonly observed compensation practices.[4] In contrast, if the exploitation value function is inverted U-shaped, our paper can provide a micro-foundation for "milestone contracts" observed in the venture capital industry. Namely, the principal (e.g., venture capitalist) promises a fixed percentage of share capital to the agent (e.g., entrepreneur) if the project meets a pre-specified milestone by a certain deadline. Thus, taking into account the structure of exploitation payoffs may help generate more realistic predictions.

Our main analysis focuses on the case where the project can only succeed when the agent exerts effort. Under this assumption, the temporary incentive compatibility constraint can only be slack at the beginning of the exploration phase, which greatly facilitates the derivation of the optimal contract. In the later part of the paper, we relax this assumption and show that the temporary incentive compatibility constraint can be slack in multiple time intervals in the optimal contract. In so doing, we provide additional insights on how the principal can slacken the temporary incentive constraints to maximize her payoff.

Our paper builds on the prior literature on contracting with exponential bandits (Bergemann and Hege, 1998, 2005; Hörner and Samuelson, 2013; Halac et al., 2016). In these works, the initial breakthrough generates a fixed reward and leads to the termination of the relationship afterward. In contrast, we consider a general specification of the principal's exploitation value function that incorporates both the prior literature and the dynamic corporate finance literature. By doing so, we show that the temporary incentive constraint may not always be tight in the optimal contract and develop a general procedure that characterizes the optimal contract.

Green and Taylor (2016) analyze contracting for a multistage project in the absence of learn-

[4] In the venture capital industry, entrepreneurs' shares tend to become diluted whenever new funds are provided. see, for instance, Sahlman (1990); Lerner (1995); Gompers (1995).

ing. They focus on incentivizing the agent to truthfully report an intermediate breakthrough, which is privately observed by the agent. In contrast, we consider the case where the breakthrough is publicly observable and instead face the challenge that the agent can privately learn about the profitability of the project off the equilibrium path. Furthermore, we identify conditions for which the temporary incentive compatibility constraint should be slack in this environment.

Finally, Khalil et al. (2020) also study contracting for a multistage project with exponential bandits. More specifically, they study an adverse selection problem in which the agent has private information about its effort cost, whereas we consider a moral hazard problem in which the agent's past effort choices are unobservable to the principal. Therefore, the two papers are concerned with different information frictions.

2 Setup

2.1 Environment

Time is continuous and indexed by $t \in [0, \infty)$. A deep-pocketed principal (referred to as "she") can contract with a cashless agent (referred to as "he") to operate a project of unknown profitability, denoted by $\theta \in \{G, B\}$. A project with $\theta = G$ is called a *good project,* whereas a project with $\theta = B$ is called a *bad project.* Both players are risk-neutral, live forever, discount future payoffs at a common discount rate $r > 0$, and share a prior belief $\mathbb{P}_0(\theta = G) = \pi_0 \in (0, 1)$.

At $t = 0$, the principal offers a long-term contract to the agent, who can either accept or reject it. If the agent accepts the contract, the principal will be committed to the contractual terms afterwards. Let us heuristically outline how the "stage game" proceeds during the infinitesimal time interval $[t, t + dt]$. At time t, the principal invests $i > 0$ per unit time to run the project. After the investments are sunk, the agent privately chooses to either exert effort ($a_t = 1$) or shirk ($a_t = 0$). His flow effort cost at time t is given by $ca_t dt$ for some $c > 0$.

The project can either fail or succeed over the interval $(t, t + dt]$. Project success is modelled as a single jump process whose time-t instantaneous success rate is $\lambda a_t \times \mathbf{1}_{\{\theta=G\}}$ with $\lambda > 0$. Therefore, the project can potentially succeed at time t only if (1) the agent exerts effort at time t (i.e., $a_t = 1$), and (2) Nature chooses a good project (i.e., $\theta = G$). Let τ denote the random time at which success occurs, with $\tau = \infty$ if the project never succeeds. In Section 5, we extend our analysis to the model in which the project can potentially succeed even without the agent's effort.

If the project fails, it generates zero flow revenues. Following a failure, the project can either be continued or irreversibly terminated. We denote termination time by T. Upon termination, each player earns a zero reservation payoff.[5]

If the project succeeds, it leads to a *breakthrough* in the contractual relationship. Let us provide an abstract, reduced-form representation of continuation play in the exploitation phase. In particular, in case success occurs at time t, a player's time-t continuation payoff (i.e., the expected value of the player's discounted future payoff beginning at time t) shall be expressed as

$$W_t^S \text{ for the agent, } V(W_t^S) \text{ for the principal.} \tag{1}$$

Henceforth, we refer to W_t^S as the agent's exploitation utility upon success at time t, and $V(\cdot)$ as the principal's exploitation value function. Verbally, $V(W_t^S)$ represents the principal's maximal continuation profit in the exploitation phase that delivers the agent W_t^S upon success at time t. We require that $W_t^S \geq 0$, which reflects either the agent's interim participation constraint or the limited liability constraint. This requirement precludes the first-best outcome in which the principal "sells the enterprise" to the agent.

Throughout the paper, we impose the following regularity Assumption on the exploitation value function:

Assumption 0. *The principal's exploitation value function $V : \mathbb{R}_+ \to \mathbb{R}$ has the following properties:*

(i) The function $V(\cdot)$ is concave and differentiable with $V' \geq -1$.

(ii) The function $V(\cdot)$ has a unique maximizer $W^* \geq 0$.

Both Conditions (i) and (ii) are not overly restrictive and satisfied in many canonical models of moral hazard.[6] These regularity conditions ensure that the principal's optimization program has a well-defined solution.

Let us formally define the players' strategies in the exploration phase. A contract Γ specifies termination time T and a continuation plan $W^S := \{W_t^S\}_{t \leq T}$ in the exploitation phase, which is a sequence of the agent's exploitation utility $W_t^S \geq 0$ upon success at time t.[7] In response,

[5] All our main results continue to hold even when (1) the players' reservation payoffs are assumed to be weakly positive, (2) when the project generates positive flow outputs upon failure, or both.

[6] See, for instance, Quadrini (2004) and Hörner (2013) for a lucid exposition as to why value functions in many dynamic contracting papers are often concave and have a derivative weakly greater than -1.

[7] Since $V' \geq -1$, it entails no loss of generality to backload compensation until the exploitation phase. Thus, even if we allowed the principal to pay nonnegative fixed wages to the agent in the exploration phase,

the agent chooses his action strategy $a := \{a_t\}_{t \leq \tau \wedge T}$ in the exploration phase.[8] In keeping with the prior literature on dynamic contracting with learning (DeMarzo and Sannikov, 2017; He et al., 2017), we confine our attention to pure strategies so that in equilibrium, all players make the same inference about the past path of play. Moreover, for technical purposes, we require that an action strategy a must be $\mathscr{F}$-predictable, where $\mathscr{F} := \{\mathscr{F}_t\}_{t \geq 0}$ is the filtration generated by the outcome process $\{\mathbf{1}_{\{\tau \leq t\}}\}_{t \geq 0}$. This implies that the agent's action choice a_t at time t depends only on the outcome history up to (but excluding) time t. Also, to maintain tractability, we assume that the exploitation utility W_t^S must be absolutely continuous with respect to time.[9]

2.2 Contracting Problem

As is standard in the dynamic contracting literature, we focus on contracts with *a front-loaded effort recommendation*, under which the agent is recommended to exert effort throughout the exploration phase (i.e., $a_t = 1$ for each $t \leq \tau \wedge T$). This restriction entails no loss of generality if the principal's investment cost $i > 0$ per unit time is sufficiently large.[10] At the beginning of the supergame, the principal offers a contract Γ that solves the following optimization program:

$$\max_{a,T,W^S} \mathbb{E}_a \left[\exp(-r\tau) V(W_\tau^S) \times \mathbf{1}_{\{\tau \leq T\}} - i \int_0^{\tau \wedge T} \exp(-rt) a_t dt \right] \tag{2}$$

$$\begin{aligned} \text{subject to:} \quad & \mathbb{E}_a \left[\exp(-r\tau) W_\tau^S \times \mathbf{1}_{\{\tau \leq T\}} - c \int_0^{\tau \wedge T} \exp(-rt) a_t dt \right] \\ \geq & \mathbb{E}_{\hat{a}} \left[\exp(-r\tau) W_\tau^S \times \mathbf{1}_{\{\tau \leq T\}} - c \int_0^{\tau \wedge T} \exp(-rt) \hat{a}_t dt \right] \quad \text{for any action strategy } \hat{a}, \end{aligned} \tag{IC}$$

where $\mathbb{E}_a$ is the expectation operator associated with the probability measure induced by an action strategy a and the prior belief that $\mathbb{P}_0(\theta = G) = \pi_0$. Henceforth, the constraint (IC) shall be referred to as the *full incentive compatibility constraint*, which requires that the agent must find it optimal to adhere to the strategy recommended by the principal.[11]

the principal would find it suboptimal to do so.

[8]For any pair (t_1, t_2), we denote the minimum of t_1 and t_2 by $t_1 \wedge t_2$, and the maximum of t_1 and t_2 by $t_1 \vee t_2$.

[9]This requirement is useful when applying Pontryagin's maximum principle, as it rules out discontinuity with respect to time.

[10]Note that even for small $i > 0$, the principal always recommends a front-loaded effort recommendation under fully observable effort, so it is natural to restrict our attention to this class of recommendations.

[11]By the full incentive compatibility constraint (IC), the agent's equilibrium continuation utility must be higher than his continuation utility from always shirking, which yields the agent a continuation utility weakly

3 Reformulation of Contracting Problem

In Section 3, we modify the martingale approach à la Sannikov (2008) to reformulate the incentive compatibility constraint. This alternative representation allows us recast the contracting problem as an optimal control problem. In so doing, we can make use of standard techniques in optimal control theory to derive the optimal contract.

Let us first introduce additional notations. Denote by $\mathbb{P}_a^G$ the probability measure induced by the agent's action strategy a from the perspective of someone that knows that the project is good. Let $\mathbb{E}_a^G$ denote the expectation operator associated with the probability measure $\mathbb{P}_a^G$. The probability measure $\mathbb{P}_a^G$ is a convenient tool to reformulate the incentive compatibility constraint in the exploration phase because the agent's past deviation has no effect on this probability measure. Therefore, it allows us to apply the martingale approach even when the agent's deviation can create a divergence between the two players' belief.

For any $t \leq \tau \wedge T$, define

$$W_t^G(\Gamma, a) := \mathbb{E}_a^G \left[\exp(-r(\tau - t)) W_\tau^S \times \mathbf{1}_{\{\tau \leq T\}} - c \int_t^{\tau \wedge T} \exp(-r(s-t)) a_s ds \middle| \mathscr{F}_t \right], \tag{3}$$

to be the agent's time-t equilibrium continuation utility under contract Γ, from the perspective of someone that knows that the project is good. Moreover, the process $W^G(\Gamma, a) := \{W_t^G(\Gamma, a)\}_{t \leq \tau \wedge T}$ is formally $\mathscr{F}$-adapted, implying that the continuation utility under the probability measure $\mathbb{P}_a^G$ given in Equation (3) is evaluated after the realization of the project outcome at time t. Since success perfectly reveals that the project is good, we must have $W_t^G(\Gamma, a) = W_t^S$ upon success at time $t \leq T$.

The remainder of Section 3 is organized as follows. In Subsection 3.1, we derive the "promise-keeping constraint" under the probability measure $\mathbb{P}_a^G$, which describes the law of motion for the agent's continuation utility $W_t^G(\Gamma, a)$ under the probability measure $\mathbb{P}_a^G$. In Subsection 3.2, we reformulate the full incentive compatibility constraint (IC) for contracts with a front-loaded effort recommendation. In Subsection 3.3, we rely on these contractual constraints to recast the original contracting problem (2) as an optimal control program, and consider another optimal control program that represents the more constrained version of the problem (2).

higher than his reservation utility. Therefore, any incentive compatible contract must always satisfy the agent's ex interim participation constraints.

3.1 Promise-Keeping Constraint Under Probability Measure $\mathbb{P}_a^G$

We next derive a bookkeeping equation that describes the evolution of the agent's continuation utility $W_t^G(\Gamma, a)$ under the probability measure $\mathbb{P}_a^G$. We rely on the martingale techniques developed in the continuous-time contracting literature (Sannikov, 2008; Biais et al., 2010). For any $t \leq \tau \wedge T$, consider an auxiliary variable that represents the time-t expectation of his lifetime utility under the probability measure $\mathbb{P}_a^G$:

$$U_t^G(\Gamma, a) := \mathbb{E}_a^G \left[\exp(-r\tau) W_\tau^S \times \mathbf{1}_{\{\tau \leq T\}} - c \int_0^{\tau \wedge T} \exp(-rt) dt \middle| \mathscr{F}_t \right]. \tag{4}$$

Note that this expression makes use of the front-loaded nature of the effort recommendation, so that $a_t = 1$ for all $t \leq \tau \wedge T$. By the Law of Iterated Expectations, the process $U^G(\Gamma, a) := \{U_t^G(\Gamma, a)\}_{t \leq \tau \wedge T}$ must be a martingale under the probability measure $\mathbb{P}_a^G$ with respect to the filtration $\mathscr{F}$. Since this filtration is induced by outcome histories, we can apply the martingale representation theorem for point processes to $U^G(\Gamma, a)$ to yield the following result:[12]

Lemma 1. *For any contract Γ, there exists an $\mathscr{F}$-predictable process $\beta^G(\Gamma, a) := \{\beta_s^G(\Gamma, a)\}_{t \leq \tau \wedge T}$ by which the posterior lifetime utility process $U^G(\Gamma, a)$ under the probability measure $\mathbb{P}_a^G$ can be represented as follows:*

$$U_t^G(\Gamma, a) = U_0^G(\Gamma, a) + \int_0^t \exp(-rs) \beta_s^G(\Gamma, a) \Big\{ dN_s - \lambda ds \Big\}, \tag{5}$$

for any $t \leq \tau \wedge T$.

The proof of Lemma 1 is standard and can be found in, for example, Biais et al. (2010).

The "martingale representation" given by Equation (5) has the following economic interpretation. After appropriately adjusting for discounting, the instantaneous change in U_t^G can be expressed as the product of two terms: (1) an increment $dN_t - \Lambda_t dt$ that reflects whether there was a jump at time t, and (2) a term β_t^G that reflects the sensitivity of the agent's continuation utility under the probability measure $\mathbb{P}_a^G$ to success at time t.

Taken together, two Equations (3) and (5) imply:

$$U_t^G(\Gamma, a) = -c \int_0^t \exp(-rt) ds + \exp(-rt) W_t^G(\Gamma, a),$$

for any $t \leq \tau \wedge T$. After plugging the "martingale representation" given in (5) into the left

[12]See, for instance, Theorem T9 on page 64 in Chapter 3 of Brémaud (1981).

side of the equation above, differentiation yields the following law of motion:

$$dW_t^G(\Gamma, a) = \Big[rW_{t^-}^G(\Gamma, a) - \lambda\beta_t^G(\Gamma, a) + c\Big]dt + \beta_t^G(\Gamma, a)dN_t, \tag{6}$$

for any $t \le \tau \wedge T$, where $W_{t^-}^G(\Gamma, a) := \lim_{s \uparrow t} W_s^G(\Gamma, a)$ represents the left-limit of the process $W^G(\Gamma, a)$ with respect to time at t.

With a slight abuse of notation, we suppress the dependence on the contract Γ and let W_t^G denote the agent's equilibrium continuation utility under the probability measure $\mathbb{P}_a^G$ after the project has failed up to and including time t. Also, define $\frac{\partial W_t^G}{\partial t}$ to be its time derivative given by the drift term in Equation (6), and β_t^G to be the increase in the agent's time-t continuation utility under the probability measure $\mathbb{P}_a^G$ upon success at time t.

In this notation, Equation (6) states that the agent's equilibrium continuation utility W_t^G increases to $W_t^G + \beta_t^G$ upon success at time t, which coincides with W_t^S by the definition given in (6). Moreover, $W_T^G = 0$, which reflects the agent's reservation payoff upon termination. In summary, we have:

$$\begin{aligned} &\frac{\partial W_t^G}{\partial t} = rW_t^G - \lambda\beta_t^G + c \qquad \text{for any } t \le T, \qquad W_T^G = 0; \\ &W_t^S = W_t^G + \beta_t^G \qquad\qquad\qquad \text{for any } t \le T. \end{aligned} \tag{PKG}$$

Henceforth, we refer to the system of equations in (PKG) as the *promise-keeping constraint under the probability measure* $\mathbb{P}_a^G$. The constraint keeps track of the principal's utility promise to the agent under the continuation contract, when evaluated on the basis of the probability measure $\mathbb{P}_a^G$.

Since the boundary value problem in the first line of the constraint (PKG) is linear, its solution can be explicitly expressed as follows:

$$W_t^G = \int_t^T \exp(-r(\tau - t))\Big(\Lambda_s\beta_\tau^G - ca_\tau\Big)d\tau. \tag{7}$$

Hence, once the termination time $T > 0$ and the sensitivity process $\beta^G := \{\beta_t^G\}_{t \le T}$ are both determined, we may pin down the continuation plan $W^S := \{W_t^S\}_{t \le T}$ in the exploitation phase based on the integral expression in (7) and the promise-keeping constraint in (PKG).

3.2 Temporary Incentive Compatibility Constraints

Let us make use of the sensitivity process β^G to reformulate the full incentive compatibility constraint (IC) under a contract with a front-loaded effort recommendation. To facilitate the reader's intuition, we begin with heuristic arguments and then continue by stating the more formal result.

Consider a "discretized" version of our model in which the agent can adjust his actions only at time points in $\mathcal{T}(\Delta) := \{0, \Delta, 2\Delta, ...\}$, where $\Delta > 0$ represents a sufficiently short duration of time. Fix any contract Γ and an action strategy $\hat{a}$ in which the agent may potentially deviate from the recommended action until time $\hat{t} \in \mathcal{T}(\Delta)$ and exerts effort afterward. Under the contract-action pair $(\Gamma, \hat{a})$, for any outcome history in which the project has failed up to and including time t ($\leq T$), define π_t^A to be the agent's time-t belief that the project is good, and π_t^P to be the principal's belief that the project is good. By Bayes' Rule, the belief pair (π_t^P, π_t^A) obeys the following law of motion:

$$\begin{aligned} d\pi_t^P =& -\lambda \pi_t^P (1-\pi_t^P)dt, \;\; d\pi_t^A = -\hat{\Lambda}_t \pi_t^A (1-\pi_t^A)dt \qquad \text{for any } t \leq T; \\ \pi_0^P =& \pi_0^A = \pi_0, \end{aligned} \tag{8}$$

where $\hat{\Lambda}_t := \lambda \hat{a}_t$ is the instantaneous success rate of a good project at time t implied by the agent's action strategy $\hat{a}$. Note that the principal can never observe the agent's action strategy, so she updates her belief based on her recommendation.

