

Discussion Paper No. 1015

**DYNAMIC PERFORMANCE EVALUATION
WITH DEADLINES:
THE ROLE OF COMMITMENT**

Chia-Hui Chen
Junichiro Ishida

October 2017

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Dynamic Performance Evaluation with Deadlines: The Role of Commitment*

Chia-Hui Chen[†] and Junichiro Ishida[‡]

October 20, 2017

Abstract

We consider an environment in which a principal hires an agent and evaluates his productivity over time in an ongoing relationship. The problem is embedded in a continuous-time model with both hidden action and hidden information, where the principal must induce the agent to exert effort to facilitate her learning process. The value of committing to a deadline is examined in this environment, and factors which make the deadline more profitable are identified. Our framework generates a unique recursive equilibrium structure under no commitment which can be exploited to obtain a full characterization of equilibrium. The analysis allows us to evaluate the exact value of commitment for any given set of parameters and provides insight into when it is beneficial to commit to an evaluation deadline at the outset of a relationship.

JEL Classification Number: D82, D83

Keywords: dynamic agency, deadlines, experimentation, commitment, up-or-out contracts.

*The first author acknowledges financial support from JSPS KAKENHI Grant-in-Aid for Scientific Research (B) 16H03615 and Young Scientists (B) 17K13699. The second author acknowledges financial support from JSPS KAKENHI Grant-in-Aid for Scientific Research (S) 15H05728, (A) 20245031, (B) 16H03615 and (C) 24530196 as well as the program of the Joint Usage/Research Center for Behavioral Economics at ISER, Osaka University.

[†]Institute of Economic Reserach, Kyoto University. Email: chchen@kier.kyoto-u.ac.jp

[‡]Institute of Social and Economic Research, Osaka University. Email: jishida@iser.osaka-u.ac.jp

1 Introduction

Suppose that an employer needs to hire a worker to carry out a project over time. The project is ability-intensive in that the worker can successfully complete the project only if he is sufficiently productive. As is often the case, however, the worker's sheer productivity is not directly observable to the employer, who must instead make an inference from a sequence of observed outputs. Since success cannot be achieved overnight, the employer must exercise some patience even when things do not appear to be in good shape. Excessive tolerance for failure, however, diminishes the worker's incentive to take costly actions that are indispensable to be successful. A number of issues arise in this dynamic environment. How much time should the employer give the worker before she terminates the project? Should the employer commit to a deadline at the outset? If so, under what conditions?

In this paper, we attempt to address these issues by analyzing a situation where a principal hires an agent and evaluates his productivity over time in an ongoing relationship. The problem is embedded in a continuous-time framework with both hidden action and hidden information. At each instant, the agent privately chooses how much effort to supply. The outcome is either a success or a failure, depending on his effort choice as well as his productivity type. The game ends immediately when the agent achieves a success (or a "breakthrough"). The principal's task in this environment is to determine when to terminate the project, conditional on no success having occurred. Within this setup, we analyze two distinct cases to illustrate the role of commitment: in one case, the principal sets a deadline and commits to it at the outset, and the project is terminated automatically when the deadline is reached without attaining a success; in the other, the principal makes no such commitment, thereby retaining discretion to terminate the project at any instant, and simply terminates the project when the continuation payoff is not high enough to justify further experimentation. By directly comparing these two cases, we evaluate the extent to which the principal benefits from committing to an evaluation deadline in this dynamic environment.

The driving force of our analysis is a dynamic strategic interaction between the agent's effort choice and the principal's termination strategy. On one hand, the agent's effort choice depends clearly on how much time is left until the project is terminated: since the net value of achieving a success is low when the project is still far from termination, the agent tends to start off with low effort and gradually shift to higher effort as the expected termination date approaches. On the other hand, the principal's willingness to terminate the project depends also on the agent's effort choice: when the agent is less motivated and exerts low effort, less information is revealed about his type, which makes the principal more reluctant

to terminate the project. This strategic interaction can generate a vicious cycle where the principal's reluctance to terminate the project diminishes the agent's motivation, which in turn makes the principal even more reluctant. It is in general profitable to commit to a deadline when each player's incentive to "procrastinate" is sufficiently strong.

The main contribution of the paper is that we devise an analytical framework that is tractable enough to admit a complete characterization of pure-strategy equilibria, both with and without commitment, while capturing this dynamic interaction. Our framework thus allows us to evaluate the exact value of commitment to an evaluation deadline for any given set of parameters. To this end, our main analytical focus is on the no-commitment case, which is generally far more challenging than the commitment case when dynamic interactions are considered. In the no-commitment case, the principal's termination strategy must be sequentially rational along the way, and the belief off the equilibrium path plays a crucial role. Even in this case, we can show that our framework generates a unique recursive equilibrium structure which can be exploited to establish the existence and uniqueness of (pure-strategy) perfect Bayesian equilibrium. We then build on this result to derive a necessary and sufficient condition under which the principal can strictly benefit from committing to an evaluation deadline.

As can be expected, the lack of commitment to a deadline entirely alters the dynamic allocation of effort as well as the timing of project termination. When the principal sets a deadline at the outset, she does so by taking into account how it affects the agent's entire effort sequence. In particular, since extending a deadline in general relaxes the agent's incentive compatibility constraint and reduces his total effort supply, the deadline must be set at a point where the expected benefit (of achieving a success at the next instant) equals the cost of a decrease in total effort. The situation changes rather drastically when she makes no such commitment, because her termination decision would have no influence on the agent's past behavior. As such, she terminates the project whenever the continuation payoff is about to turn negative while taking the agent's effort strategy as given. This fact implies that the cost of extending a deadline is smaller under no commitment and tends to give her an excessive incentive to wait for a success.

Given this incentive structure, one may expect that the principal would always wait longer for a success in equilibrium under no commitment. As it turns out, though, this conjecture does not always hold true in the current setup. Among other things, an interesting, and somewhat counterintuitive, property of our model is that the average duration of the project may not be monotonic with respect to the initial prior belief under no commitment, whereas it is always weakly increasing under commitment. As a direct consequence of this, the average

duration can be either shorter or longer with commitment than without: in other words, there exists an equilibrium in which the principal prematurely terminates the project when she does not commit to a deadline.

This result is somewhat surprising, provided that the inefficiency of the problem stems from the principal's reluctance to terminate the project in the first place. The principal tends to terminate the project too early when the agent's productivity under low effort is sufficiently small. For the sake of argument, suppose that the agent fails almost surely when he exerts low effort, in which case the expected instantaneous payoff is negative while the principal can learn almost nothing from failures (during the phase where the agent is supposed to exert low effort). The principal's belief then declines very slowly over time, forcing her to incur a large amount of loss if she is to wait until she totally loses her confidence in the agent's ability. If this expected loss is prohibitively large, the principal may find it optimal to terminate the project even when the belief is still relatively high.

Of course, in equilibrium, the agent correctly anticipates this reaction and adjusts his effort allocation accordingly. We show that the equilibrium timing of project termination can be pinned down by backward induction, where we start from the final critical time (to be derived) and solve backward. This process gives rise to the aforementioned recursive equilibrium structure which allows us to establish the uniqueness of equilibrium and also obtain a diverse set of equilibrium dynamics under no commitment.

The implications of our analysis can be applied broadly to a range of circumstances in which a principal (an evaluator) must assess an agent's upside potential that is only gradually revealed in an ongoing relationship, e.g., a manager who must evaluate subordinates, a professor who must evaluate graduate students, a head coach in professional sports who must evaluate players, and so on. Among those possibilities, the most prominent example of evaluation schemes with deadlines is perhaps the "up-or-out system," which is widely observed in academia and professional service industries such as law, accounting, and consulting. As a specific application, our framework offers some insight for when up-or-out contracts are more valuable by identifying several key factors – such as high ability intensity, stable job descriptions, and similar jobs across ranks – which favor the use of an evaluation deadline from a previously unexplored channel. Each of these factors intensifies either the agent's incentive to delay exerting high effort for a given deadline or the principal's incentive to delay terminating the project for a given effort sequence (or both), thereby rendering it more profitable to set a deadline at the outset.