After the project has failed up to time $t - \Delta$, the agent believes that a good project succeeds once over $[t - \Delta, t)$ with approximate probability $(1 - \exp(-\hat{\Lambda}_t \Delta))$, and fails with residual probability. Therefore, the agent's expected continuation utility after the project has failed up to time $\hat{t}$ can be approximated as follows:

$$\begin{aligned} &-c\hat{a}_t + \pi_t^A \exp(-r\Delta)\Big((1 - \exp(-\hat{\Lambda}_t\Delta)(W_t^G + \beta_t^G) + \exp(-\hat{\Lambda}_t\Delta)W_t^G\Big) \\ &-(1-\pi_t^A)c\int_t^T \exp(-r(s-t))ds + o(\Delta), \end{aligned}$$

where the two approximations of the agent's continuation value at time t in the first line (i.e., $W_t^G + \beta_t^G$ and W_t^G) are given by the promise-keeping constraint (PKG) under the probability measure $\mathbb{P}_a^G$. Therefore, the marginal increase in the agent's time-$(t-\Delta)$ expected continuation utility from exerting effort over $[t-\Delta, t)$ can be approximately as follows:

$$\pi_t^A \Big(1 - \exp(-\lambda\Delta)\Big) \exp(-r\Delta)\beta_t^G - c + o(\Delta). \tag{9}$$

After dividing by $\Delta > 0$ and letting $\Delta \to 0$, the expression above tends to:

$$\lambda \pi_t^A \beta_t^G - c,$$

which represents the continuous-time limit of the marginal increase in the agent's expected continuation utility from exerting effort at time t.[13]

Assume that for almost every $t \leq T$,[14]

$$\beta_t^G \geq \frac{c}{\lambda \pi_t^P} \tag{10}$$

at time t. Suppose, to the contrary, that the agent's optimal strategy $\hat{a}$ induces him to shirk for a positive duration of time until time $\hat{t}$ and exert effort afterward. It entails no loss of generality to assume that the strategy $\hat{a}$ induces the agent to shirk at time $\hat{t}$ because we may always redefine $\hat{t}$ this way. Since the agent shirks for a positive duration of time before time $\hat{t}$, the agent must be more optimistic about profitability than the principal under the contract-action pair $(\Gamma, \hat{a})$ after the project has failed up to time $\hat{t}$ (i.e., $\pi_{\hat{t}}^A > \pi_{\hat{t}}^P$). Therefore, whenever Condition (10) is satisfied, the marginal increase in the agent's expected continuation utility from exerting effort $\lambda \pi_{\hat{t}}^A \beta_{\hat{t}}^G - c$ at time $\hat{t}$ is strictly positive, implying that the agent is strictly better off by exerting effort at time $\hat{t}$. However, this contradicts with the hypothesis that the strategy $\hat{a}$ is optimal for the agent. Therefore, there exists no profitable deviation from the principal's effort recommendation.

Conversely, assume that the principal's effort recommendation is incentive compatible. Therefore, it must be sequentially rational for him to exert effort throughout the exploration phase. Therefore, along the equilibrium path of play, the agent's decision to exert effort must weakly increase his expected continuation utility at time $\hat{t}$, which implies that the inequality in (10) must hold true for any $t \leq T$.

The next lemma formalizes these intuitive arguments:

Lemma 2. *For any contract Γ, it is optimal for the agent to always exert effort in the exploration phase if and only if*

$$\beta_t^G \geq \frac{c}{\lambda \pi_t^P} \tag{TIC}$$

for almost every $t \leq T$.

[13]Note that since effort is $\mathscr{F}$-predictable, the agent's effort is chosen prior to the realization of the outcome at time t.

[14]Consistent with the standard terminology in measure theory, "for almost every $t(\leq \tau)$" means every time except potentially on a time set with zero Lebesgue measure.

Proof. In the Appendix. □

In the dynamic contracting literature, the condition (TIC) often is frequently called the *temporary incentive compatibility constraint* (Green, 1987; Fernandes and Phelan, 2000; Kapička, 2013). Lemma 2 asserts that when checking the incentive compatibility of a contract with a front-loaded effort recommendation, it suffices to consider only "temporary deviations," i.e., alternative strategies in which the agent shirks at time t and exerts effort all the other times in the exploration phase.

Remark 1. *It is noteworthy that under a contract with a front-loaded effort recommendation, the temporary incentive compatibility constraint* (TIC) *forms a sufficient condition for the full temporary incentive compatibility constraint* (IC). *In many models on dynamic contracting with persistent private information, it is crucial to account for the possibility of "double deviations," namely, that after having deviated from the past recommended actions, the agent may face an additional incentive to further deviate from his recommended actions.*[15] *In our framework, once the agent deviates from the principal's effort recommendation, he becomes more optimistic about profitability and has a weaker incentive to shirk in the future. Therefore, for any contract with a front-loaded effort recommendation, as long as the temporary incentive compatibility constraint* (TIC) *is satisfied in the exploration phase, it is sequentially rational for the agent to exert effort even after having shirked in the past.*

3.3 Optimal Control Problems

We recast the original contracting problem (2) based on contractual constraints (PKG), (TIC), and the front-loaded nature of effort recommendations. By the Law of Iterated Expectations and direct computations, we may transform the problem in (2) into the following optimal control problem:

$$\max_{\substack{(\beta^G, W^G, T)\\ \text{s.t. (PKG),(TIC)}}} \left[\pi_0 \int_0^T \exp(-(r+\lambda)\tau)\Big(\lambda V(W_\tau^G + \beta_\tau^G) - i\Big)d\tau - (1-\pi_0)\left(\frac{1-\exp(-rT)}{r}\right) i\right]. \tag{11}$$

In the spirit of Grossman and Hart (1983), we also consider a constrained optimization program that induces the agent to exert effort until a given termination time $T > 0$. Formally, for each $T > 0$, an *incentive scheme* $\mathcal{I}_T$ refers to a pair $\{(\beta_t^G, W_t^G)\}_{t \le T}$ satisfying both contractual constraints (PKG) and (TIC). In particular, for a given termination time $T > 0$, the *constrained optimal incentive scheme* $\mathcal{I}_T^*$ is the solution to the following constrained

[15] See, for example, Fernandes and Phelan (2000) for an earlier discussion of this issue.

optimization problem:

$$\max_{\substack{(\beta^G, W^G) \\ \text{s.t. (PKG),(TIC)}}} \left[\pi_0 \int_0^T \exp(-(r+\lambda)\tau) \Big(\lambda V(W_\tau^G + \beta_\tau^G) - i \Big) d\tau - (1-\pi_0) \left(\frac{1-\exp(-rT)}{r} \right) i \right], \tag{12}$$

where the termination time $T > 0$ is taken as exogenously fixed in Subprogram (12).

4 Derivation of Optimal Contract

In Section 4, we solve for the optimal contract, and show that the temporary incentive compatibility constraint may not necessarily bind under the optimal contract. Let us outline how we plan to proceed. In Subsection 4.1, for each termination time $T > 0$, we analyze the properties of the benchmark incentive scheme in which the temporary incentive compatibility constraints bind at all times in the exploration phase. In Subsection 4.2, we show that when the principal has a downward sloping exploitation value function, the benchmark incentive scheme considered in Subsection 4.1 is constrained optimal given a termination time $T > 0$. In Subsection 4.3, we show that when the principal has an inverted U-shaped exploitation value function, the temporary incentive compatibility constraint (TIC) may not necessarily bind at all times under the constrained optimal incentive scheme. Furthermore, we provide a sufficient condition for which the temporary incentive compatibility constraint never binds in the exploration phase.

4.1 Incentive Scheme With Binding (TIC)'s

For each termination time $T > 0$, let $\mathcal{I}_T^{\text{Binding}} := \{(\underline{\beta}_t^G, \underline{W}_t^G)\}_{t \leq T}$ denote the incentive scheme in which temporary incentive compatibility constraints always bind in the exploration phase, where

$$\underline{\beta}_t^G := \frac{c}{\lambda \pi_t^P}, \ \underline{W}_t^G := \int_t^T \exp(-r(\tau - t)) \Big(\lambda \underline{\beta}_\tau^G - c \Big) d\tau, \ \underline{W}_t^S := \underline{W}_t^G + \underline{\beta}_t^G \tag{13}$$

for each $t \leq T$.

Verbally, when payoffs are evaluated based on the probability measure $\mathbb{P}_a^G$, the term $\underline{\beta}_t^G$ is the sensitivity of the agent's continuation value to success at time t under the incentive scheme $\mathcal{I}_T^{\text{Binding}}$, whereas $\underline{W}_t^G$ is the agent's continuation value after failure up to and including time t under the incentive scheme $\mathcal{I}_T^{\text{Binding}}$. Also, $\underline{W}_t^S$ represents the agent's exploitation utility upon success at time t under the incentive scheme $\mathcal{I}_T^{\text{Binding}}$. While the variables in (13) (i.e., $\underline{\beta}_t^G, \underline{W}_t^G$, and $\underline{W}_t^S$) depend on the termination time T, we slightly abuse the notation and suppress their dependence on the termination time whenever no confusion arises.

The following lemma characterizes the agent's continuation utilities induced by the incentive scheme $\mathcal{I}_T^{\text{Binding}}$:

Lemma 3. *(1) Under an arbitrary incentive scheme $\mathcal{I}_T$ with a given termination time T, the agent's continuation value W_t^G under the probability measure $\mathbb{P}_a^G$ after failure up to and including time t is weakly higher than $\underline{W}_t^G$ induced by $\mathcal{I}_T^{Binding}$ at any $t \leq T$. Additionally, the agent's exploitation utility W_t^S upon success at time t is weakly higher than $\underline{W}_t^S$ associated with $\mathcal{I}_T^{Binding}$ for any $t \leq T$.*

(2) The agent's exploitation utility $\underline{W}_t^S$ associated with the incentive scheme $\mathcal{I}_T^{\text{Binding}}$ strictly increases in the timing of success, t.

Proof. In the Appendix. □

Lemma 3-(1) asserts that for any given termination time T, continuation payoffs $\underline{W}_t^G$ and $\underline{W}_t^S$ induced by the incentive scheme $\mathcal{I}_T^{\text{Binding}}$ impose lower bounds on time-t continuation payoffs that can be induced by an arbitrary incentive scheme under the probability measure $\mathbb{P}_a^G$. Heuristically, if the principal raises the sensitivity term β_t^G of the agent's continuation utility under the probability measure $\mathbb{P}_a^G$ at time t, she must do so by promising a higher exploitation utility W_t^S at time t, which in turn increases the agent's continuation utilities prior to time t under the probability measure $\mathbb{P}_a^G$. Therefore, slack temporary incentive compatibility constraints lead to higher "rents" (i.e., W_t^G and W_t^S) for the agent under the probability measure $\mathbb{P}_a^G$.

To develop intuition behind Lemma 3-(2), let us decompose $\underline{W}_t^S := \underline{\beta}_t^G + \underline{W}_t^G$ into two components $\underline{\beta}_t^G$ and $\underline{W}_t^G$. The first component $\underline{\beta}_t^G$ reflects the compensation for the agent's cost of effort when effort choices are publicly observable. In this first-best case, the principal can extract full surplus by promising the agent an exploitation utility of $\underline{\beta}_t^G := \frac{c}{\lambda \pi_t^P}$ upon success at time t. Since both players grow pessimistic about the likelihood of success with the passage of time in the exploration phase, the principal must raise the agent's exploitation utility $\underline{\beta}_t^G$ over time to accomodate for the lower chance of success. As a result, the first component $\underline{\beta}_t^G$ strictly increases in the timing of success, which in turn leads to the monotonicity of $\underline{W}_t^S$ as well.

The second component $\underline{W}_t^G$ reflects the agent's time-t expected information rent arising from the possibility of belief manipulations when the agent is privately informed of his past effort choices. In this case, the agent's shirking can induce the principal to become more pessimistic than the agent in the future. Thus, under the first-best continuation plan $\{\underline{\beta}_t^G\}_{t\leq T}$ in the exploitation phase, the agent can expect to earn information rents by exerting effort only when the exploitation utility level becomes sufficiently high, so he does not find it optimal to exert effort at all times until termination. Therefore, in order to incentivize the agent, the principal must promise him an additional exploitation utility $\underline{W}_t^G$ upon success at time t.

Remark 2. *If the project can potentially succeed even without the agent's effort, the agent's time-t exploitation utility $\underline{W}_t^S$ induced by binding temporary incentive compatibility constraints until a given termination time T does not necessarily increase over time. Intuitively, under this alternative specification, the agent can obtain moral hazard rents at each time even in absence of learning (i.e., $\pi_0 = 1$). Therefore, in order to incentivize the agent far from termination time T, the agent's continuation utility $\underline{W}_t^G$ tends to drop faster near the termination time T, which may give rise to the non-monotonicity of the exploitation utility $\underline{W}_t^S$ induced by binding temporary incentive compatibility constraints until the termination time T.*

4.2 Case 1: Model With Downward Sloping $V(\cdot)$

In Section 4.2, we consider the case in which the principal has a downward sloping exploitation value function (as illustrated in Figure-1-(a)).

Definition 1. *The principal's exploitation value function $V(\cdot)$ is* ***downward sloping*** if it strictly decreases in the agent's exploitation utility: $V'(w) < 0$ for any $w \in \mathbb{R}_+$.

Definition 1 generalizes a salient property of the principal's exploitation value function in existing contracting models with exponential bandits (Bergemann and Hege, 1998, 2005; Hörner and Samuelson, 2013; Halac et al., 2016). In these works, project success generates a fixed surplus, which is subsequently shared between two players based on the initial contract. Therefore, the principal earns a strictly lower exploitation profit if she promises the agent a larger exploitation utility upon success.

The next lemma shows that when the exploitation value function is assumed to be downward sloping, the temporary incentive compatibility constraint (TIC) must bind throughout the exploration phase under the constrained optimal contract:

Lemma 4. *If the principal's exploitation value function is downward sloping, for each termination time $T > 0$, the incentive scheme $\mathcal{I}_T^{Binding}$ with binding temporary incentive compatibility constraints is constrained optimal.*

Proof. In the Appendix. □

The intuition behind the lemma is as follows. If the principal has a downward sloping exploitation value function and promises a strictly higher exploitation utility upon success at time t, she must earn a strictly lower exploitation profit upon success at time t. Accordingly, the principal finds it optimal to implement the incentive scheme $\mathcal{I}_T^{\text{Binding}}$, which delivers the agent the minimal exploitation utility required to induce him to exert effort at any given time by Lemma 3-(1).

The next proposition summarizes the main results of Section 4.2:

Proposition 1. *Suppose that the principal's exploitation value function is downward sloping. Then, for any given termination time $T > 0$, the temporary incentive compatibility constraints bind throughout the exploration phase under the constrained optimal incentive scheme. Additionally, the agent's exploitation utility $\underline{W}_t^S$ in (13) strictly increases in the timing of success, t.*

Proof. Immediate from Lemmata 3-(2) and 4. □

4.3 Case 2: Model With Inverted U-Shaped $V(\cdot)$

In Section 4.3, we show that when the principal has an inverted U-shaped exploitation value function (as illustrated in Figure-1-(b)), the temporary incentive compatibility constraint (TIC) may not necessarily always bind at all times in the optimal contract.

Definition 2. *The principal's exploitation value function $V(\cdot)$ is* ***inverted U-shaped*** *if its unique maximizer W^* is strictly higher than the principal's reservation payoff: $W^* > 0$.*

By the regularity Assumption 0, it is straightforward to see that an inverted U-shaped exploitation value function is strictly increasing on $[0, W^*]$ and strictly decreasing on $[W^*, \infty)$. This shape can be motivated by appealing to prior works on dynamic moral hazard (Quadrini, 2004; Clementi and Hopenhayn, 2006; DeMarzo and Sannikov, 2006; DeMarzo and Fishman, 2007a,b; Vereshchagina and Hopenhayn, 2009; DeMarzo et al., 2012; Fuchs et al., 2022).

The rest of Section 4.3 is organized as follows. First, in 4.3.1, we derive the general form of the constrained optimal incentive scheme for any given termination time $T > 0$, under which temporary incentive compatibility constraints may not necessarily bind at all times in the exploration phase. Second, in 4.3.2, we identify sufficient conditions for which temporary incentive compatibility constraints never bind under the optimal contract.[16]

[16]Recall from the definitions in Subsection 3.3 that a "constrained optimal incentive scheme" are not mu-

4.3.1 Constrained Optimal Incentive Scheme

Let us derive the general form of the constrained optimal incentive scheme for a given termination time $T > 0$, under which temporary incentive compatibility constraints may not necessarily bind at all times in the exploration phase. Given a termination time $T > 0$, our candidate for the constrained optimal incentive scheme is denoted as $\mathcal{I}_T^*$, under which the agent expects to earn the following exploitation utility upon success at time t:

$$W_t^{S*} := \max\{W^*, \underline{W}_t^S\}. \tag{14}$$

Figure 2 shows the time trajectories of the optimal exploitation utility W_t^{S*} under hidden effort, the minimal exploitation utility $\underline{W}_t^S$ induced by the incentive scheme $\mathcal{I}_T^{\text{Binding}}$, and the exploitation utility $\frac{c}{\lambda\pi_t^P}$ under observable effort.

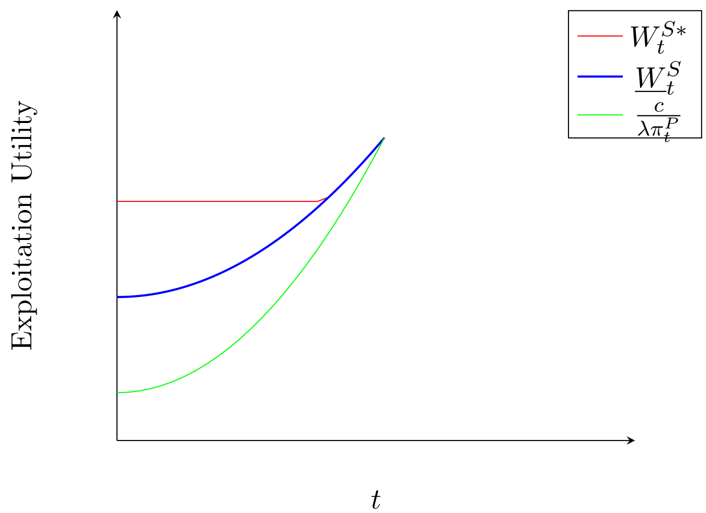


Figure 2: Time Trajectories of W_t^{S*} and $\underline{W}_t^S$

Based on the exploitation utility given in (14), the time-t contracting variables β_t^{G*} and W_t^{G*} associated with the candidate optimal incentive scheme $\mathcal{I}_T^*$ can be computed as follows:

$$W_t^{G*} = \int_t^T \exp(-(r+\lambda)(\tau - t))(\lambda W_\tau^{S*} - c)d\tau, \;\; \beta_t^{G*} = W_t^{S*} - W_t^{G*} \tag{15}$$

for each $t \leq T$. Note that the first equality follows from the definition given in (3) and the front-loaded nature of the principal's effort recommendation, whereas the second equality follows from the second line of the promise-keeping constraint (PKG) under the probability measure $\mathbb{P}_a^G$.

tually exchangeble with an "optimal contract," under which the termination time is also optimally chosen.

As a preliminary step, the next lemma shows that our candidate for the constrained optimal incentive scheme $\mathcal{I}_T^*$ is indeed incentive compatible:

Lemma 5. *Under the candidate $\mathcal{I}_T^*$ for the constrained optimal incentive scheme, the temporary incentive compatibility constraint is slack at any time t with $\underline{W}_t^S < W^*$, and binds at all other times prior to termination.*

Proof. In the Appendix. □

Intuitively, if the principal promises a higher exploitation utility at time t than the minimal possible level $\underline{W}_t^S$ and fixes the level of the agent's exploitation utility for a positive duration of time, the agent finds it less attractive to wait for the exploitation utility to go up at time t. As a result, he has a stronger incentive to exert effort at time t.

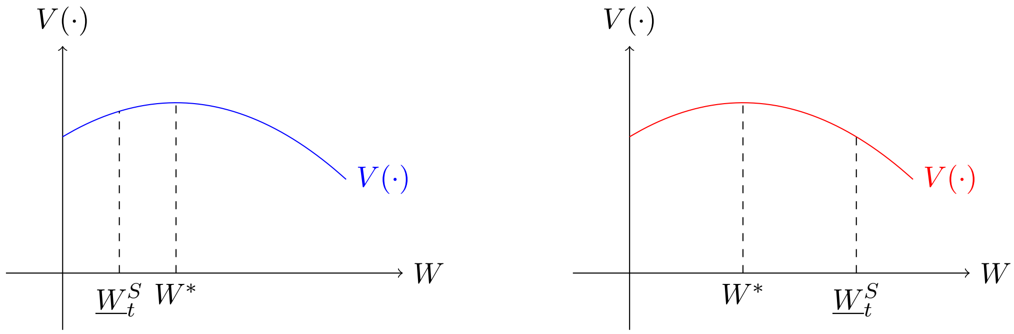
In the next proposition, we verify our conjecture that the incentive scheme $\mathcal{I}_T^*$ is indeed constrained optimal:

Proposition 2. *For each termination time $T > 0$, there exists a constrained optimal incentive scheme $\mathcal{I}_T^*$ as defined by the agent's time-t exploitation utility in* (14)*. Under the constrained optimal incentive scheme, the temporary incentive compatibility constraint is slack at time t if and only if $\underline{W}_t^S < W^*$ at time t.*

Proof. In the Appendix. □

The heuristic argument for Proposition 2 is as follows. First, if $\underline{W}_t^S < W^*$ at time t as in Panel (a) of Figure 3, the incentive scheme $\mathcal{I}_T^*$ delivers the principal her maximal exploitation profit $V(W^*)$ at time t, so no other incentive scheme can deliver a strictly higher exploitation profit at time t than $\mathcal{I}_T^*$ does. Moreover, since the principal promises strictly more than the level required to motivate the agent at any such time, Lemma 5 implies that the temporary incentive compatibility constraint should become slack.

Second, let us suppose that $\underline{W}_t^S \geq W^*$ at time t as in Panel (b) of Figure 3. By Lemma 3-(1), the principal must promise the agent weakly more than $\underline{W}_t^S$ upon success at time t under any incentive scheme. Moreover, by the concavity of the exploitation value function $V(\cdot)$, it must be strictly decreasing on $[W^*, \infty)$. Therefore, the principal's exploitation profit at time t cannot be higher than $V(\underline{W}_t^S)$, which is the level induced by the candidate for the constrained optimal incentive scheme $\mathcal{I}_T^*$.



(a) **(b)**

Figure 3: Two Cases for Proposition 2

4.3.2 Optimal Contract With Never Binding (TIC)'s

Proposition 2 shows that temporary incentive compatibility constraints may not necessarily bind at all times in the constrained optimal incentive scheme given a termination time $T > 0$. However, this still raises the possibility that when the termination time T^* is optimally chosen, temporary incentive compatibility constraints might still always bind under the optimal contract. To address this concern, the next theorem provides sufficient conditions for the existence of an optimal contract under which the temporary incentive compatibility constraint never binds:

Theorem 1. *Suppose that the principal's exploitation value function is inverted U-shaped, and that there exists a belief level* $\underline{\pi} \in (0, \pi_0)$ *such that (1)* $\lambda\underline{\pi}V(W^*) = i$, *and (2)* $\left(\frac{r}{\lambda+r}\right)W^* > \frac{c}{\lambda\underline{\pi}}$. *Then, there exists an optimal contract in which the temporary incentive compatibility constraint never binds throughout the exploration phase. The contractual relationship proceeds as follows. If the project keeps failing, the relationship terminates as soon as the players' equilibrium belief that the project is good reaches* $\underline{\pi}$. *If the project succeeds prior to termination, the principal implements her optimal continuation contract in the exploitation phase (i.e., the one that promises* W^* *to the agent in the exploitation phase) immediately upon success.*

Proof. In the Appendix. □

To better understand this result, let us consider a class of contracts in which the principal promises the agent an exploitation utility of W^* upon success, regardless of success timing. Since W^* is the unique maximizer of the exploitation value function $V(\cdot)$, if the optimal contract among this class of contracts happened to be incentive compatible, the principal would implement this particular contract.

The first condition $\lambda\underline{\pi}V(W^*) = i$ characterizes the principal's optimal contract within this class. In particular, it represents the first-order condition for the termination time, at which the principal's marginal gain $\lambda\underline{\pi}V(W^*)$ from experimentation must become equal to the investment cost i per unit time. [17] Therefore, in our candidate optimal contract, the principal always promises W^* to the agent if the project succeeds before her belief reaches $\underline{\pi}$, and terminates the relationship otherwise.

The second condition $\left(\frac{r}{\lambda+r}\right)W^* > \frac{c}{\lambda\underline{\pi}}$ ensures the incentive compatibility of this candidate optimal contract. In particular, the left side of the condition is a lower bound on the sensitivity of the agent's continuation value under the probability measure $\mathbb{P}_a^G$ to success at time t, assuming that the principal always promises the agent W^* upon success. The right side of the condition represents the upper bound on the minimal sensitivity level necessary to induce the agent to exert effort at any time prior to terimination. Therefore, the second condition requires that under the candidate optimal contract, the agent's continuation value under the probability measure $\mathbb{P}_a^G$ must be more sensitive to success than the level required to motivate the agent prior to termination. In turn, this implies that the temporary incentive compatibility constraint must be always slack throughout the exploration phase.

5 Two-Step Procedure and Extension

Section 4 shows that in the optimal contract, temporary incentive compatibility constraints may not necessarily bind at all times in the exploration phase. Therefore, the standard approach, which is based on the conjecture that temporary incentive compatibility constraints must bind at all times under the optimal contract, may not necessarily lead to the correct form of the optimal contract.

The goal of Section 5 is to develop a solution method that characterizes the optimal contract in a richer model. In particular, Section 5 concerns a setting in which the time-t instantaneous success rate is given by $(\lambda - \Delta_\lambda + \Delta_\lambda a_t) \times \mathbf{1}_{\{\theta=G\}}$ with $\lambda > \lambda - \Delta_\lambda \geq 0$. Thus, if Nature chooses a good project (i.e., $\theta = G$), (1) success rates under the two action choices (i.e., λ and $\lambda - \Delta_\lambda$) are both weakly positive; (2) effort strictly increases the success rate of the project ($\Delta_\lambda > 0$). Therefore, the model considered up to this point is a special case of this one with $\lambda = \Delta_\lambda$.

[17] In the more general model in which with the principal earns $\underline{V} \geq 0$ upon termination in the exploration phase, the first-order condition with respect to the termination belief takes the following form:

$$\lambda\underline{\pi}V(W^*) = i + r\underline{V},$$

where $r\underline{V}$ can be interpreted as the principal's forgone interest from delaying termination per unit time in the exploration phase. Hence, as long as either the opportunity cost of operation (i.e., $i + r\underline{V}$) is strictly positive in the exploration phase, there exist parametric configurations under which Theorem 1 hold true.

When $\lambda > \Delta_\lambda$, the project can potentially succeed even without the agent's effort. Thus, even in the absence of learning (i.e., $\pi_0 = 1$), the agent can enjoy a moral hazard rent prior to termination due to the unobservability of effort. As shown later in Lemma (8), this implies that the agent's exploitation utility induced by binding temporary incentive compatibility constraints is not necessarily monotone in time, which makes it more challenging to derive the optimal contract.

In this model, the time-t temporary incentive compatibility constraint (TIC) must be modified as follows:

$$\beta_t^G \geq \frac{c}{\Delta_\lambda \pi_t^P}.$$

for almost every $t \leq T$.

The rest of Section 5 proceeds as follows. In Subsection 5.1, we extend Grossman and Hart (1983)'s two-step approach to our dynamic environment. More specifically, we break down the original contracting problem (11) into two subprograms, and apply standard techniques in optimal control theory to derive the optimality conditions associated with each subprogram. This more general approach will allow us to handle the case in which the project can succeed without the agent exerting effort, which we do so in Subsection 5.2.

5.1 Solution Method

Step 1) First, we take a termination time $T > 0$ as given and solve for the constrained optimal incentive scheme that induces the agent to exert effort until a given termination time $T > 0$. Formally, in particular, for each $T > 0$, the constrained optimal incentive scheme $\mathcal{I}_T^*$ solves the the following "implementation Subprogram":

$$J(T) := \max_{\substack{(\beta^G, W^G) \text{ s.t.} \\ \text{(PKG),(TIC)}}} \int_0^T \lambda \exp(-(r+\lambda)\tau) V(W_\tau^G + \beta_\tau^G) d\tau. \tag{16}$$

under the probability measure $\mathbb{P}_a^G$.

Since Subprogram (16) is a standard optimal control problem, we can apply Pontryagin's maximum principle to derive the optimality conditions associated with it. After adjusting for discounting and uncertainty, we can define the "present-value Lagrangian" associated with the subprogram as follows:

$$\mathscr{L}(t, W_t^G, \beta_t^G, \mu_t^{\text{PKG}}, \mu_t^{\text{TIC}}) := \lambda V(W_t^G + \beta_t^G) + \mu_t^{\text{PKG}} \Big(r W_t^G - \lambda \beta_t^G + c \Big) + \mu_t^{\text{TIC}} \Big(\beta_t^G - \frac{c}{\Delta_\lambda \pi_t^P} \Big), \tag{17}$$

where μ_t^{PKG} is the present-value multiplier associated with the promise-keeping constraint (PKG) under the probability measure $\mathbb{P}_a^G$, and μ_t^{TIC} is the present-value multiplier associated with the temporary incentive compatibility constraint (TIC).

In line with pioneering works by Dorfman (1969) and Pontryagin et al. (1962), we shall interpret the multiplier μ_t^{PKG} as the principal's present-value shadow cost from delivering an additional continuation value to the agent under the probability measure $\mathbb{P}_a^G$ at time $t \leq T$. Similarly, the multiplier μ_t^{TIC} shall be interpreted as the principal's present-value shadow value of relaxing the temporary incentive compatibility constraint by a single "payoff unit" at time $t \leq T$, when the players' expected discounted payoffs are evaluated on the basis of the probability measure $\mathbb{P}_a^G$.

Note that the exploitation value function $V(\cdot)$ is concave with respect to W_t^G, and the "constraint functions" (i.e., $rW_t^G - \lambda\beta_t^G + c$ and $\beta_t^G - \frac{c}{\Delta_\lambda \pi_{t-}^P}$) are both linear with respect to W_t^G and β_t^G. Therefore, the following lemma characterizes necessary and sufficient conditions for the solution to the implementation Subprogram:

Lemma 6. *The incentive scheme* (β^{G*}, W^{G*}) *solves Subprogram* (16) *if and only if there exists a pair of multiplier processes* $(\mu^{\text{PKG}*}, \mu^{\text{TIC}*}) = \{\mu_t^{\text{PKG}*}, \mu_t^{\text{TIC}*}\}_{t\leq T}$ *satisfying the following conditions at all* $t \leq T$,

$$\underbrace{\lambda\Big(V'(W_t^{G*} + \beta_t^{G*}) - \mu_t^{\text{PKG}*}\Big) + \mu_t^{\text{TIC}*}}_{=:\frac{\partial \mathscr{L}(t, W_t^{G*}, \beta_t^{G*}, \mu_t^{\text{PKG}*}, \mu_t^{\text{TIC}*})}{\partial \beta_t^G}} = 0. \tag{FOC}$$

$$-\frac{\partial \mu_t^{\text{PKG}*}}{\partial t} = \underbrace{\lambda\Big(V'(W_t^{G*} + \beta_t^{G*}) - \mu_t^{\text{PKG}*}\Big)}_{=:\frac{\partial \mathscr{L}(t, W_t^{G*}, \beta_t^{G*}, \mu_t^{\text{PKG}*}, \mu_t^{\text{TIC}*})}{\partial W_t^G} - (\lambda + r)\mu_t^{\text{PKG}*}}. \tag{CE}$$

$$\mu_0^{\text{PKG}*} \leq 0, \quad \mu_0^{\text{PKG}*}\Big(W_0^{G*} - \underline{W}_0^G\Big) = 0. \tag{CSC$_1$}$$

$$\mu_t^{\text{TIC}*} \geq 0, \quad \mu_t^{\text{TIC}*}\Big(\beta_t^{G*} - \frac{c}{\Delta_\lambda \pi_t^P}\Big) = 0, \tag{CSC$_2$}$$

where $\underline{W}_0^G := \int_0^T \exp(-r\tau)\Big(\lambda(\frac{c}{\Delta_\lambda \pi_\tau^P}) - c\Big)d\tau$ denotes the minimal level of the agent's time-0 payoff under an incentive scheme, evaluated based on the probability measure $\mathbb{P}_a^G$. Moreover, if the exploitation value function $V(\cdot)$ is strictly concave, the constrained optimal incentive scheme yields the unique maximum of the implementation Subprogram (16).

Proof. In the Appendix. □

The first requirement given by (FOC) is the *Kuhn-Tucker first-order condition* associated with the Lagrangian (17). It states that for any $t \leq T$, the sensitivity β_t^G of the agent's continuation value at time t under the probability measure $\mathbb{P}_a^G$ must be chosen so that the principal's marginal benefit from increasing β_t^G must be equal to her marginal cost from doing so. As a result, the sum of the marginal change in the principal's value at time t with respect to β_t^G and the shadow value from relaxing the temporary incentive compatibility constraint at time t must be equal to the shadow cost from delivering an additional utility to the agent under the probability measure $\mathbb{P}_a^G$ at time t.[18]

The requirement given by (CE) is called the *co-state equation* associated with the multiplier $\mu_t^{\text{PKG}*}$. In particular, it states that for any $t \leq T$, the decrease in the agent's equilibrium time-t continuation utility per unit time under the probability measure $\mathbb{P}_a^G$ must be equal to the difference between the marginal change in the principal's value with respect to W_t^G at time t and the shadow cost from delivering the agent an additional utility at time t under the probability measure $\mathbb{P}_a^G$.

The *complementary slackness condition associated with the time-*0 *multiplier* $\mu_0^{\text{PKG}*}$ (CSC$_1$) asserts that the principal's shadow value $\mu_0^{\text{PKG}*}$ from increasing the agent's expected discounted payoff at time 0 under the probability measure $\mathbb{P}_a^G$ must be weakly negative, and can be strictly negative only if the principal promises the minimal time-0 expected payoff $\underline{W}_0^G$ required to incentivize the agent under the probability measure $\mathbb{P}_a^G$.

Similarly, the *complementary slackness condition associated with the time-t multiplier* $\mu_t^{\text{TIC}*}$ (CSC$_2$) asserts that for any $t \leq T$, the shadow cost from the temporary incentive compatibility constraint at time t must be weakly positive, and can be strictly positive only if the constraint binds under the optimal contract.

The next lemma shows that it entails no loss of generality to assume that $\mu_0^{\text{PKG}*} = 0$:

Lemma 7. *Suppose that there exists an optimal quadruple* $(\beta^{G*}, W^{G*}, \mu^{\text{PKG}*}, \mu^{\text{TIC}*})$ *satisfying conditions* (PKG), (TIC), (FOC), (CE), (CSC$_1$), *and* (CSC$_2$). *Then, there exists a pair of multiplier processes* $(\tilde{\mu}^{\text{PKG}*}, \tilde{\mu}^{\text{TIC}*})$ *such that 1)* $\tilde{\mu}_0^{\text{PKG}*} = 0$; *2) the quadruple* $(\beta^{G*}, W^{G*}, \tilde{\mu}^{\text{PKG}*}, \tilde{\mu}^{\text{TIC}*})$ *satisfies conditions* (PKG), (TIC), (FOC), (CE), (CSC$_1$), *and* (CSC$_2$).

Proof. In the Appendix. □

[18] In most papers in the continuous-time contracting literature (Sannikov, 2008; Biais et al., 2010), the optimal contract can be obtained by minimizing the sensitivity term β_t^G at any time prior to termination. However, this property does not hold true in our model because the intermediate breakthrough changes the nature of payoff structure.

Heuristically, if the the multiplier $\mu_0^{\text{PKG}*} < 0$, then the complementary slackness constraint in (CSC_1) implies that $W_0^{G*} = \underline{W}_0^G$. A straightforward generalization of Lemma 3-(1) shows that $W_0^{G*} = \underline{W}_0^G$ is equivalent to the temporary incentive compatibility constraint binding at all $t \leq T$. Therefore, the inequality $W_0^{G*} \geq \underline{W}_0^G$ can be treated as a redundant constraint and thus can be dropped, which allows us to set the Lagrange multiplier $\mu_0^{\text{PKG}*} = 0$ without loss of generality.

When coupled together with Lemma 7, the Kuhn-Tucker optimality conditions characterized in Lemma 6 can be simplified as follows:

Proposition 3. *The incentive scheme* (β^{G*}, W^{G*}) *solves Subprogram* (16) *if and only if for all* $t \leq T$,

$$\underbrace{-\lambda V'(W_t^{G*} + \beta_t^{G*}) - \int_0^t \exp(-\lambda(\tau - t))\lambda^2 V'(W_\tau^{G*} + \beta_\tau^{G*})d\tau}_{=\mu_t^{\text{TIC}*}} \geq 0, \tag{18}$$

whose left-hand side μ_t^{TIC*} *must be equal to zero for any time* t *at which the temporary incentive compatibility constraint is slack. Moreover, if the exploitation value function* $V(\cdot)$ *is strictly concave, the constrained optimal incentive scheme yields the unique maximum of the implementation Subprogram* (16).

Proof. In the Appendix. □

Step 2) The second step is to choose the optimal termination time T^* in the following problem:

$$\max_{T\geq 0}\left[\pi_0\Big\{J(T) - \Big(\frac{1-\exp(-(r+\lambda)T)}{r+\lambda}\Big)i\Big\} - (1-\pi_0)\Big(\frac{1-\exp(-rT)}{r}\Big)i\right], \tag{19}$$

where $J(T)$ is the principal's value function associated with the implementation Subprogram (16). We may apply the "dynamic envelope theorem" to obtain the following equations:[19]

$$\begin{aligned} J'(T) &= \exp(-(r+\lambda)T)\mathscr{L}(T, \underbrace{W_T^{G*}}_{=0}, \beta_T^{G*}, \mu_T^{\text{PKG}*}, \mu_T^{\text{TIC}*}) \\ &= \exp(-(r+\lambda)T)\Big(\lambda V(\beta_T^{G*}) - \mu_T^{\text{PKG}*}(\lambda\beta_T^{G*} - c)\Big) \end{aligned} \tag{20}$$

for each $T \geq 0$, where $\mu_T^{\text{PKG}*}$ is the time-T multiplier associated with the constrained optimal

[19] See, for example, Theorem 9.3 pp. 252 Foundations of Dynamic Economic Analysis Optimal Control Theory and Applications by Caputo (2005).

incentive scheme that induces the agent to exert effort until termination time T.

This result has an intuitive interpretation. By marginally increasing the duration of experimentation, the principal can expect to earn an additional profit of $\lambda V(\beta_T^{G*})$ at time T but must also bear the extra shadow cost from delivering an additional flow utility of $\lambda\beta_T^{G*} - c$ to the agent at time T, when the players' continuation payoffs are evaluated on the basis of the probability measure $\mathbb{P}_a^G$.

Taken together with the dynamic envelope result above, differentiation of the objective function with respect to termination time T in the optimization program (11) yields the following first-order condition:

$$\lambda\pi_{T^*}^P V(\beta_{T^*}^{G*}) - i = \pi_{T^*}^P \mu_{T^*}^{\mathrm{PKG}^*}(\lambda\beta_{T^*}^{G*} - c), \tag{21}$$

which must be satisfied as long as the optimal termination time T^* is strictly positive. The left side is the principal's marginal benefit from experimentation at time T^* under fully observable effort choices, whereas the right side reflects the principal's shadow cost from promising the agent an additional rent of $\lambda\beta_{T^*}^{G*} - c$ upon success at time T^*.