The paper is organized as follows. The literature review is provided in the remainder of this section. The model is presented in section 2 and analyzed in sections 3 and 4, where we

characterize both the commitment and no-commitment solutions. These solutions are then compared in section 5, in order to analyze the value of commitment and derive a necessary and sufficient condition under which it is strictly optimal to commit to a deadline. Some extensions of the baseline model and concluding remarks are offered in section 6.

Related Literature: The current analysis is most closely related to the experimentation literature in that the principal here attempts to uncover the agent’s type through a sequence of experiments.¹ From the principal’s point of view, our model can be seen as a variant of the canonical two-armed bandit problem with one safe arm (terminating the project) and one risky arm (continuing the project). Our model is particularly related to the literature on strategic experimentation which analyzes a situation where a group of individuals, rather than a single individual, face bandit problems (Bolton and Harris, 1999; Bergemann and Välimäki, 1996, 2000; Keller et al., 2005; Klein and Rady, 2011; Bonatti and Hörner, 2011). A crucial difference from this strand of literature is that experiments in our context are “intermediated,” i.e., experiments are conducted not by the principal herself but by an informed intermediary, the agent, who faces no uncertainty about the project.²

Recently, there have been increasingly many works that explore the optimal provision of incentives in bandit problems. Manso (2011) considers the classic two-armed bandit problem and shows that the optimal contract in this context must tolerate, or even reward, early failures in order to encourage exploratory activities. Bergemann and Hege (1997, 2005) and Hörner and Samuelson (2013) analyze a financing problem of a venture capitalist where the principal provides funding to the agent who conducts experiments on a project of unknown quality.³ Gerardi and Maestri (2012) consider a similar environment where an agent conducts experiments but assume that the outcome of each experiment can only be observed by the agent. The principal must hence devise a contract not only to induce costly effort but also to truthfully reveal the information. Halac et al. (2013) analyze a model of long-term contracting for experimentation with hidden information about the agent’s ability and dynamic moral hazard and obtain an explicit characterization of optimal contracts. Aside from

¹An early economic application of the bandit problem can be found in Rothschild (1974). See Bergemann and Välimäki (2008) for a survey.

²Several recent works analyze models of “delegated experimentation” where a principal delegates experimentation to an agent. Guo (2014) analyzes a situation where the principal can specify, with full commitment power, how the agent should allocate the resource in all future contingencies and solves for the optimal delegation rule. Garfagnini (2011) considers a similar setting to ours but assumes that: (i) the principal and the agent are symmetrically informed about the state of nature; and (ii) the agent’s payoff is independent of the state. Chen and Ishida (2015) consider the opposite case in which the principal, with the termination right, may be privately informed about the project quality while the agent focuses on implementing the project assigned to him.

³Also, see Buisseret (2016) who considers a two-period model of this setting with a more general, convex cost function.

some technical differences,⁴ these previous works are primarily concerned with characterizing optimal contracts. In contrast, the aim of this paper is to compare the allocations under full commitment and no commitment, while considering a less complete contractual environment, in order to evaluate the extent to which the principal can gain from committing to a deadline at the outset. Although it is hard to tell *a priori* which specification is more plausible, as it depends on the details of the underlying situation, our simple and tractable framework allows us to identify and illuminate the counterintuitive role of commitment in dynamic performance evaluation.

Several recent works examine the role of commitment in dynamic moral hazard setting. Mason and Välimäki (2015) consider a dynamic moral hazard problem and derive optimal wage contracts, both with and without commitment on wage payments. Hörner and Samuelson (2016) consider a repeated-game setting in which the principal chooses the scale of the project in addition to contingent payments and characterize the set of equilibrium payoff vectors that can be achieved without commitment.⁵ A key difference from our model is that these two works do not consider the principal’s learning, which is our main focus, with no hidden information about the agent’s type; as a consequence, the project is never terminated in their models.⁶ This stands in sharp contrast to our setting in which the principal accumulates information about the agent over time, and the project must be terminated in finite time.⁷

It is well known that players often wait until the deadline to reach an agreement in finite-horizon models. This behavior, which is referred to as the deadline effect, is a topic of utmost concern in many bargaining and war of attrition models (Hendricks et al., 1988; Spier, 1992; Fershtman and Seidman, 1993; Hörner and Samuelson, 2011; Chen, 2012; Damiano et al., 2012; Fuchs and Skrzypacz, 2013). Some recent works also explore the role of deadlines in dynamic problems with multiple agents. Bonatti and Hörner (2011) analyze a dynamic

⁴As a key technical difference, we consider a case where the agent knows his own productivity (or the project quality), so that our model belongs to the class of dynamic signaling (with stochastic signals), rather than of experimentation, from the agent’s point of view.

⁵In their model, outcomes are privately observed by the agent, and the moral-hazard problem regards the truthful disclosure of this private information.

⁶Georgiadis et al. (2014) analyze the role of commitment in a dynamic contribution games where the manager has the decision right over the project size. With multiple agent types and dynamic learning, Bonatti and Hörner (2017) analyze a (symmetric-information) experimentation model in which wages are determined competitively a la Holmström (1999) and characterize effort and wage dynamics with an exogenous termination date.

⁷A subtle technical difference which directly arises from this fact lies in the ways in which to construct an equilibrium. Mason and Välimäki (2016) and Hörner and Samuelson (2017) construct an equilibrium via arguments found in the analysis of infinitely repeated games (reversion to the worst continuation equilibrium), and as such, their analyses yield a non-degenerate set of equilibria. In our analysis, an equilibrium is obtained via backward induction which gives rise to a unique equilibrium.

moral-hazard problem with a team in which multiple agents work on a project of unknown quality and briefly discuss the optimal deadline in this context. Campbell et al. (2013) consider a similar environment where two agents work jointly on a project. They assume that there is only one project type but assume that one’s own outcomes are his private information. Each agent can exert effort to produce a breakthrough individually, and a successful agent can reveal that he has been successful. The focus of these works is placed on the interaction between the agents, especially the freerider problem, whereas ours is on the dynamic interaction between the principal’s termination decisions and the agent’s effort choices.

2 A dynamic model of performance evaluation

Environment: We employ a continuous-time model because of its greater tractability. Consider a situation where a principal (female) hires an agent (male) to complete a project. The game ends either when the agent attains a success (a “breakthrough”) or when the principal terminates the project. The agent is either good with prior probability $p_0 \in (0, 1)$ or mediocre with probability $1 - p_0$. The ability type is the agent’s private information and is not directly observable to the principal who must instead evaluate it from a sequence of observed outcomes.

Production: The agent makes unobservable effort $a_t \in \{l, h\}$ at each instant t .⁸ We interpret that low effort ($a_t = l$) refers to the minimum level of effort that can be induced via input monitoring while high effort ($a_t = h$) refers to any part of effort that cannot be directly monitored by any means. The instantaneous cost of effort a is denoted by d_a where $d_h = d > 0$ and $d_l = 0$. A success arrives stochastically, depending on the effort choice as well as the agent’s type. If the good type chooses $a_t = a$ over time $[t, t + dt)$, he attains a success with probability $\lambda_a dt$ where $\lambda_h > \lambda_l > 0$. In contrast, the mediocre type can never succeed with any effort level.⁹ Define $\Delta_\lambda := \lambda_h - \lambda_l$.

Payoff: We consider an incomplete-contracting environment where contingent rewards on the arrival of a success cannot be enforced.¹⁰ A success yields a net present value of $y > 0$

⁸Our focus on binary effort reflects our implicit presumption that the effort cost and success probability are linear, as usually assumed in this literature (e.g., Keller et al, 2005; Bonatti and Horner, 2011, for most of their analysis). As long as this structure is maintained, an extension to continuous effort, say $a_t \in [l, h]$, yields exactly the same allocation and is hence irrelevant.

⁹Our model specification is thus the “breakthrough” type in which one success can resolve all the uncertainty regarding the agent’s type – an assumption that is predominant in the experimentation literature. See, for instance, Keller et al. (2005) and Bonatti and Hörner (2011, 2017).