5.2 Extension: Model With $\lambda > \Delta_\lambda$

Up to Section 4, we have assumed that $\lambda = \Delta_\lambda$, so that the project can succeed at time t only if the agent exerts effort at time t. In the current Subsection, we instead assume that $\lambda > \Delta_\lambda$, so that the project may succeed even if the agent does not exert effort.

5.2.1 Incentive Scheme With Binding TICs

Let us first chracterize the benchmark incentive scheme induced by binding temporary incentive compatibility constraints. As in Equation (13), for each termination time $T > 0$, let $\mathcal{I}_T^{\text{Binding}} := \{(\underline{\beta}_t^G, \underline{W}_t^G)\}_{t \le T}$ denote the incentive scheme in which temporary incentive compatibility constraints always bind in the exploration phase. The next lemma characterize the time trajectory of the exploitation utility induced by the incentive scheme $\mathcal{I}_T^{\text{Binding}}$:

Lemma 8. *Under the incentive scheme $\mathcal{I}_T^{Binding}$, there exists $\tau \in [0, T)$ such that the exploitation utility $\underline{W}_t^S$ initially strictly increases in time over $[0, \tau)$, and strictly decreases in time over $[\tau, T]$. Moreover, for any $\epsilon > 0$, there exists a $\overline{\Delta}_\lambda \in (0, \lambda)$ such that for all $\Delta_\lambda \in (\overline{\Delta}_\lambda, \lambda)$, the exploitation utility $\underline{W}_t^S$ initially strictly increases in time over $[0, T - \epsilon]$.*

Proof. In the Appendix. □

Since $\tau \in [0, T)$, the exploitation utility $\underline{W}_t^S$ must eventually decrease in time. Therefore, the incentive scheme $\mathcal{I}_T^{\text{Binding}}$ generates one of the following paths: 1) the exploitation utility $\underline{W}_t^S$ initially strictly increases in time up to τ, and strictly decreases in time afterward; 2) the exploitation utility $\underline{W}_t^S$ strictly decreases in time.

To better understand the intuition behind Lemma 8, let us decompose the time-t exploitation utility $\underline{W}_t^S$ under the incentive scheme $\mathcal{I}_T^{\text{Binding}}$ as follows:

$$\begin{aligned}\underline{W}_t^S &:= \frac{c}{\Delta_\lambda \pi_t^P} + \int_t^T \exp(-r(\tau - t))\Big(\lambda\left(\frac{c}{\Delta_\lambda \pi_t^P}\right) - c\Big) d\tau \\ &= \underbrace{\frac{c}{\Delta_\lambda \pi_t^P}}_{\text{Static Agency Cost}} + \underbrace{c\left(\frac{\lambda - \Delta_\lambda}{\Delta_\lambda}\right)\left(\frac{1 - \exp(-r(T-t))}{r}\right)}_{\text{Intertemporal Agency Cost}}, \\ &\quad + \underbrace{c\left(\frac{\lambda}{\Delta_\lambda}\right)\left(\frac{1 - \pi_t^P}{\pi_t^P}\right)\left(\frac{1 - \exp(-(r-\lambda)(T-t))}{r - \lambda}\right)}_{\text{Informational Agency Cost}},\end{aligned} \tag{22}$$

where the second equality follows from direct computations.

In the terminology of Bergemann and Hege (1998), the *static agency cost* refers to the agent's minimal continuation value upon success if the players interact for an infinitesimally short time. This cost increases in the timing of success. Intuitively, both players become more pessimistic as the project keeps failing, so the principal may need to raise the exploitation utility over time in order to compensate the agent for the lower likelihood of success.

The *intertemporal agency cost* is the expectation of the agent's all future rents from the possibility of success, when the players have common knowledge about the profitability of the project. This cost decreases in the timing of success. When $\lambda > \Delta_\lambda$, the agent enjoys a moral hazard rent at each time from the possibility of success even without exerting effort. Thus, as the time left until termination diminishes, so does the expectation of the agent's future rents as well.

Finally, the *informational agency cost* is the agent's expected information rent from any potential divergence in the players' beliefs. This cost is generally non-monotonic in the timing of success. Intuitively, when the players' beliefs π_t^P and π_t^A are sufficiently close to 1, they update slowly with respect to time. Therefore, at such times, the agent gains little from creating a difference in the two players' beliefs. Also, when there is little time left until termination, the probability of eventual success also becomes small, so the agent's expected future gain from any discrepancy in beliefs must approach zero.

5.2.2 Case 1) Model With Downward Sloping $V(\cdot)$

For each termination time $T > 0$, we study the constrained optimal incentive scheme when the principal's exploitation value function is assumed to be downward sloping. The following analog of Lemma 4 holds true when $\lambda > \Delta_\lambda$:

Lemma 9. *If the principal's exploitation value function is downward sloping, for each termination time $T > 0$, the incentive scheme $\mathcal{I}_T^{Binding}$ with binding temporary incentive compatibility constraints is constrained optimal.*

Proof. In the Appendix.[20] □

A heuristic argument is as follows. By hypothesis, the principal's exploitation profit increases in the agent's exploitation utility. Moreover, a straightforward generalization of Lemma 3-(1) shows that the incentive scheme $\mathcal{I}_T^{\text{Binding}}$ induces the minimal exploitation utility required to incentivize the agent throughout the exploration phase. In turn, for a given termination time $T > 0$, it is constrained optimal for the principal to implement the incentive scheme $\mathcal{I}_T^{\text{Binding}}$.

5.2.3 Case 2) Model With Inverted U-shaped $V(\cdot)$

As illustrated in Lemma 8, the exploitation utility $\underline{W}_t^S$ induced by binding temporary incentive compatibility constraints is not necessarily monotone in time. Therefore, we use an alternative way to derive the constrained optimal incentive scheme for a given termination time $T > 0$. In particular, for each termination time $T > 0$, we construct three candidates for the constrained optimal incentive scheme, and show that one of the candidates must be constrained optimal.

First Candidate Under the first candidate $\mathcal{I}_T^{\text{Type 1}}$ for the constrained optimal incentive scheme, the principal always promises the exploitation utility of W^* to the agent, if the project succeeds prior to termination time $T > 0$. Since W^* is the unique maximizer of $V(\cdot)$, no incentive scheme can deliver a strictly higher time-0 payoff to the principal than this first candidate $\mathcal{I}_T^{\text{Type 1}}$. Hence, $\mathcal{I}_T^{\text{Type 1}}$ must be a constrained optimal incentive scheme if and only if it satisfies the temporary incentive compatibility constraint at every $t \leq T$.

Second Candidate In our analysis of the model with $\lambda = \Delta_\lambda$, we modified the incentive scheme $\mathcal{I}_T^{\text{Binding}}$ with binding temporary incentive compatibility constraints so that the time-0 exploitation utility $W_0^{S*} \geq W^*$ under the constrained optimal incentive scheme. The next

[20] We prove both Lemmata 4 and 9 at the same time because the proof of the two Lemmata are effectively identical.

lemma suggests that in the model with $\lambda > \Delta_\lambda$, we need to make similar adjustments to the incentive scheme $\mathcal{I}_T^{\text{Binding}}$ to obtain a candidate constrained optimal incentive scheme:

Lemma 10. *In any constrained optimal incentive scheme, $V'(W_0^{S*}) \leq 0$. If the temporary incentive compatibility constraint is slack at time 0 (i.e., $\mu_0^{\text{TIC}*} = 0$), $V'(W_0^{S*}) = 0$.*

Proof. In the Appendix. □

To better understand Lemma 10, suppose, to the contrary, that $V'(W_0^{S*}) < 0$ in the optimal contract. If the principal raises the agent's exploitation utility at time 0, she can obtain a higher exploitation profit at time 0. This would slacken the temporary incentive constraint at time 0, and would have no effect on temporary incentive constraints in the future. Hence, since this contract is improvable, this contradicts the optimality hypothesis.

By Lemma 10, we consider an incentive scheme $\mathcal{I}_T^{\text{Type 2}}$ in which the agent earns the following time-t exploitation utility:

$$W_t^S := \begin{cases} W^* \text{ if } t \leq \underline{t}, \\ \underline{W}_t^S \text{ if } t > \underline{t}, \end{cases} \tag{23}$$

where $\underline{t}$ denotes the first time at which the exploitation utility $\underline{W}_t^S \geq W^*$ under the original incentive scheme $\mathcal{I}_T^{\text{Binding}}$ (if the exploitation utility $\underline{W}_t^S < W^*$, set $\underline{t} = \infty$). By arguining similarly as in Lemma 5, it is straightforward to see that the temporary incentive compatibility constraint is slack prior to $\underline{t}$ and binds afterwards.

Third Candidate It is also possible that the first candidate $\mathcal{I}_T^{\text{Type 1}}$ may fail to satisfy the incentive compatible constraint, and the second candidate $\mathcal{I}_T^{\text{Type 2}}$ is suboptimal. To see why, suppose that (a) the continuation plan $\{\underline{W}_t^S\}_{t \leq T}$ in the exploitation phase induced by $\mathcal{I}_T^{\text{Type 2}}$ is strictly decreasing over time as in Panel (a) of Figure 4; (b) the principal's exploitation value function $V(\cdot)$ increases steeply over $[0, W^*]$ and decreases gradually over $[W^*, \infty)$ with $\underline{W}_T^S < W^* < \underline{W}_0^S$, as in Panel (b) of Figure 4.

Under these assumptions, it is not incentive compatible for the agent to exert effort under the first candidate $\mathcal{I}_T^{\text{Type 1}}$. By Lemma 3-(1), it is straightforward to show that the agent's exploitation utility under any incentive incentive scheme must exceed $\underline{W}_0^S$. However, since the agent only earns $W^*(< \underline{W}_0^S)$ upon success at time 0, it cannot be sequentially rational for the agent to exert effort at time 0. Therefore, the first candidate $\mathcal{I}_T^{\text{Type 1}}$ cannot be a constrained optimal incentive scheme.

Moreover, the incentive scheme $\mathcal{I}_T^{\text{Type 2}}$ is suboptimal. By hypothesis, the principal's ex-

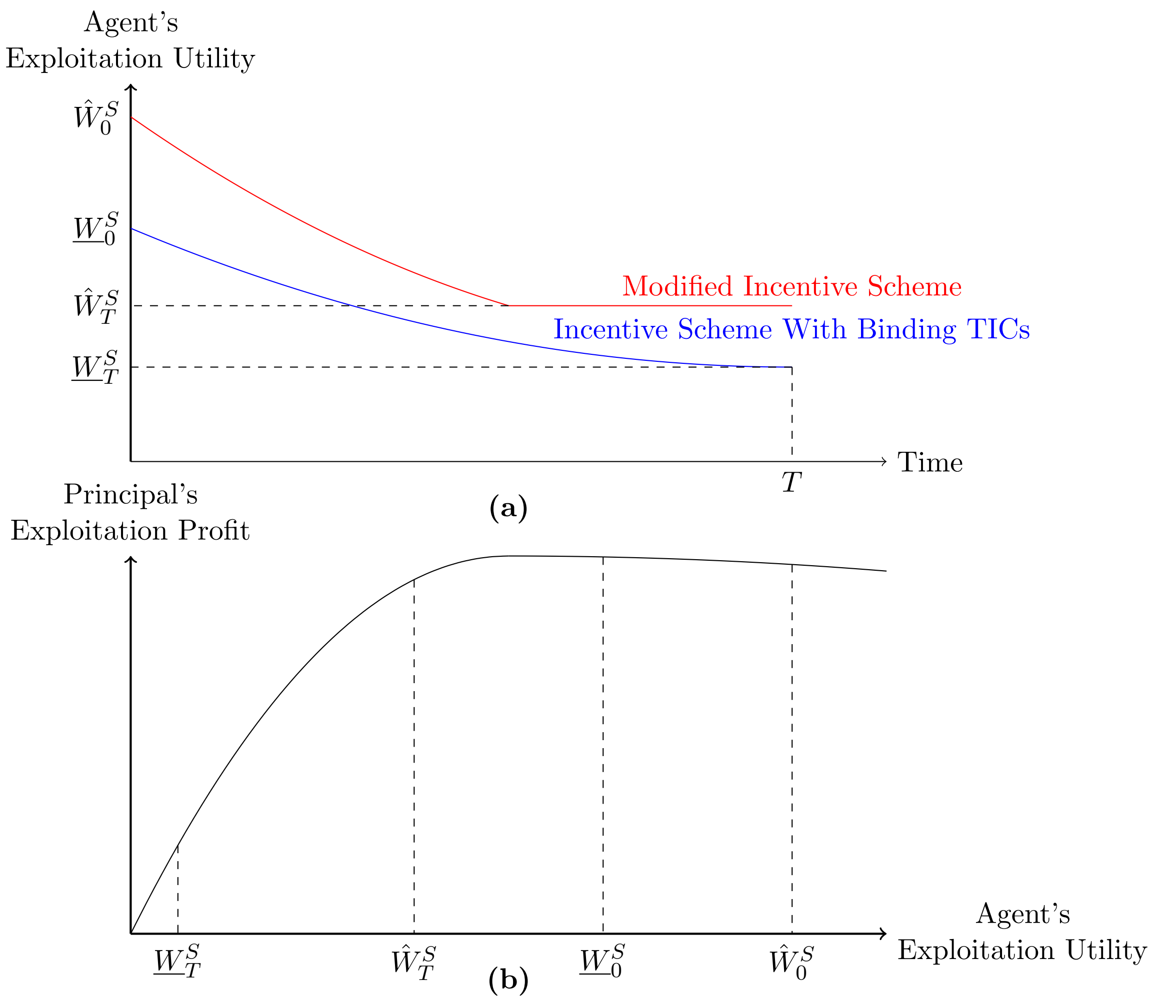


Figure 4: Case Where Neither $\mathcal{I}_T^{\text{Type 1}}$ nor $\mathcal{I}_T^{\text{Type 2}}$ is Constrained Optimal

ploitation value function increases steeply over $[0, W^*]$ and decreases gradually over $[W^*, \infty)$. Thus, as shown in Panel (b) of Figure 4, the principal's exploitation profit can become very low near the termination time T under the incentive scheme $\mathcal{I}_T^{\text{Type 2}}$. Therefore, the principal can instead modify the incentive scheme $\mathcal{I}_T^{\text{Type 2}}$ so that the temporary incentive compatibility constraint binds in the earlier part of the exploration phase, and is slack in the later part of the exploration phase. By doing so, the principal bears a relatively small cost from promising the agent an additional exploitation utility in the earlier part of the exploration phase, and reduces the cost in the later part of the exploration phase.

This case suggests that the principal can potentially benefit from further slackening the temporary incentive compatibility constraints. The next lemma shows that the temporary incentive compatibility constraint can be slack only if the principal can weakly gain from delivering additional exploitation utility to the agent:

Lemma 11. *In any optimal contract, $V'(W_t^{S*}) \geq 0$ for any time t at which the temporary incentive compatibility constraint is slack.*

Proof. In the Appendix. □

To see why Lemma 11 holds true, suppose, to the contrary, that the temporary incentive compatibility constraint is slack at some t with $V'(W_t^{S*}) < 0$ under the optimal incentive scheme. If the principal reduces the time-t exploitation utility W_t^{S*} by $\epsilon > 0$, the agent still has a sufficient incentive to exert effort at time t. Furthermore, the agent becomes more motivated to exert effort prior to time t because reducing the time-t exploitation utility makes it less attractive for the agent to wait until time t. Therefore, all the temporary incentive compatibility constraints are satisfied under this modified incentive scheme. However, the incentive scheme makes the principal strictly better off without violating the incentive compatibility condition, which contradicts the optimality hypothesis.

In the next proposition, we construct a constrained optimal incentive scheme for the case in which neither the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ nor $\mathcal{I}_T^{\text{Type 2}}$ is constrained optimal:

Proposition 4. *Suppose that (a) the principal's exploitation value function is inverted U-shaped, and (b) neither the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ nor $\mathcal{I}_T^{\text{Type 2}}$ is a constrained optimal incentive scheme. Then, there exists a constrained optimal incentive scheme $\mathcal{I}_T^{\text{Type 3}}$ such that for some time pair $(\underline{t}, \bar{t})$ with $0 \leq \underline{t} < \bar{t} < T$, the temporary incentive compatibility constraint binds over $[\underline{t}, \bar{t}]$ and is slack in $[0, \underline{t}) \cup (\bar{t}, T]$. Furthermore, there exists a $w_{slack} \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$ such that in the constrained optimal incentive scheme $\mathcal{I}_T^{\text{Type 3}}$, the agent's exploitation utility $W_t^{S*} = w_{slack}$ for all $t \in [\bar{t}, T]$.*

Proof. In the Appendix. □

In the constrained optimal incentive scheme in Proposition 4, the temporary incentive constraint binds for a positive duration of time, and becomes slack in the later part of the exploration phase. Moreover, if the temporary incentive constraint is slack prior to time $\underline{t} > 0$ under the incentive scheme, the agent's time-t exploitation utility under the incentive scheme must be equal to W^* for any $t < \underline{t}$.

The next lemma shows how the shape of the principal's exploitation value function affects the design of the constrained optimal incentive scheme for a given termination time $T > 0$. In particular, when the principal finds it more costly to lower the agent's exploitation utility (i.e., $V(\cdot)$ increases steeper over $[0, W^*)$, the temporary incentive constraint can be slack for a longer duration of time near termination:

Proposition 5. *Consider two inverted U-shaped exploitation value functions V_1, and V_2 such that 1) both are maximized at W^*, 2) V_2 has a strictly higher derivative than V_1 on $[0, W^*]$ and the same derivative as V_1 on $[W^*, \infty)$, and 3) are strictly concave on $\mathbb{R}_+$. Suppose that when the principal's exploitation value function is V_1, there exists $t_1 > 0$ such that the temporary incentive compatibility constraint binds on $[t_1 - \epsilon, t_1]$ and is slack over $(t_1, T]$ in a constrained optimal incentive scheme. Then, when the principal's exploitation value function is V_2, the constrained optimal incentive scheme under V_1 is no longer constrained optimal, and there exists $t_2 \in (0, t_1)$ such that the temporary incentive compatibility constraint is slack on $[t_2, T]$ in a constrained optimal incentive scheme.*

Proof. In the Appendix. □

6 Discussion

6.1 Implication For Termination Time

Let us explore how the shape of the principal's exploitation value function affects her optimal termination time. The next proposition shows that even if an inverted U-shaped value function is strictly lower than a downward-sloping value function at every point, the inverted U-shaped value function can still induce a longer termination time in the optimal contract than the downward sloping value function:

Proposition 6. *Suppose that when the principal's exploitation value function is given by $V_D(W_t^S) = y - W_t^S$ for some $y > 0$, it is optimal to terminate the relationship at time $T_D^* > 0$ in the exploration phase. Then, there exists an inverted U-shaped exploitation value function V_I such that 1) $V_I(W_t^S) < V_D(W_t^S) = y - W_t^S$ at every $W_t^S \geq 0$, and 2) when the principal's exploitation value function is given by $V_I(W_t^S)$, it is optimal to terminate the relationship at time $T_I^* > T_D^*$ in the exploration phase.*

Proof. In the Appendix. □

The intuition behind this unexpected result is as follows. When the principal increases termination time, the agent can potentially gain more from creating a divergence between the two players' belief, so the principal tends to bear an extra shadow cost from delivering a higher exploitation utility to the agent. If the principal has a downward sloping exploitation value function, the shadow cost is strictly positive because she finds it costly to deliver an additional exploitation utility to the agent. The positive shadow cost depresses her benefit from implementing a longer termination time, so the principal has an incentive to prematurely terminate the relationship.