¹⁰In reality, we rarely observe complete wage contracts in industries comprised of professionals. For instance, few academic institutions, if any, offer rewards specifically contingent on verifiable measures of output: salaries

to the principal and $b > 0$ to the agent;¹¹ otherwise, they both receive zero. Aside from this, the principal must also pay a flow wage $w > 0$ to the agent as long as the project continues.¹² We assume that w is exogenously given for most part; the case with an endogenous wage is briefly discussed in the concluding section. The reservation payoff is assumed to be zero for both players. The common discount rate is denoted by $r \in (0, \infty)$.

Contract: The only contractible decision for the principal in this environment is whether to set a deadline, and if so, at what point. If the principal commits to a deadline, she terminates the project at the deadline (but never before) if the agent has not attained a success up to that point. If the principal chooses not to commit to any specific deadline, on the other hand, she retains discretion to terminate the project at any instant. As mentioned earlier, our analytical focus is on the latter case which requires that both players' strategies be sequentially rational.

3 Analysis

3.1 Agent's effort decision

The agent decides whether or not to exert high effort at each instant. To analyze this problem, it is important to note that the principal's belief affects the agent's payoff only through her termination decision. The agent's optimal effort choice thus depends only on the remaining time to the termination date (hereafter, simply the remaining time), i.e., the maximum length of time for which the principal continues the project without attaining a success. The remaining time is obvious when the principal sets a deadline at time τ , in which case the remaining time at time t is simply given by $\tau - t$. Even without such an explicit commitment, however, the remaining time can be computed from the principal's equilibrium strategy in essentially the same manner as we shall discuss below. For now, we proceed with the presumption that the remaining time exists and is well-defined at any given point in time.

are determined through bilateral negotiations, often dictated by market forces, in some countries whereas they are subject to bureaucratic regulations in others. One possible reason for the lack of complete wage contracts in those industries is that it is often difficult, and perhaps prohibitively costly, to measure the exact value of a "success" in a verifiable manner. Additionally, this assumption makes our analytical framework applicable to a wider range of circumstances. One such possibility is that the benefit of achieving a success accrues from non-transferrable (psychological) gains such as prestige, authority, and the sense of achievement, and extrinsic rewards are hence of secondary importance. There are also many cases where contingent monetary transfers are neither feasible nor desirable, as in a professor-student relationship.

¹¹An obvious interpretation is that y and b represent the continuation payoffs of achieving a success, including not only the intrinsic value of a successful outcome but also other benefits of identifying/signaling talent.

¹²Alternatively, w can be regarded as a flow cost of production (e.g., hiring an agent). Of course, under this interpretation, w is no longer a transfer payment to the agent and does not appear in his payoff. It is straightforward to make the setup consistent with this interpretation without having any qualitative impact on our results.

The agent's problem is rather straightforward since the mediocre type, knowing that his marginal value of effort is zero, never exerts high effort. As such, we can focus on the good type whom we refer to simply as the agent in what follows. Denote by $U(k)$ the agent's value function when the remaining time is k . Taking $k > 0$ as given, the value function can be written as

$$U(k) = \max_{a \in \{l, h\}} ((\lambda_a b - d_a + w)dt + e^{-rdt}(1 - \lambda_a dt)U(k - dt)).$$

Taking the limit $dt \rightarrow 0$, we obtain the Bellman equation:

$$rU(k) = \max_{a \in \{l, h\}} (\lambda_a b - d_a + w - \lambda_a U(k) - \dot{U}(k)), \quad (1)$$

with $\lim_{k \downarrow 0} U(k) = 0$. It is clear from this that the agent chooses high effort if and only if

$$\Delta_\lambda(b - U(k)) \geq d.$$

As usual in this type of setup (see, e.g., Bergemann and Hege, 2005), the continuation payoff $U(k)$ captures the agent's reservation payoff which he receives in case of a failure. The cost of not succeeding today is obviously small when he has a high reservation payoff. Since the continuation payoff is higher when the agent has more time to prove himself, the agent has a stronger incentive to work hard as the project approaches the termination date. This is a manifestation of the deadline effect that lies at the core of our entire analysis.

Proposition 1 *If*

$$\frac{d}{\Delta_\lambda} > \frac{rb - w}{\lambda_l + r}, \quad (2)$$

there exists k^ such that the good type exerts high effort if the remaining time is less than or equal to k^* and low effort otherwise. The threshold k^* is given by*

$$k^* = \begin{cases} -\frac{1}{\lambda_h + r} \ln\left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w}\right) & \text{if } b > \frac{d}{\Delta_\lambda}, \\ 0 & \text{if } \frac{d}{\Delta_\lambda} \geq b. \end{cases} \quad (3)$$

If (2) does not hold, the agent always exerts high effort.

Proof: See Appendix.

When $\Delta_\lambda b \leq d$, the static incentive is too weak for the agent to exert high effort for any remaining time. In contrast, when (2) fails to hold, the static incentive is strong enough to overcome the dynamic agency cost, and the agent is willing to supply high effort under any circumstance. As these cases only result in trivial solutions and are clearly of less interest for the question we pose here, we restrict our attention to the case where the strength of the static incentive lies in some intermediate range by making the following assumption.

Assumption 1 $\frac{\lambda_h b - d + w}{\lambda_h + r} > b - \frac{d}{\Delta_\lambda} > 0$.

Among other things, the assumption implies that the optimal threshold k^* is bounded from above and away from zero, i.e., $k^* \in (0, \infty)$.

3.2 Principal's termination decision

Let p_t denote the principal's belief that the agent is good at time t , conditional on no success having occurred. Given some effort sequence $\{a_s\}_{s=0}^t$, the updated belief is then given by

$$p_t = \frac{p_0 e^{-\int_0^t \lambda_{a_s} ds}}{1 - p_0 + p_0 e^{-\int_0^t \lambda_{a_s} ds}}. \quad (4)$$

Alternatively, taking the time derivative, we obtain

$$\dot{p}_t = -\lambda_{a_t} p_t (1 - p_t). \quad (5)$$

It is clear that the belief is strictly decreasing over time for any effort choice (due to the fact that $\lambda_l > 0$) until the agent attains a success, in which case the belief immediately jumps up to one.

The principal's problem is to determine when to terminate the project, conditional on no success having occurred. Now suppose that the principal intends to terminate the project at time $\tau > t$. Taking the effort sequence as given, the principal's continuation payoff can also be written as a function of the remaining time k and the current belief p_t where

$$V(k, p_t) = \int_t^{t+k} (\lambda_{a_s} p_s y - w) e^{-\int_t^s (\lambda_{a_u} p_u + r) du} ds,$$

subject to (5).

It is immediate to see that for a given effort level, there exists a threshold belief below which the principal's instantaneous payoff is strictly negative. We denote by $q_a := \min\{\frac{w}{\lambda_a y}, 1\}$ the "break-even" belief at which the instantaneous payoff equals zero under effort a . Combined with the fact that the belief is strictly decreasing over time for any effort sequence, q_h represents the absolute lower bound of the belief, as the principal clearly has no incentive to continue the project once her belief dips below this level. Since the belief must reach this level sooner or later (due to the fact that $\lambda_l > 0$ and $w > 0$), the presence of such a lower bound suggests that the game must end in some finite time. This allows us to solve the game by backward induction.

Finally, if the value of a success is too small, the model only admits a trivial solution where the principal chooses to stop immediately (or not to hire the agent in the first place). In what follows, therefore, we assume that the value of a success is large enough for the principal to hire the agent at least for some positive duration.

Assumption 2 $p_0 > q_h = \frac{w}{\lambda_h y}$.

Since the principal's continuation payoff depends on her belief, it is often convenient for the subsequent analysis to explicitly relate the current and termination beliefs to the remaining time. Suppose that the current belief is p and the principal intends to terminate the project when the belief reaches q . If the agent adopts the best response, the remaining time can be written as $K(p, q)$ such that

$$\frac{pe^{-\lambda_l \max\{K(p,q)-k^*, 0\}} - \lambda_h \min\{K(p,q), k^*\}}{1 - p + pe^{-\lambda_l \max\{K(p,q)-k^*, 0\}} - \lambda_h \min\{K(p,q), k^*\}} = q,$$

which reduces to

$$\frac{pe^{-\lambda_l K(p,q) - \Delta_\lambda k^*}}{1 - p + pe^{-\lambda_l K(p,q) - \Delta_\lambda k^*}} = q,$$

if $K(p, q) > k^*$. In addition, it is also convenient to define a backward operator $\phi(q)$ such that

$$\frac{\phi(q)e^{-\lambda_h k^*}}{1 - \phi(q) + \phi(q)e^{-\lambda_h k^*}} = q.$$

The backward operator suggests that if $p_t = \phi(q)$ and the agent exerts high effort for $t \in [t, t + k^*]$, then $p_{t+k^*} = q$; alternatively, if the agent expects the project to be terminated at $p_t = q$, $\phi(q)$ indicates the belief at which he must switch to high effort. Note that these two notions are closely related in that $K(p, q) > k^*$ if and only if $p > \phi(q)$.

4 Equilibrium

4.1 The cooperative solution: a benchmark

Before we move on to analyze our model, we first examine as a benchmark the case where both the principal and the agent attempt to maximize the sum of their individual payoffs to derive the first-best allocation. More precisely, the instantaneous payoff in the cooperative case, common to both players, is the difference between the expected benefit from a success minus the effort cost, i.e., $\lambda_a p(y + b) - d_a$ for a given effort a .¹³

The agent's problem is essentially the same and only needs a slight modification. Letting $\tilde{U}(k)$ denote the agent's continuation payoff, the agent exerts high effort if and only if

$$\Delta_\lambda((y + b) - \tilde{U}(k)) \geq d.$$

¹³If the agent shared the same objective as we assume here, his private information could in principle be induced at no cost: given that the agent's reservation payoff is zero, the mediocre type is indifferent between participating and dropping out immediately (before time 0), as he would obtain zero payoff in either case for any given deadline. As it turns out, however, the optimal deadline is independent of the initial belief p_0 and the analysis holds irrespective of whether or not the principal can immediately screen out the mediocre type.

As in the noncooperative case, there exists a threshold \tilde{k}^* such that the good type exerts high effort if and only if the remaining time is less than or equal to \tilde{k}^* . Under the maintained assumptions, one can easily verify that the agent is generally more motivated in the cooperative case than in the noncooperative case, i.e., $\tilde{k}^* > k^*$.

Given this, it is fairly straightforward to derive the cooperative solution. Since there is no social cost of hiring the mediocre type, the joint surplus from hiring the agent is strictly positive for any belief level, and as such, it is efficient for the principal to continue the project indefinitely until a success is attained. This implies that the good type will eventually succeed at some point, although he only exerts low effort all the way if \tilde{k}^* is finite.

Proposition 2 *In the cooperative case, the principal never terminates the project, and the game continues indefinitely. The agent always exerts low effort if*

$$\frac{d}{\Delta_\lambda} > \frac{r(b+y)}{\lambda_l + r}, \quad (6)$$

and high effort otherwise.

Proof: See Appendix.

Note the difference between (2) and (6). In the cooperative case, the joint expected payoff of exerting low effort indefinitely is $\frac{r(b+y)}{\lambda_l+r}$ because the value of a success is now $b+y$, instead of just b , and there is no flow cost w .

4.2 The commitment solution

The principal's problem under commitment is to choose a termination date τ which maximizes her expected payoff at time 0 while restricting attention to deterministic deadlines.¹⁴ We focus exclusively on deterministic deadlines primarily because it is generally difficult to commit to a lottery, even when it is possible to commit to a particular termination date. To see this point, consider a contract where the principal terminates the project at τ^1 or τ^2 , $\tau^1 < \tau^2$, each with probability strictly less than one. At time τ^1 , however, the principal's preferences are generically strict, i.e., it is strictly better either to stop (the continuation payoff is negative) or to continue (positive), so that there is no incentive to randomize at that point. If it is strictly better to stop at time τ^1 , for instance, the principal stops at time τ^1 with probability one but

¹⁴If stochastic deadlines are feasible, the principal may screen out the low type by offering a menu of contracts, one with a deterministic deadline and the other with stochastic ones. We do not pursue this possibility both because we do not think that it is realistic (given the enforcement problem) and because it is outside the scope of our analysis (given that our focus is on comparing the commitment and no-commitment solutions). Note that this screening cannot be done if deadlines are deterministic, because both types always strictly prefer a longer deadline.

knowing that, the contract is effectively reduced to the one with a deterministic deadline. The lack of credible enforcement is perhaps the reason why we almost never observe a contract with stochastic deadlines.

As we have already seen, the agent exerts high effort if and only if the remaining time is less than or equal to k^* . We denote by $V^*(k, p_t)$ the continuation payoff when the agent adopts this best response, where k is the remaining time and p_t is the principal's belief. This can be written as

$$\begin{aligned} V^*(k, p_t) &= \int_t^{t+\nu} (\lambda_l p_s y - w) e^{-\int_0^s (\lambda_l p_u + r) du} ds \\ &\quad + e^{-\int_t^{t+\nu} (\lambda_l p_u + r) du} \int_{t+\nu}^{t+k} (\lambda_h p_s y - w) e^{-\int_\nu^s (\lambda_h p_u + r) du} ds \\ &= \frac{p_t \pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\nu}) + \frac{p_t e^{-(\lambda_l + r)\nu} \pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)(k-\nu)}) - \frac{(1 - p_t)w}{r} (1 - e^{-rk}), \end{aligned} \quad (7)$$

where $\nu := \max\{k - k^*, 0\}$ and $\pi_a := \lambda_a y - w$. The first two terms represent the expected gain from the good type whereas the last term represents the expected loss from the mediocre type.

Since the termination date is equivalent to the remaining time at time 0, the commitment solution, denoted by τ^C , is given by

$$\tau^C(p_0) := \operatorname{argmax}_\tau V^*(\tau, p_0).$$

The continuation payoff is continuous but kinked at $\tau = k^*$. For $\tau \in (0, k^*)$, the first-order condition is given by

$$p_0 \pi_h e^{-\lambda_h \tau} - (1 - p_0)w = 0. \quad (8)$$

For $\tau \in (k^*, \infty)$, it is given by

$$M(\tau, p_0) := p_0 e^{-(\lambda_l + r)(\tau - k^*)} \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h (1 - e^{-(\lambda_h + r)k^*}) \right) - (1 - p_0)w e^{-r\tau} = 0. \quad (9)$$

To better understand this condition, it is instructive to rewrite this as

$$(p_0 \pi_h e^{-(\lambda_l(\tau - k^*) + \lambda_h k^*)} - (1 - p_0)w) e^{-r\tau} = \frac{p_0 \Delta \lambda e^{-(\lambda_l + r)(\tau - k^*)}}{\lambda_h + r} (w + ry + \pi_h e^{-(\lambda_h + r)k^*}).$$

Note that the left-hand side is the expected instantaneous payoff for the principal at the time of termination, whereas the right-hand side is the marginal cost of extending the deadline. Since the right-hand side is strictly positive, the principal must set the deadline at a point where the expected instantaneous payoff is positive if $\tau > k^*$.

Since the instantaneous payoff is negative for any effort level once the belief dips below q_h , there is clearly no incentive to set $\tau > K(p_0, q_h)$. When the initial belief is sufficiently close to q_h ($\phi(q_h) \geq p_0$), it is optimal to let the agent work until the belief reaches this lower bound. When it is far from it ($p_0 > \phi(q_h)$), on the other hand, the principal must terminate the project at a point where the belief is still above the lower bound. There are two cases, depending on the value of $M(k^*, p_0)$.¹⁵ If

$$\frac{\lambda_l + r}{\lambda_h + r} \pi_h (1 - e^{-(\lambda_h + r)k^*}) > \pi_l, \quad (10)$$

$M(k^*, p_0) < 0$ for any $p_0 \in (q_h, 1)$, and we let $\hat{p} = 1$ in this case. If (10) fails to hold, on the other hand, there exists a unique \hat{p} such that $M(k^*, \hat{p}) = 0$. Then, for $p_0 > \hat{p}$, there exists a unique interior solution $\hat{\tau}(p_0)$ such that $M(\hat{\tau}(p_0), p_0) = 0$. When the initial prior is in this range, it is not so costly to have a phase where the agent exerts low effort, and as such, the principal would wait beyond time k^* . Of course, extending the deadline beyond time k^* means that the agent would slack off at the beginning, thereby pushing back the realization of the high-effort phase. The principal trades off the gain of inducing low effort for an additional instant against the cost of realizing the high-effort phase an instant later and sets the termination date at a point where they are equalized.