By contrast, if the principal's exploitation value function is inverted U-shaped, she finds it less costly to deliver an additional utility to the agent. In turn, the principal incurs a lower shadow cost from delivering a higher exploitation utility to the agent, which reduces her incentive to prematurely terminate the relationship. Therefore, even when the principal earns a strictly lower exploitation profit at any exploitation utility promised to the agent, an inverted U-shaped exploitation value function often induces a longer termination time than a downward sloping value function.

6.2 Optimal Contract in Absence of Learning

Learning is a natural feature of many real-world applications of multi-stage projects (e.g., venture capital), so we have assumed that the players are initially uncertain of project profitability (i.e., $\pi_0 \in (0,1)$). Yet, more generally, our model can be thought of as a model with two phases, and the role played by the payoff structure in the second phase may be more critical than the one played by learning in the first phase. To explore this possibility, let us consider a model in which the players are no longer uncertain about profitability in the first phase.

Lemma 12. *Suppose that 1) both players are initially certain that the project is good in the first phase (i.e., $\pi_0 = 1$), and 2) W^* is the unique maximizer of the value function $V(\cdot)$ in the second phase. Then, in the optimal contract, the relationship is never terminated in the exploration phase, and the principal delivers the exploitation utility* $\max\{\frac{c}{\lambda}, W^*\}$ *to the agent upon the breakthrough.*

Proof. In the Appendix. □

Lemma 12 shows that the temporary incentive compatibility constraint may be slack under the optimal contract, even in the absence of learning. In particular, when the principal's value function in the second phase is inverted U-shaped with a sufficiently large maximizer W^*, the temporary incentive compatibility constraint is always slack throughout the first phase. In contrast, when the principal's value function in the second phase is downward sloping (i.e., $W^* = 0$), the temporary incentive compatibility constraint must bind at all times in the first phase.

Nevertheless, one important difference is that without learning, the relationship is never terminated in the first phase. Therefore, the model without learning cannot speak to the duration of experimentation in the first phase.

7 Conclusion

This paper examines a dynamic moral hazard problem involving an initial exploration phase followed by an exploitation phase, which allows us to integrate the experimentation literature with the dynamic corporate finance literature. Our findings highlight how the shape of the exploitation value function can determine optimal contract design and overall welfare. With a downward-sloping exploitation value function, the temporary incentive constraints must always bind, leading to additional efficiency losses from under-experimentation. In contrast, when the principal's exploitation value function is inverted U-shaped, the agent's temporary incentive compatibility constraints may become slack during the exploration phase in the optimal contract. This relaxation may occur over the entire exploration phase, and may reduce inefficiencies from under-experimentation. Surprisingly, even when the principal's attainable profits are lower with an inverted U-shaped value function, the principal may still induce more experimentation than when she has a downward sloping value function.

One fruitful extension is to consider a setting in which a bad project can potentially succeed with strictly positive probability. In such a model, the first success does not fully reveal the profitability of the project, and the players' equilibrium posterior belief about profitability keeps updating indefinitely. Therefore, solving this model is likely to involve keeping track of the equilibrium posterior belief as an additional state variable, which considerably complicates the analysis. Furthermore, there would no longer be a distinction between the exploration phase and the exploitation phase, which is the focus of this paper. That being said, we believe that many of our qualitative conclusions will continue to hold in this extension. We leave for future research to verify this conjecture.

References

Dirk Bergemann and Ulrich Hege. Venture capital financing, moral hazard, and learning. *Journal of Banking & Finance*, 22(6-8):703–735, 1998.

Dirk Bergemann and Ulrich Hege. The financing of innovation: Learning and stopping. *RAND Journal of Economics*, pages 719–752, 2005.

Bruno Biais, Thomas Mariotti, Jean-Charles Rochet, and Stéphane Villeneuve. Large risks, limited liability, and dynamic moral hazard. *Econometrica*, 78(1):73–118, 2010.

Pierre Brémaud. *Point processes and queues: martingale dynamics*, volume 66. Springer, 1981.

Michael Ralph Caputo. *Foundations of dynamic economic analysis: optimal control theory and applications*. Cambridge University Press, 2005.

Gian Luca Clementi and Hugo A Hopenhayn. A theory of financing constraints and firm dynamics. *The Quarterly Journal of Economics*, 121(1):229–265, 2006.

Peter M DeMarzo and Michael J Fishman. Agency and optimal investment dynamics. *The Review of Financial Studies*, 20(1):151–188, 2007a.

Peter M DeMarzo and Michael J Fishman. Optimal long-term financial contracting. *The Review of Financial Studies*, 20(6):2079–2128, 2007b.

Peter M DeMarzo and Yuliy Sannikov. Optimal security design and dynamic capital structure in a continuous-time agency model. *The journal of Finance*, 61(6):2681–2724, 2006.

Peter M DeMarzo and Yuliy Sannikov. Learning, termination, and payout policy in dynamic incentive contracts. *The Review of Economic Studies*, 84(1):182–236, 2017.

Peter M DeMarzo, Michael J Fishman, Zhiguo He, and Neng Wang. Dynamic agency and the q theory of investment. *The journal of Finance*, 67(6):2295–2340, 2012.

Robert Dorfman. An economic interpretation of optimal control theory. *The American Economic Review*, 59(5):817–831, 1969.

Ana Fernandes and Christopher Phelan. A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory*, 91(2):223–247, 2000.

William Fuchs, Brett Green, and David Levine. Optimal arrangements for distribution in developing markets: Theory and evidence. *American Economic Journal: Microeconomics*, 14(1):411–450, 2022.

Paul Gompers. Optimal investment, monitoring, and the staging of venture capital. *Journal of Finance*, 50(5):1461–89, 1995.

Brett Green and Curtis R Taylor. Breakthroughs, deadlines, and self-reported progress: Contracting for multistage projects. *American Economic Review*, 106(12):3660–3699, 2016.

Edward J Green. Lending and the smoothing of uninsurable income. *Contractual arrangements for intertemporal trade*, 1:3–25, 1987.

Sanford J. Grossman and Oliver D. Hart. An analysis of the principal-agent problem. *Econometrica*, 51(1):7, 1983.

Marina Halac, Navin Kartik, and Qingmin Liu. Optimal contracts for experimentation. *The Review of Economic Studies*, 83(3):1040–1091, 2016.

Zhiguo He, Bin Wei, Jianfeng Yu, and Feng Gao. Optimal long-term contracting with learning. *The Review of Financial Studies*, 30(6):2006–2065, 2017.

Johannes Hörner and Larry Samuelson. Incentives for experimenting agents. *The RAND Journal of Economics*, 44(4):632–663, 2013.

Johannes Hörner. *Comments on "Contracts"*, volume 1 of *Econometric Society Monographs*, page 172–176. Cambridge University Press, 2013. doi: 10.1017/CBO9781139060011.007.

Marek Kapička. Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach. *Review of Economic Studies*, 80(3):1027–1054, 2013.

Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

Fahad Khalil, Jacques Lawarree, and Alexander Rodivilov. Learning from failures: Optimal contracts for experimentation and production. *Journal of Economic Theory*, 190:105107, 2020.

Josh Lerner. Venture capitalists and the oversight of private firms. *The Journal of Finance*, 50(1):301–318, 1995.

L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *Mathematical theory of optimal processes.* John Wiley and Songs, 1962.

Vincenzo Quadrini. Investment and liquidation in renegotiation-proof contracts with moral hazard. *Journal of Monetary Economics*, 51(4):713–751, 2004.

William A Sahlman. The structure and governance of venture-capital organizations. *Journal of Financial Economics*, 27(2):473–521, 1990.

Yuliy Sannikov. A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, 75(3):957–984, 2008.

Galina Vereshchagina and Hugo A Hopenhayn. Risk taking by entrepreneurs. *American Economic Review*, 99(5):1808–1830, 2009.

Appendix

Proof for Lemma 2. Consider an alternative strategy $\hat{a} := \{\hat{a}_t\}_{t\geq 0}$ such that the agent may potentially deviate from the original effort recommendation up to time $\hat{t}\wedge\tau$ (i.e., for some $t \in [0, \hat{t}\wedge\tau]$, $\hat{a}_t$ may potentially differ from $a_t = 1$), and exerts effort afterward in the exploration phase, i.e., $\hat{a}_t = 1$ for all $t \in [\hat{t}\wedge\tau, T]$. Let $\hat{\pi}_t^A$ denote the agent's posterior belief at time t that $\theta = G$ induced by the strategy $\hat{a}$.

Holding the project type fixed at $\theta \in \{G, B\}$ and the agent's strategy fixed at its equilibrium level a after time $\hat{t}$, the past history of the agent's deviation has no impact on the probability distribution over the future play after time $\hat{t}$. Therefore, the agent's continuation utility from the alternative action strategy $\hat{a}$ up to and including time $\hat{t}$, and switching back to the original recommendation after time t can be written as:

$$\hat{W}_{\hat{t}}(\Gamma, \hat{a}) = \hat{\pi}_{\hat{t}}^A W_{\hat{t}}^G(\Gamma, a) + (1 - \hat{\pi}_{\hat{t}}^A) W_{\hat{t}}^B(\Gamma, a), \tag{24}$$

where $W_t^B(\Gamma, a) := c\left(\frac{1-\exp(-r(T-t))}{r}\right)$ is the agent's equilibrium continuation value at time t in case Nature chooses a bad project ($\theta = B$), for any $t \leq T$. Also, the time-$\hat{t}$ expectation of the agent's lifetime discounted payoff under the strategy $\hat{a}$ can be expressed as:

$$\hat{U}_{\hat{t}}(\Gamma, \hat{a}) = -\int_0^{\hat{t}\wedge\tau} \exp(-rs)c\hat{a}_s ds + \exp(-r(\hat{t}\wedge\tau))\hat{W}_{\hat{t}\wedge\tau}(\Gamma, a). \tag{25}$$

Similarly, let $\overline{\pi}_t^A$ denote the agent's posterior belief at time t that $\theta = G$ induced by the recommended strategy a. Then, the time-$\hat{t}$ expectation of the agent's equilibrium lifetime discounted payoff can be expressed as:

$$U_{\hat{t}}(\Gamma, a) = -\int_0^{\hat{t}\wedge\tau} \exp(-rs)c\hat{a}_s ds + \exp(-r(\hat{t}\wedge\tau))W_{\hat{t}\wedge\tau}(\Gamma, a).$$

where $W_t(\Gamma, a) := \overline{\pi}_t^A W_t^G(\Gamma, a) + (1 - \overline{\pi}_t^A) W_t^B(\Gamma, a)$ represents the agent's time-t equilibrium continuation value for each $t \leq T$. Furthermore, by Equation (24) and Lemma 1, we can show:

$$U_{\hat{t}}(\Gamma, a) = U_0(\Gamma, a) + \int_0^{\hat{t}\wedge\tau} \exp(-rs)(\beta_s^G + W_{s^-}^G - W_{s^-})\Big\{dN_s - \Lambda_s \pi_s^P ds\Big\}$$

Hence, the posterior lifetime discounted payoff given in (25) can be rewritten as:

$$\hat{U}_{\hat{t}}(\Gamma,\hat{a}) = \underbrace{\int_0^{\hat{t}} \exp(-rs)c\{a_s - \hat{a}_s\}ds + \exp(-r\hat{t})\Big(\hat{W}_{\hat{t}}(\Gamma,a) - W_{\hat{t}}(\Gamma,a)\Big)}_{=\hat{U}_{\hat{t}}(\Gamma,\hat{a})-U_{\hat{t}}(\Gamma,a) \text{ by Equation (3)}} \\ + \underbrace{U_0(\Gamma,a) + \int_0^{\hat{t}} \exp(-rs)(\beta_s^G + W_{s^-}^G - W_{s^-})\Big\{dN_s - \lambda\pi_s^P ds\Big\}}_{=U_{\hat{t}}(\Gamma,a)} \tag{26}$$

whenever $\hat{t} \leq \tau$. Furthermore, we can show

$$\begin{aligned} &\exp(-r(\hat{t}\wedge\tau))\Big(\hat{W}_{\hat{t}\wedge\tau}(\Gamma,a) - W_{\hat{t}\wedge\tau}(\Gamma,a)\Big) \\ =&\int_0^{\hat{t}\wedge\tau} \exp(-rs)\Bigg\{\Big(\Lambda_s(\pi_{s^-}^P - \hat{\pi}_{s^-}^A)\beta_s^G - \hat{\Lambda}_s\hat{\pi}_{s^-}^A(W_{s^-}^G - \hat{W}_{s^-}) + \Lambda_s\pi_{s^-}^P(W_{s^-}^G - W_{s^-})\Big)ds \\ &\qquad + (W_{s^-} - \hat{W}_{s^-})dN_s\Bigg\}, \end{aligned} \tag{27}$$

where $W_{t^-} := \pi_{t^-}^P W_{t^-}^G + (1-\pi_{t^-}^P)W_{t^-}^B$, and $\hat{W}_{t^-} := \hat{\pi}_{t^-}^A W_{t^-}^G + (1-\hat{\pi}_{t^-}^A)W_{t^-}^B$ without a slight abuse of notation. The equality follows from Ito's product rule for point processes, the definitions of W_t and $\hat{W}_t$, and the laws of motion of π_t^P and $\hat{\pi}_t^A$.

By the law of iterated expectations (henceforth "LIE"), the agent expects to earn the following lifetime discounted value under the alternative strategy profile $\hat{a}$:[21]

$$\begin{aligned} \hat{U}_0(\Gamma,\hat{a}) \underbrace{=}_{\text{"LIE"}} \mathbb{E}_{\hat{a}}\Big(\hat{U}_{\hat{t}}(\Gamma,\hat{a})\Big) =& U_0(\Gamma,a) + \underbrace{\mathbb{E}_{\hat{a}}\Big(\int_0^{\hat{t}\wedge\tau} \exp(-rs)(\beta_s^G + W_{s^-}^G - \hat{W}_{s^-})\Big\{dN_s - \hat{\Lambda}_s\hat{\pi}_{s^-}^A ds\Big\}\Big)}_{(*)} \\ &+\mathbb{E}_{\hat{a}}\Big(\int_0^{\hat{t}\wedge\tau} \exp(-rs)\Big(\lambda\hat{\pi}_{s^-}^A\beta_s^G - c\Big)(\hat{a}_s - a_s)ds\Big), \end{aligned}$$

where the second equality follows from equations derived in (26) and (27). Since the process $\{\beta_t^G + W_{t^-}^G - \hat{W}_{t^-}\}_{t\geq 0}$ is $\mathscr{F}$-predictable and the compensated process $\{N_t - \int_0^t \hat{\Lambda}_s ds\}_{t\geq t_0}$ is an $\mathscr{F}$-martingale under the probability measure $\mathbb{P}_{\hat{a}}$ and has zero mean, the optional sampling theorem implies that the underbraced expression $(*)$ is equal to zero. Hence,

$$\hat{U}_0(\Gamma,\hat{a}) = U_0(\Gamma,a) + \mathbb{E}_{\hat{a}}\Big(\int_0^{\hat{t}\wedge\tau} \exp(-rs)\Big(\lambda\hat{\pi}_{s^-}^A\beta_s^G - c\Big)(\hat{a}_s - a_s)ds\Big). \tag{28}$$

[21] Here, we make use the following fact: for any stochastic process such that $X_t = X_\tau$ for some $t > \tau$, we can write $X_t = X_0 + \int_0^{t\wedge\tau} dX_s$ for all $t \geq 0$.

Suppose that $\pi_{t^-}^P \beta_t^G \geq \frac{c}{\Delta_\lambda}$ for almost every $t \leq \tau$. Since the original effort recommendation is front-loaded, the agent is weakly more optimistic than the principal (i.e., $\pi_t^A \geq \pi_t^P$) after any private history. This implies that for almost every $t \leq \tau$, $\hat{\pi}_{t^-}^A \beta_t^G \geq \frac{c}{\Delta_\lambda}$ after any private history that can be supported under the probability measure $\mathbb{P}_{\hat{a}}$ induced by the alternative strategy profile $\hat{a}$. Hence, for almost every $t \leq \tau$, the integrand with respect to Lebesgue measure in the last line of (28) is weakly negative $\mathbb{P}_{\hat{a}}$-almost surely. By the optional sampling theorem, the expectation with respect to $\mathbb{P}_{\hat{a}}$ in the last line of (28) is weakly negative. Since our choice of $\hat{t}$, and the agent's alternative action strategy was arbitrary, it must be incentive compatible for the agent to exert effort prior to termination after any private history. This establishes sufficiency.

Alternatively, suppose that there is strictly positive probability that $\pi_{t^-}^P \beta_t^G < \frac{c}{\Delta_\lambda}$ for some $t \leq \tau$ on a time set of positive Lebesgue measure. We shall construct an alternative strategy profile $\hat{a}$ so that the integrand inside the expectation in the second line of (28) is strictly positive with strictly positive probability and nonnegative almost surely under the measure $\mathbb{P}_a$ induced by the equilibrium strategy a.

Choose $\hat{t}$ to be any time at which there is strictly positive probability $\mathbb{P}_a$ that $\pi_{t^-}^P \beta_t^G < \frac{c}{\lambda}$ for some $t \in [0, \hat{t}]$ on a set of positive Lebesgue measure. Then, there exists $\epsilon > 0$ such that $\pi_{t^-}^P \beta_t^G < \frac{c}{\lambda} - \epsilon$ for some $t \in [0, \hat{t}]$ on a set of positive Lebesgue measure with strictly positive probability under the probability measure $\mathbb{P}_a$. Fix any such $\epsilon > 0$.

For a fixed $\eta > 0$, consider an alternative strategy $\hat{a}$ under which the agent shirks at any time $t \in [\hat{t} - \eta, \hat{t}]$ with $\pi_{t^-}^P \beta_t^G < \frac{c}{\lambda} - \epsilon$, and exerts effort otherwise. Clearly, for any β_t^G, the function $\hat{\pi}_{t^-}^A \beta_t^G$ is differentiable with respect to $\hat{\pi}_{t^-}^A$. The differentiability implies that we can find a sufficiently small $\eta = \eta(\epsilon) > 0$ such that (1) for any $t \in [0, \hat{t}]$ with $\lambda \hat{\pi}_{t^-}^A \beta_t^G \geq c$, $\hat{a}_t = a_t = 1$; (2) there exists a time set of positive Lebesgue measure with $\lambda \hat{\pi}_{t^-}^A \beta_t^G < c$ and $\hat{a}_t = 0 \neq a_t = 1$. Thus, we can show that when $\eta = \eta(\epsilon) > 0$ is sufficiently small, the integrand inside the expectation given in the second line of (28) is strictly positive with strictly positive probability on a time set of positive Lebesgue measure, and nonnegative almost everywhere almost surely under the probability measure $\mathbb{P}_{\hat{a}}$ induced by the alternative strategy profile $\hat{a}$. This completes the proof. □

Proof of Lemma 3. We start by proving the first part (1). For a given termination time $T > 0$, fix an arbitrary incentive scheme $\{(\beta_t^G, W_t^G)\}_{t \leq T}$ that satisfies both Constraints (PKG) and

(TIC) at any $t \leq T$. By the integral expression given in (7), we have:

$$W_t^G := \int_t^T \exp(-r(s-t))\Big(\lambda\beta_s^G - c\Big)ds. \underbrace{\geq}_{\text{(TIC)}} \underbrace{\int_t^T \exp(-r(s-t))\Big(\lambda\underline{\beta}_s^G - c\Big)dt}_{=:\underline{W}_t^G}. \tag{29}$$

for any $t \leq T$. Combining the inequality (29) with the temporary incentive compatibility constraint (TIC), we have:

$$\underbrace{W_t^G + \beta_t^G}_{=:W_t^S} \geq \underbrace{\underline{W}_t^G + \underline{\beta}_t^G}_{=:\underline{W}_t^S} \tag{30}$$

for any $t \leq T$. This completes the proof of the first part (1).