For clarity, we summarize the set of parameters by $\Theta := (b, y, w, d, \lambda_h, \lambda_l, r)$. We can then make the following statement.

Proposition 3 *For any given set of parameters (Θ, p_0) satisfying Assumptions 1 and 2, there exists a unique commitment solution $\tau^C(p_0) \in (0, K(p_0, q_h)]$. For $\phi(q_h) \geq p_0$, the commitment solution is given by $\tau^C(p_0) = K(p_0, q_h)$. For $p_0 > \phi(q_h)$, there exists some $\hat{p} \in (q_l, 1]$ such that*

$$\tau^C(p_0) = \begin{cases} k^* & \text{if } \hat{p} \geq p_0 > \phi(q_h), \\ \hat{\tau}(p_0) & \text{if } p_0 > \hat{p}, \end{cases}$$

where: (i) $\hat{p} > q_l$ if $1 > q_l$; and (ii) $\hat{p} = 1$ if and only if (10) holds. The commitment solution is weakly increasing in p_0 .

Proof: See Appendix.

Note that $K(p_0, q_h)$ constitutes the upper bound of $\tau^C(p_0)$. Since $K(p_0, q_h)$ is bounded from above when $q_h > 0$, the proposition suggests that the project must be terminated in finite time, meaning that some good projects are bound to be terminated prematurely. This draws clear contrast with the cooperative solution where the project is never terminated.

¹⁵Note that $M(k, p_0) < 0$ for any $k > k^*$ and p_0 if and only if $M(k^*, p_0) < 0$.

4.3 The no-commitment solution

The situation becomes more complicated, and perhaps more intriguing, when the principal makes no commitment at the outset and hence her termination decision must be sequentially rational. Because of this, the principal's termination strategy now needs to be specified in a different way, as her strategy off the equilibrium path (after the project is supposed to be terminated) may matter. More precisely, the principal's strategy is given by a set of termination dates T which consists of a first termination date and further dates along off-equilibrium continuation paths. Formally, for this no-commitment case, we solve for a pure-strategy perfect Bayesian equilibrium (hereafter, simply an equilibrium) in which the principal terminates the project with probability one at each termination date (a formal definition is given below).¹⁶

To illustrate how we specify the principal's strategy, consider a simple example where the principal's strategy consists of two distinct termination dates, i.e., $T = \{\tau^1, \tau^2\}$ where $\tau^1 < \tau^2$. Under this strategy, the principal terminates the project with probability one at each τ^i , $i = 1, 2$, conditional on the continuation of the game.¹⁷ This means that the project is surely terminated at time τ^1 on the equilibrium path. To show that this constitutes an equilibrium, however, one must assure that the principal cannot profitably deviate from this strategy. To see this, suppose that the principal unexpectedly continues the project at time τ^1 . The game then enters into a new phase where the agent now expects the project to be terminated at the next termination date τ^2 and chooses his effort accordingly ($a_t = h$ if $k^* \geq \tau^2 - t$ and $a_t = l$ otherwise). Note also that the principal's belief about the agent's type is not affected by the deviation: given T and the agent's best response, we can still apply Bayes' rule to compute how the belief evolves both on and off the equilibrium path.¹⁸ This gives the continuation payoff when the principal deviates which must be non-positive.

How long does this process continue? It generally goes on until the principal has no

¹⁶We restrict our attention to pure-strategy equilibria by assuming that the principal terminates the project with probability one when the continuation payoff is zero. Since the principal is actually indifferent at this point, however, there may exist other mixed-strategy equilibria in which the principal terminates the project with probability less than one at each termination date. We do not pursue this possibility for two reasons. First, we do not believe that such an equilibrium is particularly realistic and hence appealing. Second, even with randomization, the game still must end with probability one by time τ^1 , and the continuation equilibrium is exactly the same once the game enters the final interval (τ^1, τ^2) ; solving backwards, one can see that the structure of equilibrium is essentially the same where the agent starts exerting high effort as t approaches each termination date and switches back to low effort if the project is not terminated at the termination date.

¹⁷If the game continues beyond time τ^2 , the principal terminates the project immediately as we will see below.

¹⁸Technically, we assume that the belief depends only on the agent's past effort and is given by (4) even off the equilibrium path. This is actually the "no signaling what you don't know" condition (Fudenberg and Tirole, 1991) although it is a rather obvious restriction in games with two players (it only implies that the principal's belief should not be affected by her own deviation).

incentive to continue the project regardless of the agent's effort choice or, in other words, until the belief reaches the lower bound q_h . With slight abuse of notation, we denote by p_t the equilibrium belief at time t , conditional on the continuation of the game (including the off-equilibrium path). Let τ^n denote the time at which p_t reaches q_h for some n (which is unknown at this point). We can then restrict our attention to $t \in (0, \tau^n]$, for it is a dominant strategy for the principal to terminate the project once the belief reaches this lower bound.¹⁹

Lemma 1 *In any equilibrium, T is a finite set.*

Proof: The lemma is directly implied by the fact that any two adjacent termination dates must satisfy $\tau^i - \tau^{i-1} \geq k^* > 0$. To see this, suppose otherwise, i.e., $k^* > \tau^i - \tau^{i-1}$. Suppose further that the principal deviates and continues the project at time τ^{i-1} . Then, it is strictly better for the agent to exert high effort because $k^* > \tau^i - \tau^{i-1}$. Note, however, that since $p_t > q_h$ for any $t \in (0, \tau^n)$ by definition, the instantaneous payoff is strictly positive until the belief reaches the next termination belief. This is a contradiction because the principal can strictly benefit from deviating and not stopping at time τ^{i-1} . Given this, it is clear that we can only have finitely many termination dates in $(0, \tau^n]$. ■

Given this result, we denote each element of T by τ^i where $\tau^1 < \tau^2 < \dots < \tau^n$ for some n . Accordingly, the game is divided into n distinct segments T^i , $i = 1, 2, \dots, n$ by termination dates, where $T^i = (\tau^{i-1}, \tau^i)$ for $i = 1, 2, \dots, n$ with $\tau^0 = 0$. The remaining time at time t is now given by $\tau^i - t$ for $t \in T^i$ while it is zero for all $t > \tau^n$ (see footnote 19). Let $q^i := p_{\tau^i}$ denote the corresponding termination belief. Only the first segment T^1 is actually played on the equilibrium path, as the game ends with probability one by the time the game reaches time τ^1 . For expositional clarity, we call τ^1 the no-commitment solution and denote it by $\tau^{NC}(p_0)$ to indicate its dependence on p_0 . The formal definition of our equilibrium is given below.

Definition 1 *A perfect Bayesian equilibrium in this game is a pair of strategies $\{a_s\}_{s=0}^\infty$ and $T := \{\tau^i\}_{i=1}^n$ and a belief system such that:*

- *given T , the (good-type) agent chooses a_t at each t to maximize his continuation payoff, i.e., chooses high effort if and only if the remaining time at time t is less than or equal to k^* ;*

¹⁹ Suppose that the principal deviates and continues beyond time τ^n . In this case, the principal's instantaneous payoff is negative even if the agent chooses $a_t = h$, so that it is a dominant strategy to terminate immediately for all $t > \tau^n$. Given this, since the remaining time is invariably zero, the agent always chooses $a_t = h$. This is the unique continuation equilibrium after time τ^n (although, to describe this process formally, we need to consider a discrete-time counterpart of our model where there is a minimum time unit, and take the limit as the time unit goes to zero).

- given $\{a_s\}_{s=t}^{\infty}$ and the current belief, the principal terminates the project at time t if and only if the continuation payoff is non-positive;
- The belief at time t depends only on the effort sequence $\{a_s\}_{s=0}^t$ and is given by (4), both on and off the equilibrium path.