We proceed by proving the second part (2). Taking partial differentiation of $\underline{W}_t^S$ with respect to time yields:

$$\begin{aligned}\frac{\partial \underline{W}_t^S}{\partial t} \underbrace{=}_{(13)} \frac{\partial \underline{W}_t^G}{\partial t} + \frac{\partial \underline{\beta}_t^G}{\partial t} &= \underbrace{r\underline{W}_t^G - \lambda\underline{\beta}_t^G + c}_{=\frac{\partial \underline{W}_t^G}{\partial t} \text{ by (PKG)}} + \underbrace{\Big(-\lambda\pi_t^P(1-\pi_t^P)\Big)}_{=\frac{\partial \pi_t^P}{\partial t} \text{ by (8)}} \times \underbrace{\Big(-\frac{c}{\lambda(\pi_t^P)^2}\Big)}_{=\frac{\partial \underline{\beta}_t^G}{\partial \pi_t^P}} \\ &= r\underline{W}_t^G > 0\end{aligned} \tag{31}$$

for all $t < T$, where the equality in the last line follows from direct computations, and the inequality in the last line follows from the fact that $\lambda\underline{\beta}_t^G > c$ and the definition of $\underline{W}_t^G$ in (13). This completes the proof. □

Proof of Lemmata 4 and 9. By Proposition 3, it suffices to check that the following expression is weakly positive:

$$\mu_t^{\text{TIC}*} := -\lambda V'(\underline{W}_t^S) - \int_0^t \exp(-\lambda t)\lambda^2 V'(\underline{W}_\tau^S)d\tau.$$

Since the principal's exploitation value function $V(\cdot)$ is downward sloping, it has a strictly negative derivative everywhere, which implies that $\mu_t^{\text{TIC}*} \geq 0$ for all $t \geq 0$. This implies that the multiplier process pair $(\mu^{\text{PKG}*}, \mu^{\text{TIC}*})$ constructed above satisfy all the Kuhn-Tucker conditions in Proposition 3, which completes the proof. □

Proof of Lemma 5. We first derive alternative expressions for contracting variables associated with the candidate optimal incentive scheme $\mathcal{I}_T^*$, and analogous expressions associated with the incentive scheme $\mathcal{I}_T^{\text{Binding}}$. Let W_t^{G*} denote the agent's time-t continuation utility induced by $\mathcal{I}_T^*$, evaluated after failure up to and including time t under the probability measure $\mathbb{P}_a^G$. Also, define β_t^{G*} to be the sensitivity of the agent's time-t continuation utility induced by $\mathcal{I}_T^*$, when the payoff is evaluated based on the probability measure $\mathbb{P}_a^G$. For each $t \leq T$, we have:

$$W_t^{G*} = \int_t^T \exp(-(r+\lambda)(\tau-t))(\lambda W_\tau^{S*} - c)d\tau, \ \beta_t^{G*} = W_t^{S*} - W_t^{G*}. \tag{32}$$

Note that the first inequality follows from the definition given in (3) and the front-loaded nature of the effort recommendation, whereas the second equality follows from the promise-keeping constraint under the probability measure $\mathbb{P}_a^G$ (6). Arguing similarly, we can obtain the following expressions for $\underline{W}_t^G$ and $\underline{\beta}_t^G$:

$$\underline{W}_t^G = \int_t^T \exp(-(r+\lambda)(\tau-t))\Big(\lambda \underline{W}_\tau^S - c\Big)d\tau, \ \underline{\beta}_t^G = \underline{W}_t^S - \underline{W}_t^G. \tag{33}$$

Let us proceed by showing that the temporary incentive compatibility constraint is slack at time t whenever $\underline{W}_t^S < W^*$. By definition of W_t^{S*}, we have:

$$W_t^{S*} - \underline{W}_t^S = \max\{W^* - \underline{W}_t^S, 0\}.$$

Since the agent's exploitation utility $\underline{W}_t^S$ strictly increases in time t by Lemma 3-(2), the difference $W_t^{S*} - \underline{W}_t^S$ between the agent's exploitation utilities must be weakly decreasing in time t. Moreover, we must have:

$$\begin{aligned}
\beta_t^{G*} - \underline{\beta}_t^G =& (W_t^{S*} - \underline{W}_t^S) - \lambda \int_t^T \exp(-(r+\lambda)(\tau-t))\Big(W_\tau^{S*} - \underline{W}_\tau^S\Big)d\tau. \\
\geq& (W_t^{S*} - \underline{W}_t^S) - \lambda \int_t^T \exp(-(r+\lambda)(\tau-t))d\tau \times (W_t^{S*} - \underline{W}_t^S). \\
\geq& \frac{r}{\lambda+r}(W^* - \underline{W}_t^S) > 0.
\end{aligned} \tag{34}$$

for any t with $\underline{W}_t^S < W^*$. The equality in the first line follows from the difference between β_t^{G*} and $\underline{\beta}_t^G$ in (32) and (33), whereas the inequality in the second line follows from the fact that the difference $W_t^{S*} - \underline{W}_t^S$ is weakly decreasing in time t. The inqualities in the last line follow from the facts that (1) the integral in the second line is less than $\int_t^\infty \exp(-(r+\lambda)(\tau-t))d\tau = \frac{1}{r+\lambda}$ and (2) $W_t^{S*} = W^*$ whenever $\underline{W}_t^S < W^*$. The chain of inequalities above show that $\beta_t^{G*} > \underline{\beta}_t^G := \frac{c}{\lambda \pi_t^P}$, so the temporary incentive compatibility constraint is slack at time t.

It remains to show that the temporary incentive compatibility constraint is slack at any time t whenever $\underline{W}_t^S \geq W^*$. By the definition given in (14), $W_t^{S*} = \underline{W}_t^{S*}$. Furthermore, by Lemma 3-(2), $W_\tau^{S*} = \underline{W}_\tau^S$ for any $\tau \in [t, T]$. Therefore, by (32) and (33), we must have

$\beta_t^{G*} = \underline{\beta}_t^G := \frac{c}{\lambda \pi_t^P}$, which completes the proof. $\square$

Proof of Proposition 2. Since Lemma 5 shows the slackness of the incentive scheme $\mathcal{I}_T^*$ at any time with $\underline{W}_t^S < W^*$, it remains to establish its optimality. It suffices to show that $V(W_t^{S*}) \geq V(W_t^S)$, where W_t^S represents the time-t exploitation utility induced by an arbitrary incentive scheme for a given termination time $T > 0$.

Let us consider two cases: 1) $\underline{W}_t^S < W^*$; 2) $\underline{W}_t^S \geq W^*$. If $\underline{W}_t^S < W^*$, we must have $W_t^{S*} = W^*$ by construction. Since the value function $V(\cdot)$ is uniquely maximized at W^* by hypothesis (Assumption 0-(ii)), we must have $V(W_t^{S*}) = V(W^*) \geq V(W_t^S)$ in this case. Alternatively, assume that $\underline{W}_t^S \geq W^*$. Therefore, we must have $\underline{W}_t^{S*} = \underline{W}_t^S$, which represents the minimal level of the agent's exploitation utility upon success at time t by Lemma 3-(1). Moreover, since $V(\cdot)$ is a concave function with a unique maximizer at W^* by Assumption 0 and $W^*(\geq \underline{W}_t^S)$ by hypothesis, the function $V(\cdot)$ must be downward sloping in $[\underline{W}_t^S, \infty)$. This implies that the maximizer of $V(\cdot)$ over the set $[\underline{W}_t^S, \infty)$ is $\underline{W}_t^S$, which completes the proof of Proposition 2. $\square$

Proof of Theorem 1. We begin by proving that the contract described in Theorem 1 yields the principal a weakly higher payoff than any other incentive-compatible contracts. Consider a relaxed version $(*)$ of the implementation subprogram (16) without temporary incentive compatibility constraints:

$$\underbrace{\max_{\substack{(\beta^G, W^G) \\ \text{s.t. (PKG)}}} \int_0^T \lambda \exp(-(r+\lambda)\tau) V(W_\tau^G + \beta_\tau^G) d\tau}_{(*)} = \left(\frac{1 - \exp(-(\lambda + r)T)}{\lambda + r} \right) \lambda V(W^*) \geq J(T)$$

for any $T \geq 0$. The equality above follows from the fact that W^* is the unique maximizer of $V(\cdot)$ and direct computation, whereas the inequality immediately follows from the fact that the new subprogram $(*)$ is a relaxed version of the implementation subprogram in (11).

By the inequality above, the following expression must be weakly greater than the objective function given in the optimization program (11):

$$\pi_0 \left(\frac{1 - \exp(-(r+\lambda)T)}{r+\lambda} \right) \Big(\lambda V(W^*) - i \Big) - (1 - \pi_0) \Big(\frac{1 - \exp(-rT)}{r} \Big) i \tag{35}$$

for any given $T \geq 0$. Verbally, the expression given in (35) represents the principal's maximal expected discounted payoff from a contract (which need not be incentive compatible in the exploration phase) with termination time T.

Since the principal's expected discounted payoff in (11) is strictly concave in termination time T, we can use the following first-order condition to characterize the optimal termination time T^*_{Relaxed} that maximizes the expected discounted payoff in (11) with respect to T:

$$\lambda \pi_{T^*_{\text{Relaxed}}} V(W^*) = i. \tag{36}$$

The first-order condition above asserts that in the relaxed contracting problem $(*)$ without the incentive compatibility constraint, it would be optimal for the principal to terminate the relationship as soon as her equilibrium belief π_t^P that the project is good reaches $\underline{\pi}$. Since this contract solves the relaxed program $(*)$, it delivers the principal an expected discounted payoff weakly higher than any incentive compatible contract. Furthermore, this contract also coincides with the one described in Proposition 1. This establishes the first part of the proof.

We proceed by verifying that under the contract described in Theorem 1, the temporary incentive compatibility constraint is always slack throughout the exploration phase. Let us invoke the definition given in (3) to compute the agent's time-t continuation utility from this contract under the probability measure $\mathbb{P}_a^G$:

$$\begin{aligned} W_t^G &:= \mathbb{E}_a^G \left[\exp(-r(\tau - t))W^* \times \mathbf{1}_{\{\tau \leq T^*\}} - c \int_t^{\tau \wedge T^*_{\text{Relaxed}}} \exp(-r(s-t))ds \middle| \mathscr{F}_t \right], \\ &= \left(\frac{1 - \exp(-(\lambda + r)(T^*_{\text{Relaxed}} - t))}{\lambda + r} \right) (\lambda W^* - c) \end{aligned} \tag{37}$$

for any $t \leq T^*_{\text{Relaxed}}$.

By the promise-keeping constraint (PKG) under the probability measure $\mathbb{P}_a^G$, we can calculate the sensitivity term β_t^G under this contract:

$$\begin{aligned} \beta_t^G :=& W^* - W_t^G \\ =& \frac{r}{r+\lambda} W^* + \underbrace{\left(\frac{\lambda}{\lambda + r}\right) \exp(-(\lambda + r)(T^*_{\text{Relaxed}} - t))W^* + \left(\frac{1 - \exp(-(\lambda + r)(T^*_{\text{Relaxed}} - t))}{\lambda + r} \right) c}_{>0} \\ \geq& \frac{r}{r+\lambda} W^* > \frac{c}{\lambda \underline{\pi}}. \end{aligned}$$

for any $t \leq T^*_{\text{Relaxed}}$, where the last inequality follows from the initial hypothesis given in Theorem 1. The inequality chain above shows that in the contract described in Theorem 1,, all the temporary incentive compatibility constraints are slack throughout the exploration phase. This completes the proof. □

Proof of Lemma 6. We consider the following auxiliary optimization program with an exogenously fixed level of the agent's time-0 expected discounted utility:

$$J(w,T) := \max_{\substack{(\beta^G, W^G) \text{ s.t.} \\ \text{(PKG),(TIC),} \\ W_0^G = w}} \int_0^T \lambda \exp(-(r+\lambda)\tau) V(W_\tau^G + \beta_\tau^G) d\tau. \tag{38}$$

for each pair of termination time T and the agent's time-0 expected discounted utility w under the probability measure $\mathbb{P}_a^G$.

As the first step, we solve the auxiliary program (38). Since the exploitation value function $V(\cdot)$ is concave and the constraint functions are linear in β_t^G and W_t^G, Mangasarian sufficiency theorem[22] implies that the incentive scheme pair (β^{G*}, W^{G*}) that satisfies (PKG), (TIC), $W_0^{G*} = w$ solves Subprogram (38) if and only if there exists a pair of multiplier processes $(\mu^{\text{PKG}*}, \mu^{\text{TIC}*}) = \{(\mu_t^{\text{PKG}*}, \mu_t^{\text{TIC}*})\}_{t \leq T}$ satisfying requirements given by (FOC), (CE), (CSC_1), and (CSC_2) for a given pair (w, T).

As the second step, let us optimize over the agent's time-0 expected discounted payoff w. By Lemma 3-(1), $\underline{W}_0^G$ is the infimum of the time-0 expected discounted payoff that can be promised to the agent, thus implying the additional constraint $w \geq \underline{W}_0^G$. Moreover, since the exploitation value function $V(\cdot)$ is concave and the constraint set is convex, the function $J(w,T)$ derived from (38) must be concave in w. Therefore, the following Kuhn-Tucker conditions are both sufficient and necessary conditions that characterize the optimal level of the agent's time-0 expected discounted payoff w^*:

$$\frac{\partial J(w^*, T)}{\partial w} \leq 0, \ \frac{\partial J(w^*, T)}{\partial w}(w^* - \underline{W}_0^G) = 0$$

By the dynamic envelope theorem,[23] we must have $\frac{\partial J(w^*,T)}{\partial w} = \mu_0^{\text{PKG}*}$ when the latter is evaluated at $w = w^*$. Substituting this expression into the Kuhn-Tucker conditions above establishes (CSC_1). If the exploitation value function is assumed to be strictly concave, it also follows from Mangasarian sufficiency theorem that the solution described above is unique, which completes the proof. □

Proof of Lemma 7. Suppose that there exists an optimal quadruple $(\beta^{G*}, W^{G*}, \mu^{\text{PKG}*}, \mu^{\text{TIC}*})$ satisfying conditions (PKG), (TIC), (FOC), (CE), (CSC_1), and (CSC_2). By Equations (FOC), and (CE), we have:

$$\mu_t^{\text{PKG}*} = \exp(\lambda t)\mu_0^{\text{PKG}*} - \int_0^t \exp(-\lambda(s-t)) V'(W_s^S) dt, \ \mu_t^{\text{TIC}*} = -\lambda V'(W_t^S) + \lambda \mu_t^{\text{PKG}*}.$$

[22]See, for example, Theorem 6.2 in Caputo (2005).
[23]See, for example, Theorem 9.3 in Caputo (2005).

Define

$$\tilde{\mu}_t^{\text{PKG}*} := -\int_0^t \exp(-\lambda(s-t))V'(W_s^S)dt$$
$$\tilde{\mu}_t^{\text{TIC}*} := -\lambda V'(W_t^S) + \lambda\tilde{\mu}_t^{\text{PKG}*} = \mu_t^{\text{TIC}*} - \lambda\exp(\lambda t)\mu_0^{\text{PKG}*} \geq 0,$$

where the last inequality follows $\mu_t^{\text{TIC}*} \geq 0$ and $\mu_0^{\text{PKG}*} \leq 0$ by complementary slackness conditions (CSC_1) and (CSC_2). Straightforward calculations show that $\tilde{\mu}_0^{\text{PKG}*} = 0$, and that the quadruple $(\beta^{G*}, W^{G*}, \tilde{\mu}^{\text{PKG}*}, \tilde{\mu}^{\text{TIC}*})$ satisfies all the conditions in (PKG), (TIC), (FOC), (CE), (CSC_1), and (CSC_2), which completes the proof. □

Proof of Proposition 3 . It is immediate from Lemmata 6 and 7 that the incentive scheme (β^{G*}, W^{G*}) solves Subprogram (16) if and only if there exists a pair of multiplier processes $(\mu^{\text{PKG}^*}, \mu^{\text{TIC}^*}) = \{\mu_t^{\text{PKG}^*}, \mu_t^{\text{TIC}^*}\}_{t\leq T}$ *satisfying the following conditions at all* $t \leq T$,

$$\begin{aligned}
&\lambda\left(V'(W_t^{G*} + \beta_t^{G*}) - \mu_t^{\text{PKG}^*}\right) + \mu_t^{\text{TIC}^*} = 0. \\
&-\frac{\partial \mu_t^{\text{PKG}^*}}{\partial t} = \lambda\left(V'(W_t^{G*} + \beta_t^{G*}) - \mu_t^{\text{PKG}^*}\right),\ \mu_0^{\text{PKG}^*} = 0. \\
&\mu_t^{\text{TIC}^*} \geq 0, \quad \mu_t^{\text{TIC}^*}\left(\beta_t^{G*} - \frac{c}{\Delta_\lambda \pi_t^P}\right) = 0.
\end{aligned} \tag{39}$$

Assume that there exists a pair of multiplier processes $(\mu^{\text{PKG}^*}, \mu^{\text{TIC}^*})$ satisfying the three conditions given by (39). Since the initial value problem described by the co-state equation in the second line of (39) is a linear differential equation with an initial condition, its solution $\mu_t^{\text{PKG}*}$ must be uniquely given by the following expression:

$$\mu_t^{\text{PKG}*} := -\int_0^t \exp(-\lambda(\tau-t))\lambda V'(W_\tau^{G*} + \beta_\tau^{G*})d\tau.$$

for all $t \leq T$. Plugging this expression into the Kuhn-Tucker first-order condition (FOC) shows that

$$\mu_t^{\text{TIC}*} = -\lambda V'(W_t^{G*} + \beta_t^{G*}) - \int_0^t \exp(-\lambda(\tau-t))\lambda^2 V'(W_\tau^{G*} + \beta_\tau^{G*})d\tau,$$

which must be non-negative by the complementary slackness condition (CSC_2).

Conversely, assume that $\mu_t^{\text{TIC}^*}$ as defined as above is weakly positive for all $t \leq T$. It is easy to check that the multiplier pair process $\{(\mu_t^{\text{PKG}*}, \mu_t^{\text{TIC}*})\}_{t\leq T}$ as defined above satisfies the three conditions given by (39), which completes the proof. □

Proof of Lemma 8. Taking partial differentiation of $\underline{W}_t^S$ with respect to time yields:

$$\frac{\partial \underline{W}_t^S}{\partial t} \underbrace{=}_{(13)} \frac{\partial \underline{W}_t^G}{\partial t} + \frac{\partial \underline{\beta}_t^G}{\partial t} = \underbrace{r\underline{W}_t^G - \lambda \underline{\beta}_t^G + c}_{=\frac{\partial \underline{W}_t^G}{\partial t} \text{ by (PKG)}} + \underbrace{\Big(-\lambda \pi_t^P (1-\pi_t^P)\Big)}_{=\frac{\partial \pi_t^P}{\partial t} \text{ by (8)}} \times \underbrace{\left(-\frac{c}{\Delta_\lambda (\pi_t^P)^2}\right)}_{=\frac{\partial \underline{\beta}_t^G}{\partial \pi_t^P}} \tag{40}$$

$$= r\underline{W}_t^G - \left(\frac{\lambda - \Delta_\lambda}{\Delta_\lambda}\right) c,$$

where the last equality follows from direct computations. Also, we have:

$$\begin{aligned} \underline{W}_t^G &:= \int_t^T \exp(-r(\tau - t))\Big(\lambda\left(\frac{c}{\Delta_\lambda \pi_t^P}\right) - c\Big) d\tau, \\ &= c\left(\frac{\lambda}{\Delta_\lambda}\right)\left(\frac{1-\pi_t^P}{\pi_t^P}\right)\left(\frac{1-\exp(-(r-\lambda)(T-t))}{r-\lambda}\right) + c\left(\frac{\lambda - \Delta_\lambda}{\Delta_\lambda}\right)\left(\frac{1-\exp(-r(T-t))}{r}\right), \end{aligned} \tag{41}$$

where the equality in the second line also follows from direct computations. After plugging in the explicit expression for $\underline{W}_t^G$ given in (41) into the last line of (40) and rearranging its terms as necessary, we can see that $\frac{\partial \underline{W}_t^S}{\partial t}$ is positive if and only if:

$$\left(\frac{\exp((r-\lambda)(T-t)) - 1}{r-\lambda}\right) - \left(\frac{\exp(-\lambda T)}{r}\right)\left(\frac{\pi_0^P}{1-\pi_0^P}\right)\left(\frac{\lambda - \Delta_\lambda}{\lambda}\right) > 0 \tag{42}$$

Since the left hand side is strictly decreasing in t over the entire real line, the derivative $\frac{\partial \underline{W}_t^S}{\partial t}$ switches its sign at most once. In addition, as $t \to T$, the first term $\left(\frac{\exp((r-\lambda)(T-t))-1}{r-\lambda}\right)$ approaches zero, so the left side of the expression above becomes strictly negative. This implies that for any t sufficiently close to T, the derivative $\frac{\partial \underline{W}_t^S}{\partial t}$ is strictly negative. This completes the first part of the statement.

Moreover, for a given $\epsilon > 0$, whenever Δ_λ is sufficiently close to λ, the expression in (42) is strictly positive, which is equivalent to the fact that the derivative $\frac{\partial \underline{W}_t^S}{\partial t}$ must be strictly positive at $t = T - \epsilon$. Also, since the expression in (42) strictly decreases in t, the derivative $\frac{\partial \underline{W}_t^S}{\partial t}$ must be strictly positive at any $t \in [0, T-\epsilon]$, which completes the proof. □

Proof of Lemma 10. Fix any incentive scheme associated for a given termination time $T > 0$. Let us show that $\mu_t^{\text{PKG}*} \geq 0$ for all $t \leq T$. Since $\mu_0^{\text{PKG}*} = 0$ by the co-state equation (CE), it remains to show that $\mu_t^{\text{PKG}*}$ weakly increases over time. By Proposition 3, $\mu_t^{\text{PKG}*}$ must be differentiable in time t. Also, the Kuhn-Tucker optimality conditions in Proposition 3 yields

$$\frac{\partial \mu_t^{\mathrm{PKG}*}}{\partial t} \underbrace{=}_{\text{(CE)}} \lambda \left(\mu_t^{\mathrm{PKG}*} - V'(W_t^{G*} + \beta_t^{G*}) \right) \underbrace{=}_{\text{(FOC)}} \mu_t^{\mathrm{TIC}*} \underbrace{\geq}_{(\mathrm{CSC}_2)} 0$$

at any $t \leq T$. Therefore, $\mu_t^{\mathrm{PKG}*}$ weakly increases over time.

By the Kuhn-Tucker first-order condition, we have:

$$\lambda \left(\mu_t^{\mathrm{PKG}*} - V'(W_t^{G*} + \beta_t^{G*}) \right) \underbrace{=}_{\text{(FOC)}} \mu_t^{\mathrm{TIC}*}$$

By hypothesis, $\mu_t^{\mathrm{TIC}*} = 0$ whenever the temporary incentive compatibility constraint is slack at time t. Thus, for any such t, we must have $V'(W_t^{G*}) = \mu_t^{\mathrm{PKG}*}$, which was already shown to be non-negative. This completes the proof. □

Proof of Lemma 11. Fix any optimal contract and the incentive scheme associated with it. By the Kuhn-Tucker first-order condition, we have:

$$\lambda \left(\mu_t^{\mathrm{PKG}*} - V'(W_t^{G*} + \beta_t^{G*}) \right) \underbrace{=}_{\text{(FOC)}} \mu_t^{\mathrm{TIC}*}$$

By hypothesis, $\mu_t^{\mathrm{TIC}*} = 0$ whenever the temporary incentive compatibility constraint is slack at time t. Thus, for any such t, we must have $V'(W_t^{G*}) = \mu_t^{\mathrm{PKG}*}$, which was already shown to be non-negative. This completes the proof. □

Proof of Proposition 4. Let us outline the proof of Proposition 4. In **Steps 1 and 2**, we construct a family of incentive schemes indexed by the agent's exploitation utility $w_{slack} \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$. For each incentive scheme, there exists some time $\tilde{t}$ after which the agent's exploitation utility stays constant at $w_{slack} \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$. Moreover, the temporary incentive compatibility constraint must be slack whenever $t \in (\tilde{t}, T]$, and binds at $t \in (\tilde{t} - \epsilon, \tilde{t})$ for some $\epsilon > 0$. In particular, for each $w_{slack} \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$, we establish the existence of a unique $\tilde{t}$ satisfying the requirements above. In **Step 2,** we describe the continuation plan in the exploitation phase associated with each incentive scheme, and indeed shows that the temporary incentive compatibility is satisfied at each $t \leq T$. Next, in **Step 3,** we show that under this family of incentive schemes, the agent's exploitation utility must continue to weakly decrease over time once it started decreasing. In **Step 4,** we rely on the property derived in **Step 3** to select a particular incentive scheme as a candidate solution. In **Step 5**, we establish the optimality of our candidate solution based on the sufficient and necessary condition described in Proposition 3.

Step 1) For each $w_{slack} \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$, there exists a unique time $\bar{t}(w_{slack}) > 0$ such that:
a) there exists $\epsilon > 0$ such that $w_{slack} - \left(\frac{1-\exp(-(\lambda+r)(T-t))}{r+\lambda}\right)(\lambda w_{slack} - c) < \frac{c}{\Delta_\lambda \pi_t^P}$ for infinitely many $t \in [\bar{t}(w_{slack}) - \epsilon, \bar{t}(w_{slack})]$.
b) $w_{slack} - \left(\frac{1-\exp(-(\lambda+r)(T-t))}{r+\lambda}\right)(\lambda w_{slack} - c) \geq \frac{c}{\Delta_\lambda \pi_t^P}$ for all $t \in [\bar{t}(w_{slack}), T]$.
Furthermore, the mapping $\bar{t} : [\frac{c}{\Delta_\lambda \pi_T^P}, W^*] \to [0, T]$ that sends w_{slack} to the unique time satisfying the conditions above is differentiable and decreasing on $[\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$ with $\bar{t}(\frac{c}{\Delta_\lambda \pi_T^P}) = 0$.

Define an auxiliary function:

$$h(t, w_{slack}) := w_{slack} - \left(\frac{1-\exp(-(\lambda+r)(T-t))}{r+\lambda}\right)(\lambda w_{slack} - c) - \frac{c}{\Delta_\lambda \pi_t^P}.$$

Let us first establish uniqueness. Suppose, to the contrary, that there exists a pair (t_1, t_2) with $t_1 < t_2$ satisfying the properties in **Step 1.** Then, there exists a small $\epsilon > 0$ such that $h(t_2 - \epsilon, w_{slack}) < 0$, which contradicts the fact that $h(t_2 - \epsilon, w_{slack}) \geq 0$ by the first property in **Step 1** for time t_1.

We proceed by establishing the existence of $\bar{t}(w_{slack})$. By construction, $h(T, w_{slack}) = w_{slack} - \frac{c}{\Delta_\lambda \pi_T^P} \geq 0$. Also, since the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ is also not constrained optimal, there must exist a time t at which the inequality in the next line holds true, so that the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ is not incentive compatible:

$$\begin{aligned} 0 >& W^* - \left(\frac{1-\exp(-(\lambda+r)(T-t))}{r+\lambda}\right)(\lambda W^* - c) - \frac{c}{\Delta_\lambda \pi_t^P}. \\ \geq& w_{slack} - \left(\frac{1-\exp(-(\lambda+r)(T-t))}{r+\lambda}\right)(\lambda w_{slack} - c) - \frac{c}{\Delta_\lambda \pi_t^P} = h(t, w_{slack}) \end{aligned} \tag{43}$$

where the inequality in the second line follows from the fact that $w_{slack} \leq W^*$ and $\left(\frac{\lambda(1-\exp(-(\lambda+r)(T-t)))}{r+\lambda}\right) < 1$. Hence, for a fixed $w_{slack} \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$, the set $\{t : h(t, w_{slack}) < 0\}$ is non-empty with an upper bound T. It is straightforward to check that the supremum of this set satisfies both properties (a) and (b) required in **Step 1.**

Let $\bar{t}(\cdot)$ denote the function that sends $w_{slack} \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$ to the supremum of the set $\{t : h(t, w_{slack}) < 0\}$. We wish to prove that the function $\bar{t}(\cdot)$ is differentiable. Since the function $h(t, w_{slack})$ is continuous in both arguments, assuming $h(\bar{t}(w_{slack}), w_{slack}) \neq 0$ leads to an immediate contradiction to the fact that $\bar{t}(w_{slack}) = \sup\{t : h(t, w_{slack}) < 0\}$, so we must have $h(\bar{t}(w_{slack}), w_{slack}) = 0$. Furthermore, suppose, to the contrary, that $\frac{\partial h(\bar{t}(w_{slack}), w_{slack})}{\partial t} =$

0. Direct computations show that

$$\frac{\partial^2 h(\bar{t}(w_{slack}), w_{slack})}{\partial t^2} = \frac{\partial}{\partial t}\left(\frac{\partial h(\bar{t}(w_{slack}), w_{slack})}{\partial t}\right)$$

$$= \frac{\partial}{\partial t}\left(\exp(-(\lambda+r)(T-t))(\lambda w_{slack} - c) - \underbrace{\left(\frac{1-\pi_t^P}{\pi_t^P}\right)}_{=\exp(\lambda t)\frac{1-\pi_0}{\pi_0}} \frac{\lambda c}{\Delta_\lambda}\right)\Big|_{t=\bar{t}(w_{slack})}$$

$$= \lambda \underbrace{\left(\frac{\partial h(\bar{t}(w_{slack}), w_{slack})}{\partial t}\right)}_{=0 \text{ by hypothesis}} + r\exp(-(\lambda+r)(T-t)) \underbrace{(\lambda w_{slack} - c)}_{>0 \because w_{slack} \geq \frac{c}{\Delta_\lambda \pi_T^P}} > 0$$

Therefore, since $\frac{\partial h(\bar{t}(w_{slack}), w_{slack})}{\partial t} = 0$ and $\frac{\partial^2 h(\bar{t}(w_{slack}), w_{slack})}{\partial t^2} > 0$, $\bar{t}(w_{slack})$ must be a local minimizer of $h(\cdot, w_{slack})$ for a fixed w_{slack}. However, since $\bar{t}(w_{slack})$ is the supremum of the set $\{t : h(t, w_{slack}) < 0\}$, for every $\epsilon > 0$, there exists $\tilde{t} \in [\bar{t} - \epsilon, \bar{t}]$ such that $h(\tilde{t}, w_{slack}) < 0 = h(\bar{t}(w_{slack}), w_{slack})$, which contradicts the fact that $\bar{t}(w_{slack})$ must be a local minimizer of $h(\cdot, w_{slack})$ for a fixed w_{slack}. Therefore, we must have $\frac{\partial h(\bar{t}(w_{slack}), w_{slack})}{\partial t} \neq 0$ for any $w_{slack} \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$. Thus, we may apply the implicit function theorem to prove that $\bar{t}(\cdot)$ is a differentiable function in its domain $[\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$. Furthermore, direct computations yield $h(T, \frac{c}{\Delta_\lambda \pi_T^P}) = 0$ and $\frac{\partial h(T, \frac{c}{\Delta_\lambda \pi_T^P})}{\partial t} = \left(\frac{\lambda - \Delta_\lambda}{\Delta_\lambda}\right) c > 0$, which implies that $\bar{t}(\frac{c}{\Delta_\lambda \pi_T^P}) = T$.

Let us show that the mapping $\bar{t}(\cdot)$ is a strictly decreasing function. Fix any w_1, w_2 with $\frac{c}{\Delta_\lambda \pi_T^P} \leq w_1 < w_2 \leq W^*$. Then, by the continuity of $h(\cdot,\cdot)$ in both arguments, $h(\bar{t}(w_2), w_2) = 0$. Moreover, since the function $h(t, w)$ strictly increases in w, we must have $h(\bar{t}(w_2), w_1) < 0$. Hence, for any $\tilde{t}$ in a sufficiently small neighborhood of $\bar{t}(w_2)$, $h(\tilde{t}, w_1) < 0$ by the continuity of $h(\cdot,\cdot)$ in both arguments. Then, by the construction of $\bar{t}(\cdot)$, we must have $\bar{t}(w_1) > \bar{t}(w_2)$.. This completes the proof of **Step 1.**

Step 2) For each $w_{slack} \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$, we construct a candidate for the optimal continuation plan $\{W_t^S(w_{slack})\}_{t \leq T}$ in the exploitation phase as follows. Then, for each $t \in [0, T]$, define:
(a) For any $t \in [\bar{t}(w_{slack}), T]$, set $\tilde{W}_t^S(w_{slack}) = w_{slack}$.
(b) For any $t < \bar{t}(w_{slack})$, construct $\tilde{W}_t^S(w_{slack})$ so that the temporary incentive compatibility constraint binds at t :

$$\tilde{W}_t^S(w_{slack}) = \frac{c}{\Delta_\lambda \pi_t^P} + \underbrace{\int_t^{\bar{t}(w_{slack})} \exp(-r(\tau - t))(\lambda \left(\frac{c}{\Delta_\lambda \pi_\tau^P}\right) - c)d\tau + \exp(-r(\bar{t}(w_{slack}) - t))\tilde{W}_{\bar{t}}^G}_{=\tilde{W}_t^G},$$

where $\tilde{W}_{\bar{t}(w_{slack})}^G = \int_{\bar{t}(w_{slack})}^T \exp(-(r+\lambda)(\tau - \bar{t}(w_{slack})))(\lambda w_{slack} - c)d\tau$ is the agent's time-$\bar{t}(w_{slack})$ continuation value under the probability measure $\mathbb{P}_a^G$. By construction, the ex-

ploitation utility process $\{\tilde{W}_t^S(w_{slack})\}_{t\leq T}$ satisfies the temporary incentive compatibility constraint at any $t \leq T$.
(c) If $\tilde{W}_0^S < W^*$, then define:

$$W_t^S(w_{slack}) := \begin{cases} W^* \text{ if } t < \underline{t}(w_{slack}), \\ \tilde{W}_t^S \text{ if } t \geq \underline{t}(w_{slack}), \end{cases} \tag{44}$$

where $\underline{t}(w_{slack})$ denotes the first time at which the exploitation utility $\tilde{W}_t^S \geq W^*$ under the original incentive scheme $\mathcal{I}_T^{\text{Binding}}$ (with $\underline{t}(w_{slack}) := 0$ if $\tilde{W}_0^S \geq W^*$). An argument similar to the proof of Lemma 5 shows that the exploitation utility process $\{W_t^S(w_{slack})\}_{t\leq T}$ satisfies the temporary incentive compatibility constraint at any $t \leq T$.

Step 3) We claim that for each $w_{slack} \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$, if $\frac{\partial W_t^S(w_{slack})}{\partial t}\Big|_{t=\hat{t}} < 0$ for some $\hat{t} \in [\underline{t}(w_{slack}), \bar{t}(w_{slack})]$, then $\frac{\partial W_t^S(w_{slack})}{\partial t} < 0$ for all $t \in [\hat{t}, \bar{t}(w_{slack})]$.

Suppose, to the contrary, that there exists $w_{slack} \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$ such that

$$\frac{\partial W_t^S(w_{slack})}{\partial t}\Big|_{t=t_1} < 0 \leq \frac{\partial W_t^S(w_{slack})}{\partial t}\Big|_{t=t_2}$$

for some pair (t_1, t_2) with $t_1 < t_2$. It entails no loss of generality to assume that t_2 is the first time at which $\frac{\partial W_t^S(w_{slack})}{\partial t}$ becomes weakly positive after time t_1. Therefore, $\frac{\partial W_t^S(w_{slack})}{\partial t} < 0$ for all $t \in (t_1, t_2)$. Furthermore, for all $t \in [\underline{t}(w_{slack}), \bar{t}(w_{slack})]$, define $W_t^G(w_{slack}) := W_t^S(w_{slack}) - \frac{c}{\Delta_\lambda \pi_t^P}$ to be the agent's equilibrium continuation value after failure up to time t, evaluated from the perspective of somebody who knows the project is good. Since $\frac{c}{\Delta_\lambda \pi_t^P}$ strictly increases in t on $[\underline{t}(w_{slack}), \bar{t}(w_{slack})]$, the fact that the exploitation utility $W_t^S(w_{slack})$ is strictly decreasing on $[t_1, t_2]$ implies that $W_t^G(w_{slack})$ must be strictly decreasing on $[t_1, t_2]$.

The same set of computations as in (40) yield:

$$\frac{\partial W_t^S(w_{slack})}{\partial t} = rW_t^G(w_{slack}) - \left(\frac{\lambda - \Delta_\lambda}{\Delta_\lambda}\right)c,$$

for all $t \in [\underline{t}(w_{slack}), \bar{t}(w_{slack})]$. Since $\frac{\partial W_t^S(w_{slack})}{\partial t}\Big|_{t=t_1} < 0 \leq \frac{\partial W_t^S(w_{slack})}{\partial t}\Big|_{t=t_2}$ by hypothesis, we must have $W_{t_1}^G(w_{slack}) < \left(\frac{\lambda-\Delta_\lambda}{\Delta_\lambda}\right)\frac{c}{r} \leq W_{t_2}^G(w_{slack})$ by the computation above. However, this contradicts the fact that W_t^G strictly decreases on $[t_1, t_2]$, which completes the proof of **Step 3.**

Step 4) For each $w_{slack} \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$, we also define the candidate multiplier pair associated with the solution as follows:

$$\mu_t^{PKG}(w_{slack}) := \begin{cases} \exp(-\lambda(\bar{t}(w_{slack}) - t))V'(w_{slack}) + \int_t^{\bar{t}(w_{slack})} \exp(-\lambda(\tau - t))\lambda V'(W_\tau^S(w_{slack}))d\tau \\ \qquad\qquad\qquad\qquad\qquad \text{if } t \leq \bar{t}(w_{slack}) \\ V'(w_{slack}) \qquad\qquad\qquad\quad \text{if } t > \bar{t}(w_{slack}) \end{cases}$$
$$\mu_t^{TIC}(w_{slack}) := -\lambda V'(W_t^S(w_{slack})) + \lambda\mu_t^{PKG}(w_{slack}) \qquad \text{if } t \leq T. \tag{45}$$

Let us show the existence of a $w_{slack}^* \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$ such that $\mu_0^{PKG}(w_{slack}^*) = 0$. In particular, we first prove that $\mu_0^{PKG}(W^*) < 0$ and $\mu_0^{PKG}(\frac{c}{\Delta\pi_T^P}) > 0$ and apply the intermediate value theorem to establish the existence of our desired w_{slack}^*. Then, we select $\{W_t^S(w_{slack}^*)\}_{t\leq T}$ as our candidate for the constrained optimal continuation plan in the exploitation phase.