To pin down the equilibrium termination dates, we need to pay closer attention to the principal's continuation payoff. Since the agent's best response is the same, the principal's continuation payoff can still be written as $V^*(k, p_t)$. The only difference is that the principal's strategy must now be sequentially rational, which amount to the following two equilibrium conditions which the principal's termination strategy T must satisfy.

Condition T: (i) $q^n = q_h$ and (ii) if $n > 1$, for each $i = 2, \dots, n$, $V^*(K(q^{i-1}, q^i), q^{i-1}) = 0$.

Condition C: For each $i = 1, 2, \dots, n$, $V^*(\tau^i - t, p_t) > 0$ for all $t \in T^i$.

Condition T is the usual indifference condition that requires the principal to terminate the project when the continuation payoff (when she deviates) is non-positive at each τ^i . This indifference condition alone is in general not sufficient because the instantaneous payoff may not be monotonically decreasing over time: in any segment T^i , the agent may start off with low effort, during which the instantaneous payoff could be so small that the principal is tempted to stop prematurely. Condition C assures that the principal does not stop before the intended termination date τ^i is reached.

The next result establishes that there exists a perfect Bayesian equilibrium that is always unique even under no commitment.

Proposition 4 *For any given set of parameters (Θ, p_0) satisfying Assumptions 1 and 2, there exists a (generically) unique pure-strategy equilibrium. Given Θ , the belief space $(q_h, 1)$ is partitioned into $m(\Theta) \geq 1$ distinct intervals $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ where $P^{m(\Theta)} = q_h$ and $P^0 = 1$. For $p_0 \in (P^j, P^{j-1})$, $j = 1, 2, \dots, m(\Theta)$, the equilibrium strategies are as follows:*

- The principal's equilibrium strategy is characterized by a set of $n = m(\Theta) - j + 1$ distinct termination dates $\{\tau^i\}_{i=1}^n$ where $q^i = P^{i+j-1}$ and $\tau^i - \tau^{i-1} > k^*$,²⁰
- The agent exerts high effort if the remaining time is less than or equal to k^* and low effort otherwise.

Moreover, if $m(\Theta) > 1$, then $q_l > P^1$.

²⁰As discussed in footnote 19, $a_t = 0$ for all $t \geq \tau^n$.

Proof: See Appendix A.

As the proposition suggests, the structure of equilibrium is thoroughly characterized by how the belief space is partitioned into intervals. A crucial determinant of the partition $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ is the profitability under low effort which is captured by π_l .²¹ Below, we first briefly illustrate how we pin down $m(\Theta)$ and $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ from a given Θ (in what follows, we simply denote $m = m(\Theta)$ to save notation). Once they are obtained, it is quite straightforward to derive the equilibrium strategy T for a given p_0 .

Suppose first that the profitability under low effort is sufficiently high, so that the principal's instantaneous payoff is positive even under low effort. In this case, the principal is not tempted to stop prematurely, and only Condition T is sufficient to pin down the equilibrium. Figure 1 depicts this situation where $m = 1$: the agent starts off with low effort and switches to high effort when the remaining time is k^* ; the principal stops when the belief reaches the lower bound $P^1 = q_h$ at time τ^1 . Formally, $m = 1$ and hence $n = 1$ for all $p_0 \in (q_h, 1)$ if and only if $V^*(K(q_l, q_h), q_l) > 0$, i.e.,

$$\frac{q_l \pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\nu}) + \frac{q_l e^{-(\lambda_l + r)\nu} \pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)(k - \nu)}) > \frac{(1 - p_t)w}{r} (1 - e^{-rk}), \quad (11)$$

where $\mu = \max\{K(q_l, q_h) - k^*, 0\}$ (see the proof of Proposition 4, especially Lemma 2, for more detail). Note that (11) depends only on Θ . The no-commitment solution $\tau^{NC}(p_0)$ is monotonic in p_0 if and only if this condition holds.

[Figure 1 about here]

In contrast, the instantaneous payoff may become negative under low effort when the success probability under low effort is relatively low, in which case we may have a situation where the principal's continuation payoff also becomes negative before the belief reaches q_h . Figures 2 and 3 show the evolution of the belief and the expected payoff when $m = 2$ while fixing the principal's strategy at $T = \{\tau^1\}$. In the figures, q_l is so high that the instantaneous payoff is negative for the entire interval during which the agent exerts low effort, and there exists a point τ' (in Figure 3) such that the continuation payoff is negative for $t \in (0, \tau')$. This implies that $T = \{\tau^1\}$ does not constitute an equilibrium as it violates Condition C. In this case, the game is divided into two segments, T^1 and T^2 , as illustrated in Figure 4. The pair of strategies now satisfies the equilibrium conditions since the instantaneous payoff is always positive in T^1 . Formally, if $V^*(K(q_l, q_h), q_l) < 0$, there must exist $p' \in (q_h, q_l)$ such

²¹Note that the partition $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ is determined solely by Θ while q^i depends also on the initial prior p_0 .

that $V^*(K(p', q_h), p') = 0$, in which case we redefine $P^1 = p'$ and $P^2 = q_h$. As above, $m = 2$ if and only if $V^*(K(q_l, P^1), q_l) > 0$.

[Figures 2-4 about here]

We can continue this process until we find P^1 such that $V^*(K(q_l, P^1), q_l) > 0$.²² Once the partition is pinned down, we can then easily derive the equilibrium strategies: for $p_0 \in (P^j, P^{j-1})$, the principal's equilibrium strategy consists of $n = m - i + 1$ termination dates where $q^i = P^{i+j-1}$ and

$$\tau^{NC} = \tau^1 = K(p_0, q^1), \quad \tau^i = \sum_{i'=1}^{i-1} \tau^{i'} + K(q^{i-1}, q^i) \text{ for } i = 2, 3, \dots, n.$$

Note also that $q^{i-1} > \phi(q^i) > q^i$ because the continuation payoff is always strictly positive when the belief is in $(q^i, \phi(q^i))$, the range where the remaining time is less than k^* and the agent exerts high effort. This alternatively means $\tau^i - \tau^{i-1} > k^*$, i.e., the length between any two adjacent termination dates must be larger than k^* .

Finally, we would also like to note that the constructed equilibrium is generically unique for any given set of parameters. This uniqueness result stems crucially from the fact that there is a lower bound of the belief q_h below which the principal would never continue the project. As stated above, together with the fact that $\lambda_l > 0$, this implies that the game must end in some finite time, which allows us to solve the game via backward induction analogously to the ‘‘gap case’’ of the durable-good monopoly problem (Fudenberg et al., 1985). More precisely, since the agent's strategy depends only on the remaining time, we know exactly how the game must end as the belief approaches the lower bound. Applying this reasoning backward, we can identify a unique continuation equilibrium for each $p_t > q_h$ and all the way back to the initial prior p_0 .

5 Discussion

5.1 The value of commitment

Our framework yields unique commitment and no-commitment solutions which enable us to directly assess the value of commitment to an evaluation deadline. Given that both τ^C and τ^{NC} can be written as functions of p_0 , the expected equilibrium payoffs can also be written as functions of p_0 . Let

$$V^C(p_0) := V^*(\tau^C(p_0), p_0), \quad V^{NC}(p_0) := V^*(\tau^{NC}(p_0), p_0).$$

²²When $q_l = 1$, there is no such q^1 and m will go to infinity. If $q_l < 1$, on the other hand, this process must converge after a finite number of rounds.

The value of commitment is then defined as

$$\Delta V(p_0) := V^C(p_0) - V^{NC}(p_0).$$

In principle, we can compute the exact value of $\Delta V(p_0)$ for any given p_0 . The following statement is the main result of the paper which analytically characterizes how the value of commitment varies with respect to p_0 .

Proposition 5 (i) $\Delta V(p_0) = 0$ for $p_0 \in (P^m, \phi(P^m))$ where $P^m = q_h$. (ii) If $m > 1$, for each $j = 2, 3, \dots, m$, ΔV is strictly increasing in p_0 for $p_0 \in (\phi(P^j), P^{j-1})$ and strictly decreasing for $p_0 \in (P^{j-1}, \phi(P^{j-1}))$ with $\Delta V(\phi(P^j)) = 0$. (iii) $\Delta V(\phi(P^1)) = 0$ if and only if $\hat{p} \geq \phi(P^1)$ where \hat{p} is the threshold defined in Proposition 3. (iv) $\lim_{p_0 \rightarrow 1} \Delta V(p_0) = 0$ if and only if $1 > \hat{p}$.