Let us show that the continuation plan $W_t^S(W^*) \geq W^*$ in the exploitation phase at all $t \leq T$. Suppose, to the contrary that $W_{\hat{t}}^S(W^*) < W^*$ for some $\hat{t} \leq T$. Since $W_t^S(W^*) = W^*$ for all $t \geq \bar{t}(W^*)$ by definition, $\hat{t} \leq \bar{t}(W^*)$. Since $W_{\bar{t}(W^*)}^S(W^*) = W^*$ and $W_{\hat{t}}^S(W^*) < W^*$, the mean-value theorem implies that there exists $\tilde{t}$ such that $\frac{\partial W_t^S(W^*)}{\partial t}\Big|_{t=\tilde{t}} > 0$ for some $\tilde{t} \in [\hat{t}, \bar{t}(W^*)]$, which in turn implies that $\frac{\partial W_t^S(W^*)}{\partial t} > 0$ for all $t \in [\underline{t}(W^*), \bar{t}(W^*)]$ by **Step 3.** However, this is absurd. If $\underline{t}(W^*) = 0$, then the strict monotonicity of $W_t^S(W^*)$ on $[\underline{t}(W^*), \bar{t}(W^*)]$ implies that $W_t^S(W^*) < W^*$ for all $[\underline{t}(W^*), \bar{t}(W^*)]$, which contradicts the definition of $\underline{t}(W^*)$. Alternatively, if $\underline{t}(W^*) > 0$, $W_{\underline{t}(W^*)}^S(W^*) = W_{\bar{t}(W^*)}^S(W^*) = W^*$. However, this is impossible because $W_t^S(W^*)$ strictly increases in t on $[\underline{t}(W^*), \bar{t}(W^*)]$. Hence, $W_t^S(W^*) \geq W^*$ at all $t \leq T$.

Furthermore, $W_t^S(W^*) > W^*$ for a positive duration of time by the hypothesis that the first candidate $\mathcal{I}_T^{\text{Type 1}}$ for the constrained optimal incentive scheme is not implementable. Therefore, $V'(W_t^S(W^*))$ is weakly negative at all $t \leq T$ and strictly negative for a positive duration of time, which implies that

$$\mu_0^{PKG}(W^*) = \exp(-\lambda(\bar{t}(w_{slack}) - t))V'(W^*) + \int_t^{\bar{t}(w_{slack})} \exp(-\lambda(\tau - t))\lambda V'(W_\tau^S(W^*))d\tau < 0.$$

Let us show that $\mu_0^{PKG}(\frac{c}{\Delta_\lambda \pi_T^P}) > 0$. Observe that $\{W_t^S(\frac{c}{\Delta_\lambda \pi_T^P})\}_{t\leq T}$ is the exploitation utility process induced by the second candidate $\mathcal{I}_T^{\text{Type 2}}$ for the constrained optimal incentive scheme. Define:

$$M_t^{TIC} := -\lambda V'(W_t^S(\frac{c}{\Delta_\lambda \pi_T^P})) - \int_0^t \exp(-\lambda(\tau - t))\lambda^2 V'(W_\tau^S(\frac{c}{\Delta_\lambda \pi_T^P}))d\tau,$$

which is the "hypothetical multiplier process" induced by the second candidate $\mathcal{I}_T^{\text{Type 2}}$ for the constrained optimal incentive scheme. Since $\mathcal{I}_T^{\text{Type 2}}$ is assumed to be suboptimal, Proposition 3 implies that there must exist some $t_1 \in [0, T]$ such that $M_{t_1}^{TIC} < 0$.

As an intermediate step, we provide a "phase-diagrammatic" argument to show that $M_t^{TIC} < 0$ for all $t \in (t_1, T]$. Suppose, to the contrary that there exists $t_2 \in (t_1, T]$ such that $M_{t_2}^{TIC} \geq 0$. We may assume without loss of generality that t_2 is the first such time after t_1. Partially differentiating M_t with respect to time yields:

$$\frac{\partial M_t^{TIC}}{\partial t} := -\lambda V''(W_t^S(\frac{c}{\Delta_\lambda \pi_T^P}))\frac{\partial W_t^S(\frac{c}{\Delta_\lambda \pi_T^P})}{\partial t} + \lambda M_t^{TIC}.$$

We wish to show that $\frac{\partial M_t^{TIC}}{\partial t} < 0$ for all $t \in (t_1, t_2)$ to derive a contradiction. Observe that $V''(\cdot) \leq 0$ by concavity. Furthermore, let us show that $\frac{\partial W_t^S(\frac{c}{\Delta_\lambda \pi_T^P})}{\partial t} \leq 0$ for all $t \in (t_1, t_2)$. If $\frac{\partial W_t^S(\frac{c}{\Delta_\lambda \pi_T^P})}{\partial t} > 0$ were to hold true for some $t' \in (t_1, t_2)$, **Step 3** would imply $W_t^S(\frac{c}{\Delta_\lambda \pi_T^P})$ is weakly increasing in t over $[0, t']$, which in turn implies that $W_t^S \geq W^*$ for all $[0, t']$. Therefore, $V'(W_t^S) \leq 0$ for all $[0, t']$. Hence, the definition of $M_{t_1}^{TIC}$ implies that it must be weakly positive, which contradicts our hypothesis that it is strictly negative. Therefore, $\frac{\partial W_t^S(\frac{c}{\Delta_\lambda \pi_T^P})}{\partial t} \leq 0$ for all $t \in (t_1, t_2)$. Finally, since t_2 is the first time such that $M_{t_2}^{TIC} \geq 0$ after t_1, $M_{t_2}^{TIC} < 0$ for all $t \in (t_1, t_2)$. Hence, $\frac{\partial M_t^{TIC}}{\partial t} < 0$ for all $t \in (t_1, t_2)$.

However, since $M_{t_1}^{TIC}$ is strictly negative and $\frac{\partial M_t^{TIC}}{\partial t} < 0$ for all $t \in (t_1, t_2)$, $M_{t_2}^{TIC}$ is strictly negative by the mean-value theorem. However, this contradicts our hypothesis that $M_{t_2}^{TIC} \geq 0$, which shows that $M_t^{TIC} < 0$ for all $t \in (t_1, T]$. Hence, $M_T^{TIC} < 0$, which implies:

$$\mu_0^{PKG}(\frac{c}{\Delta_\lambda \pi_T^P}) = \exp(-\lambda T) \underbrace{\left(V'(\frac{c}{\Delta_\lambda \pi_T^P}) + \int_0^T \exp(-\lambda\tau)\lambda V'(W_\tau^S(\frac{c}{\Delta_\lambda \pi_T^P}))d\tau\right)}_{=-M_T^{TIC}} > 0. \quad (46)$$

Hence, $\mu_0^{PKG}(\frac{c}{\Delta_\lambda \pi_T^P}) > 0$, and $\mu_0^{PKG}(W^*) < 0$. By construction, $\mu_0^{PKG}(w_{slack})$ is continuous with respect to w_{slack}. Hence, the intermediate value theorem implies that there exists $w_{slack}^* \in (\frac{c}{\Delta_\lambda \pi_T^P}, W^*)$ such that $\mu_0^{PKG}(w_{slack}^*) = 0$.

Step 5) By Proposition 3, it suffices to show that $\mu_t^{TIC}(w_{slack}^*) \geq 0$ for all $t \in [0, T]$. Suppose, to the contrary, that there exists $t_1 \in [0, \bar{t}(w_{slack}^*)]$ such that $\mu_{t_1}^{TIC}(w_{slack}^*) < 0$. Under this hypothesis, almost exactly the same "phase-diagrammatic" arguments used in **Step 3** to show $M_t^{TIC} < 0$ for all $t \in (t_1, T]$ can be carried over to show that $\mu_{\bar{t}(w_{slack}^*)}^{TIC}(w_{slack}^*) < 0$. However, this contradicts the fact that $\mu_{\bar{t}(w_{slack}^*)}^{TIC}(w_{slack}^*) = 0$ by the definition in (45), which completes the proof. $\square$

Proof of Proposition 5. The proof of Proposition 5 heavily relies on the construction and methods given in the proof of Proposition 4, so it is useful to read the proof of Proposition 4 before going over the remaining part of the proof of Proposition 5. In **Step 1,** we show that when the principal's exploitation value function is V_2, neither the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ nor $\mathcal{I}_T^{\text{Type 2}}$ is a constrained optimal incentive scheme. This allows us to apply Proposition 4 to construct the unique constrained optimal incentive scheme under V_2. In **Step 2,** we show that the temporary incentive constraint is slack on $[t_i, T]$ under the unique constrained optimal incentive scheme under principal's exploitation value function V_i for each $i = 1, 2,$, and also that $t_2 < t_1$.

Step 1) We claim that when the principal's exploitation value function is V_2, neither the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ nor $\mathcal{I}_T^{\text{Type 2}}$ is a constrained optimal incentive scheme. Suppose, to the contrary, the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ is constrained optimal when the principal's exploitation value function is V_2. Hence, the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ must satisfy the incentive compatibility constraint. However, this implies that the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ is constrained optimal even when the principal's exploitation value function is V_1. If this is true, then the constrained optimal incentive scheme under V_1 cannot be optimal because $\mathcal{I}_T^{\text{Type 1}}$ does strictly better, which contradicts our hypothesis.

Suppose, to the contrary, the incentive scheme $\mathcal{I}_T^{\text{Type 2}}$ is constrained optimal when the principal's exploitation value function is V_2. For each $t \leq T$, let $W_t^{S,Type2}$ denote the agent's time-t exploitation utility process associated with the incentive scheme $\mathcal{I}_T^{\text{Type 2}}$. Since $V_1(\cdot)$ is strictly concave, Proposition 3 implies that the incentive scheme described in the statement of Proposition 5 is the unique solution to the implementation Subprogram. The uniqueness of the solution implies that the incentive scheme $\mathcal{I}_T^{\text{Type 2}}$ cannot be constrained optimal when principal's exploitation value function is V_1, so there must exist $\tilde{t} < 0$ such that the expression in the first line must be strictly negative:

$$
\begin{aligned}
& -\lambda V_1'(W_{\tilde{t}}^{S,Type2}) - \int_0^{\tilde{t}} \exp(-\lambda(\tau - \tilde{t}))\lambda^2 V_1'(W_\tau^{S,Type2}) d\tau \\
\geq & -\lambda V_2'(W_{\tilde{t}}^{S,Type2}) - \int_0^{\tilde{t}} \exp(-\lambda(\tau - \tilde{t}))\lambda^2 V_2'(W_\tau^{S,Type2}) d\tau
\end{aligned}
$$

where the inequality in the second line follows from the initial hypothesis that $V_1'(\cdot)$ is weakly greater than $V_2'(\cdot)$ at all points. Therefore, since the first line is strictly negative, the inequalities above imply that the last line must be strictly negative. However, since the incentive scheme $\mathcal{I}_T^{\text{Type 2}}$ is assumed to be constrained optimal, Proposition 5 implies that the last line must be weakly positive. This contradiction proves **Step 1.**

Step 2) Since neither the incentive scheme $\mathcal{I}_T^{\text{Type 1}}$ nor $\mathcal{I}_T^{\text{Type 2}}$ is a constrained optimal incentive scheme under a strictly concave exploitation value function V_2, there exists a unique constrained optimal incentive scheme as described in Proposition 4 under V_2. Moreover, the proof of Proposition 4 shows that both the constrained optimal incentive scheme under V_1 and the one under V_2 must belong to the family of incentive schemes as described in **Step 2)** of the proof of Proposition 4. In particular, there exists $w_1^* \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$ such that the incentive scheme $\{W_t^S(w_1^*)\}_{t\le T}$ is constrained optimal under the exploitation value function V_1. This implies that when the principal's exploitation value function is V_1, the time-0 multiplier associated with the promise-keeping constraint under the incentive scheme $\{W_t^S(w_1^*)\}_{t\le T}$ is zero, which can be formally expressed as follows:

$$\begin{aligned}0 =& \exp(-\lambda\bar{t}(w_1^*))V_1'(w_1^*) + \int_0^{\bar{t}(w_1^*)} \exp(-\lambda\tau)\lambda V_1'(W_\tau^S(w_1^*))d\tau \\ <& \exp(-\lambda\bar{t}(w_1^*))V_2'(w_1^*) + \int_0^{\bar{t}(w_1^*)} \exp(-\lambda\tau)\lambda V_2'(W_\tau^S(w_1^*))d\tau,\end{aligned}$$

where the strict inequality follows from the fact that 1) $w_1^* \in [\frac{c}{\Delta_\lambda \pi_T^P}, W^*]$ and 2) $V_2'(w)$ is weakly greater than $V_1'(w)$ for all $w \ge 0$, and strictly greater whenever $w \in [0, W^*]$. This implies that when the principal's exploitation value function is V_2, the time-0 multiplier associated with the promise-keeping constraint under the incentive scheme $\{W_t^S(w_1^*)\}_{t\le T}$ is strictly higher than zero. Moreover, by arguing analogously to **Step 4** of Proposition 4, we can establish the following inequality:

$$\exp(-\lambda\bar{t}(w_1^*))V_1'(W^*) + \int_0^{\bar{t}(W^*)} \exp(-\lambda\tau)\lambda V_1'(W_\tau^S(W^*))d\tau < 0$$

Therefore, by the intermediate value theorem, there must exist $w_2^* \in (w_1^*, W^*)$ such that $\exp(-\lambda\bar{t}(w_2^*))V_2'(w_2^*) + \int_0^{\bar{t}(w_2^*)} \exp(-\lambda\tau)\lambda V_2'(W_\tau^S(w_2^*))d\tau = 0$. Arguing as in **Step 5** of Proposition 4 shows that the incentive scheme $\{W_t^S(w_2^*)\}_{t\le T}$ is uniquely constrained optimal.

Moreover, by the definition given in **Step 2** of the proof of Proposition 4, for each $i = 1, 2$, the temporary incentive constraint is slack on the time interval $[\bar{t}(w_i^*), T]$ under the incentive scheme $\{W_t^S(w_i^*)\}_{t\le T}$. Hence, for each $i = 1, 2$, $t_i = \bar{t}(w_i^*)$ in the statement of Proposition 5. Since $\bar{t}$ is a strictly decreasing function by **Step 1** of Proposition 4 and $w_1^* < w_2^*$, we must have $t_2 < t_1$, which completes the proof.

□

Proof of Proposition 6. First, let us assume that the principal's exploitation value function is $V_D(W_t^S) = y - W_t^S$ for some $y > 0$. Since the value function is downward sloping in W_t^S, the temporary incentive compatibility constraint always binds throughout the exploration phase in the optimal contract. Therefore, the sensitivity of the agent's exploitation utility to success must be equated to the minimal possible level: $\beta_{T*}^{G*} = \frac{c}{\lambda \pi_{T^*}^P}$. Furthermore, since $V_D'(W_t^S) = -1$ for all $W_t^S \geq 0$, we can use the co-state equation (CE) to obtain:

$$\mu_T^{\text{PKG}*} = \exp(\lambda T) - 1 \tag{47}$$

for any given $T > 0$. Therefore, the first-order condition in (21) can be written as follows:

$$\lambda \pi_{T_D^*}^P y - c - i = c(\exp(\lambda T_D^*) - 1)(1 - \pi_{T_D^*}^P). \tag{48}$$

when T_D^* is the optimal termination time under the downward sloping value function $V_D(W_t^S)$. Since the shadow cost from delivering an additional utility on the right hand side of the last equation is strictly positive, the principal prematurely terminates the relationship when her belief $\pi_{T_D^*}^P$ is strictly above the first best threshold level $\pi_{T^*_{\text{First Best}}}^P = \frac{c+i}{\lambda y}$.

Fix any $\underline{\pi} \in (\pi_{T^*_{\text{First Best}}}^P, \pi_{T_D^*}^P)$ and let $\underline{y} := \frac{c+i}{\lambda \underline{\pi}}$. Observe that

$$y = \frac{c+i}{\lambda \pi_{T^*_{\text{First Best}}}^P} < \frac{c+i}{\lambda \underline{\pi}} = \underline{y},$$

where the strictly inequality follows from $\pi > \pi_{T^*_{\text{First Best}}}^P$. Hence, $y > \underline{y}$. Define an inverted value function $V_I(\cdot)$ as follows:

$$V_I(W_t^S) := \begin{cases} -a_1(W_t^S - W^*)^2 + \underline{y} - W^* & \text{if } W_t^S \leq W^* \\ -a_2(W_t^S - W^*)^2 + \underline{y} - W^* & \text{if } W_t^S \in (W^*, W_1) \\ \frac{1}{2}(\underline{y} + y) - W_t^S & \text{if } W_t^S \geq W_1, \end{cases}$$

where the parameters W^*, W_1, a_1, and a_2 are respectively given by:

$$W^* = \frac{c}{\lambda \underline{\pi}}, \; W_1 = W^* + \frac{y - \underline{y}}{3}, \; a_1 = \frac{\underline{y} - W^*}{(W^*)^2}, \; a_2 = \frac{3}{2(y - \underline{y})}.$$

It is easy to verify that the function V_I is concave and differentiable everywhere, has the unique maximizer $W^* = \frac{c}{\lambda \underline{\pi}}$, and is strictly less than the function $V_D(W_t^S) = y - W_t^S$ for every $W_t^S \geq 0$.

Let us show that under the value function V_I, the relationship proceeds as follows in the optimal contract. If the agent succeeds, the principal implements her optimal exploitation contract (i.e., delivers W^* to the agent upon success) immediately upon success. The relationship is terminated if the agent keeps failing until the principal's belief is equal to $\underline{\pi} < \pi^P_{T^*_D}$. This implies that the termination time under the value function V_I must be strictly longer than the termination time under the value function V_D.

Since the minimal sensitivity to induce the agent to exert effort at the termination time is equal to $\frac{c}{\lambda\underline{\pi}} = W^*$, the exploitation utility induced by the contract described above coincides with the one given in (14). Therefore, the contract is incentive compatible by Lemma 5.

It remains to show that the contract does better than any incentive compatible contract. For any fixed termination time, an incentive compatible contract can do no better than a contract contracts that always delivers W^* to the agent upon success. Moreover, among all the contracts that always deliver W^* to the agent upon success, the optimal termination belief satisfies the following first condition with respect to termination time:

$$\lambda\underline{\pi}\underbrace{V(W^*)}_{=\frac{i}{\lambda\underline{\pi}}} = i.$$

Since the contract above satisfies the optimality condition above, it does strictly better than any incentive compatible contract, which completes the proof. □

Proof of Lemma 12. First, we show that it is optimal for the agent to exert effort under the contract described above. By computation, the agent's equilibrium continuation value after failing up to and including time t can be calculated as:

$$W^G_t := \int_0^\infty \exp(-(\lambda + r)t)(\lambda \max\{W^*, \frac{c}{\lambda}\} - c)dt = \frac{\max\{\lambda W^* - c, 0\}}{r + \lambda}.$$

If $W^* < \frac{c}{\lambda}$, the sensitivity $\beta^G_t := \frac{c}{\lambda} - W^G_t$ of the agent's time-t equilibrium continuation value to success is equal to $\frac{c}{\lambda}$, which is exactly the minimal level required to incentivize the agent. If $W^* > \frac{c}{\lambda}$, the sensitivity $\beta^G_t := W^* - W^G_t$ of the agent's time-t equilibrium continuation value to success is equal to $\frac{rW^*+c}{\lambda+r}$, which is greater than $\frac{c}{\lambda}$ by the hypothesis that $W^* > \frac{c}{\lambda}$. Therefore, the contract described above is incentive compatible.

We verify whether the contract described above is indeed optimal. First, let us assume that $W^* < \frac{c}{\lambda}$. By this hypothesis, the principal's value function V in the second phase must be strictly decreasing on the interval $[\frac{c}{\lambda}, \infty)$. Since the agent must be promised at least $\frac{c}{\lambda}$ upon success to exert effort, this implies that there cannot be any incentive scheme that improves upon the one that promises $\frac{c}{\lambda}$ upon success. Second, let us assume that $W^* > \frac{c}{\lambda}$. Since W^* is the unique maximizer of $V(\cdot)$, no incentive scheme can improve upon one that promises W^* upon success. This completes the proof. □