Proof: See Appendix.

The proposition suggests that the value of commitment changes in a non-monotonic way as p_0 increases, precisely due to the non-monotonic nature of V^{NC} . Despite this, we can still find some regularities except for the last segment $(\phi(P^1), 1)$.²³ Take an interval $(\phi(P^j), \phi(P^{j-1}))$ for $j = 2, 3, \dots, m$ where $\phi(P^{j-1}) > P^{j-1} > \phi(P^j)$. In each of these intervals, (i) the value of commitment is single-peaked and maximized at P^{j-1} (at which point the no-commitment solution yields a zero payoff); (ii) the value of commitment is zero at both ends of the interval (at which point the two solutions are identical). Important features of Proposition 5 are captured by Figure 5 which shows a typical path of the value of commitment when $m = 2$.²⁴

[Figure 5 about here]

To see how ΔV changes with respect to p_0 , observe that the only feasible history at any continuation game is the one consisting only of failures up to that point, and the principal hence has no additional information other than the fact that the agent has achieved no success. This immediately implies that the no-commitment solution can yield no higher payoff than the commitment solution, i.e., $\Delta V(p_0) \geq 0$ for all p_0 . Clearly, the no-commitment solution

²³In the last segment, the profitability under low effort can be high enough (if $p_0 > q_l$), and the expected payoff can go up even during the low-effort phase. Moreover, in this range, $\tau^C(\phi(P^1)) = \tau^{NC}(\phi(P^1))$ may not even be satisfied since $\tau^C(\phi(P^1))$ can be larger than k^* . As a consequence, the value of commitment can either go up or down (and can even be non-monotonic). See Appendix B for more detail.

²⁴For the figure, the parameters are set as follows: $\lambda_h = 0.1$, $\lambda_l = 0.04$, $b = y = 1$, $w = 0.035$, $d = 0.02$, $r = 0.05$.

gives the principal a different, and necessarily lower, payoff when $\tau^C \neq \tau^{NC}$. Still, the no-commitment solution can replicate the same allocation as the commitment solution when $\tau^C = \tau^{NC}$, because Proposition 1 ensures the unique equilibrium given the same horizon. It follows from Proposition 5 that the commitment solution generically yields a strictly higher expected payoff for the principal than the no-commitment solution if and only if $p_0 > \phi(q_h)$,²⁵ i.e.,

$$p_0 > \frac{q_h}{q_h + (1 - q_h)e^{-\lambda_h k^*}} \Leftrightarrow \frac{1 - q_h}{q_h} \left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right)^{\frac{\lambda_h}{\lambda_h + r}} > \frac{1 - p_0}{p_0}. \quad (12)$$

Conversely, when this condition fails to hold, the absence of commitment entails no additional cost, and the no-commitment solution becomes strictly profitable in the presence of a (possibly very small) commitment cost.²⁶

The proposition states that the principal benefits from committing to a deadline if the initial prior belief p_0 exceeds a certain threshold $\phi(q_h)$, i.e., if the initial prior is so high that the belief at time k^* is above the lower bound q_h even if the agent exerts high effort for $t \in [0, k^*]$. The value of a deadline depends crucially on each player's propensity to "procrastinate." In short, the threshold is lower and hence favors the commitment solution for a given initial belief either when the agent delays exerting high effort for a given deadline or when the principal delays terminating the project for a given effort sequence. Below, we briefly summarize how each player's propensity to procrastinate is determined:

1. The agent's propensity to procrastinate is determined by the tradeoff between the current gain of attaining a success and the potential loss of future payoffs. For a given deadline, the agent tends to procrastinate more (a small k^*) when the potential loss of future payoffs is large relative to the current gain.
2. The principal's propensity to procrastinate depends on the break-even belief and the likelihood of a success under high effort. The break-even belief is evidently the major force determining how patient the principal can be, where the principal is more tempted to wait for an eventual success when q_h is low. Further, for a given q_h , the principal tends to wait longer when λ_h is small, because the information about the agent's type is revealed only slowly in that case.

²⁵To be more precise, as Proposition 5 suggests, there could be some non-generic cases where the expected payoffs are identical.

²⁶An obvious commitment cost is that the principal must give up flexibility to adjust *ex post* to any uncertainty that may resolve during the course of play. Although we have thus far assumed away this aspect for clarity, we extend the analysis to incorporate uncertainty into the current setup in section 6: there, we show that the commitment solution is not always weakly optimal. The commitment cost may also arise from the cost of writing and enforcing a formal contract.

Before we move on, it is important to note that our characterization result owes largely to the fact that we consider a binary effort choice or, more importantly, that there is an upperbound for effort. In contrast, if the effort domain is unbounded with a convex effort cost, the commitment solution in general yields a strictly higher payoff than the no-commitment solution.²⁷ Even in this case, however, the nature of the problem remains largely intact. The intuition provided above carries over to this more general case as well because the value of a deadline is still determined in essentially the same way, although the analysis would be substantially more complicated.

5.2 Too early or too late?

In the absence of commitment, the principal often fails to stop at the right time because she has a strong incentive to wait for a success. This reasoning seems to indicate that the principal always terminates the project earlier with commitment than without, or equivalently, $\tau^{NC}(p_0) \geq \tau^C(p_0)$ for all p_0 . As indicated in Figures 3 and 4, however, this does not necessarily hold in the current setup. When the principal knows that she cannot stop soon enough, and it is too costly for her, she may terminate the project even when the instantaneous payoff is still strictly positive.

From Proposition 3, it is conceptually straightforward to compute the commitment solution τ^C . Figures 6a and 6b illustrate τ^C as a function of p_0 , using the same parameters as in Figure 5 except for λ_l . The qualitative nature of τ^C depends largely on the productivity under low effort as captured by λ_l . Figure 6a is the case where λ_l is relatively high. In this case, the principal does not lose much during the low-effort phase, and is willing to extend beyond k^* when the initial prior is sufficiently high. In contrast, when λ_l is relatively low, as depicted in Figure 6b, the principal sets $\tau^C \leq k^*$ to assure that the agent always exerts high effort. In either case, for $p_0 > \phi(q_h)$, τ^C is set strictly below the upper bound $K(p_0, q_h)$ so that the project is terminated before the belief reaches q_h .

[Figures 6a and b about here]

As we show in the proof of Proposition 4, we can also explicitly derive the no-commitment solution $\tau^{NC}(p_0)$ although it may take a much more complicated form. Figure 7 depicts $\tau^{NC}(p_0)$ along with $\tau^C(p_0)$, which confirms that the average duration of the project can be longer with commitment than without, using the same parameters as in Figure 5. The figure also reveals that the no-commitment may not be monotonically increasing in p_0 , in

²⁷On the other hand, we can obtain essentially the same result even if the effort level is continuous as long as the feasible effort level is bounded from above.

clear contrast to the commitment solution which is always weakly increasing. When the initial belief is relatively high, it takes long for the belief to reach the lowerbound q_h , and the principal may have to incur large losses if she is to wait until the moment at which the instantaneous payoff equals zero. To avoid this situation and discipline the agent, the principal must act early, i.e., she may need to pull the trigger earlier for *ex ante* more promising agents in some cases. This is more likely to be the case when λ_l is low, in which case the principal must incur losses in the low-effort phase. Formally, we can make the following statement.

Proposition 6 $\tau^{NC}(p_0)$ can be larger or smaller than $\tau^C(p_0)$. Moreover, $\tau^C(p_0) > \tau^{NC}(p_0)$ for almost all $p_0 \in (\phi(q_h), 1)$ as λ_l tends to zero.

Proof: It is clear that $\tau^{NC}(p_0)$ can be larger than $\tau^C(p_0)$ for some p_0 , because $\tau^{NC}(p_0) > k^*$ for p_0 slightly larger than $\phi(q_h)$ while $\tau^C(p_0) = k^*$ for $p_0 \in (\phi(q_h), \hat{p})$. We can also show that $\tau^{NC}(p_0)$ can be smaller if $m \geq 2$, because $\lim_{p_0 \downarrow P^1} \tau^{NC}(p_0) = 0 < k^* \leq \tau^C(P^1)$.

As for the second statement, It follows from Proposition 3 that $\tau^C(p_0) \geq k^*$ for $p_0 \in (\phi(q_h), 1)$. To prove the proposition, it thus suffices to show that $k^* > \tau^{NC}(p_0)$ in this range. Suppose otherwise, i.e., $\tau^{NC}(p_0) \geq k^*$ for some $p_0 \in (\phi(q_h), 1)$. This happens if and only if $p_0 \in (\phi(P^j), P^{j-1})$ for $j = 1, 2, \dots, m$ or, in other words, P^{j-1} must be bounded away from $\phi(P^j)$. This is a contradiction, however, as $\lim_{\lambda_l \rightarrow 0} V^*(K(P^{j-1}, P^j), P^{j-1}) < 0$ for any $P^{j-1} > \phi(P^j)$. ■

[Figure 7 about here]

Our results yield some empirical implications. An obvious one is that the initial prior p_0 is not necessarily a reliable predictor of future success under no commitment. Aside from this, it is also worth emphasizing that the equilibrium allocation depends crucially on what happens off the equilibrium path. As a consequence, two organizations that are observationally similar may exhibit drastically different equilibrium outcomes. To put this idea in context, consider two distinct organizations, A and B , each characterized by λ_l^i , $i = A, B$, and the corresponding no-commitment solution $\tau^{NC,i}$ (while assuming that the two organizations are equivalent in every other dimension). Suppose further that $\lambda_l^A > \lambda_l^B$ and $\lambda_l^B \rightarrow 0$. In this case, as Proposition 6 suggests, $\tau^{NC,B}(p_0) < k^*$ for almost all p_0 , meaning that we almost never observe the low-effort phase in organization B , and hence λ_l^B cannot be estimated from actual data. The two organizations are thus observationally indistinguishable even though their respective no-commitment solutions can be totally different from each other.

5.3 An application: when is up-or-out optimal?

The most prominent example of evaluation schemes with deadlines is arguably what is called the “up-or-out system” – a promotion scheme that is widely observed in academia and professional service industries such as law, accounting, and consulting (Lazear and Gibbs, 2014; Eriksson, 2016).²⁸ Under a typical up-or-out contract, the employer sets a deadline by which a promotion must occur, and the worker must leave the firm if he is not promoted by the deadline. In academia, it is often the case that incoming assistant professors must be promoted within a certain period of time (typically around six years). Outside of academia, a leading example of industries with up-or-out is the legal service industry. Many law firms traditionally adopt a set of managerial practices, known as the Cravath system, which include up-or-out (or “partnership track”) where incoming associates must make partner within a certain period of time (roughly seven to ten years). This contrasts sharply with other typical firm organizations which almost never specify deadlines by which workers must be promoted.

Here, we attempt to shed some light on this issue by examining under what conditions (12) is likely to be satisfied. There are of course some caveats. First and foremost, our theoretical framework is perhaps too simplistic to capture all the important details of this diverse set of industries; some of our arguments may thus be applied only to a subset of those industries.²⁹ Second, there is also a danger in comparing allocations under different incentives structures. One particular concern is the potential endogeneity of parameter values that are taken as exogenous here.³⁰ With those qualifications in mind, our framework can still provide some useful, though preliminary, insight for when it pays to commit to a promotion deadline as in the case of up-or-out.

Ability intensity: It is often argued that knowledge intensity is one of the most fundamental characteristics of professional service industries (von Nordenflycht, 2010). Combined with the fact that they are also less capital intensive,³¹ the productivity of an organization depends

²⁸For brevity, we refer to those industries characterized by up-or-out, including academia, broadly as professional service industries. The military is another example which is characterized by up-or-out, but we do not consider this case because the underlying structure appears to be quite different. See Kahn and Huberman (1988), Waldman (1990), O’Flaherty and Siow (1992), and Ghosh and Waldman (2010) for formal analyses of up-or-out contracts.

²⁹For instance, our exponential specification implies that success is very rare and information is coarse, as assumed in Bonatti and Hörner (2017), which might be applicable to some industries but not to some others. Also, team production may play a bigger role in some industries than in others.

³⁰It is likely that y and b may vary with the contract duration when we interpret them as the net continuation payoffs of achieving a success. In particular, we normalize the continuation payoff at zero when the project is terminated, but this could well be a function of the contract duration (or the market belief at the time of termination).

³¹High knowledge intensity and low capital intensity are not equivalent in the strict sense of the word because, as argued by von Nordenflycht (2010), an industry can be both knowledge- and capital-intensive at the same time. For the purpose of this study, however, we do not make any distinction between them.

crucially on the extent of knowledge embodied in individuals (Starbuck, 1992; Winch and Schneider, 1993). A likely consequence of this fact is that innate ability matters and creates value that cannot be easily substituted by sheer effort, at least in the short run – a feature which we call ability intensity for expositional clarity.

Since the mediocre type can never attain a success in the current setup, one way to measure the extent of ability intensity is by the (good type’s) success probabilities. We say that the project is more ability-intensive when λ_l and λ_h are larger with a fixed Δ_λ . Note that an increase in λ_h , while fixing Δ_λ , yields three conflicting effects. First, it raises the agent’s productivity at the margin, which in turn lowers the threshold q_h . Second, it gives the agent a stronger incentive to procrastinate (a lower k^*) as it raises his expected future payoff under high effort. Finally, it also facilitates the principal’s learning as each failure reveals more information. The first two effects raises the value of up-or-out while the last lowers it, such that the overall impact is not immediately clear.

One can readily show, however, that there exists some threshold $\bar{\lambda}$ such that it is strictly optimal to set a deadline, thereby favoring the use of up-or-out, for $\lambda_h > \bar{\lambda}$. To see this, we rewrite (12) as

$$\frac{p_0}{1 - p_0} > \frac{q_h}{(1 - q_h)e^{-\lambda_h k^*}} = \frac{w}{(\lambda_h y - w)e^{-\lambda_h k^*}}. \quad (13)$$

Under the maintained assumptions, $\lambda_h > \underline{\lambda} := \max\{\frac{w}{p_0 y}, \frac{\Delta_\lambda}{d}(r(b - \frac{d}{\Delta_\lambda}) + d - w)\}$. If $\frac{\Delta_\lambda}{d}(r(b - \frac{d}{\Delta_\lambda}) + d - w) > \frac{w}{p_0 y}$, we have $\lim_{\lambda_h \rightarrow \underline{\lambda}} k^* = \infty$, and hence, (13) is never satisfied if λ_h is small. In contrast, we have

$$\lim_{\lambda_h \rightarrow \infty} \frac{q_h}{(1 - q_h)e^{-\lambda_h k^*}} = 0,$$

implying that (13) must hold if λ_h is sufficiently large.

Stable job descriptions: Another distinctive feature of professional service industries is a professionalized workforce that builds on a particular knowledge base (Torres, 1991; von Nordenflycht, 2010). This in turn creates well-defined job boundaries and stable job descriptions. Due to various professional requirements and accreditation processes, workers in professional service industries are typically responsible only for a narrow and clear set of tasks, compared to workers in other industries. This implies that the nature of tasks that they are expected to carry out is very stable over time, and the production environment is relatively immune to stochastic shocks. For instance, a demand or technology shock that entirely changes the job description of a lawyer or a college professor is highly unlikely.

We can interpret r as the rate at which the project is terminated for exogenous reasons,

e.g., the arrival of a stochastic shock that makes the project completely worthless.³² Define \bar{r} such that

$$\frac{\lambda_h b - d + w}{\lambda_h + \bar{r}} = b - \frac{d}{\Delta_\lambda}.$$

Assumption 1 then implies that r must be bounded between $(0, \bar{r})$. It is clearly not profitable to commit to a deadline if r is sufficiently close to \bar{r} because $\lim_{r \rightarrow \bar{r}} k^* = \infty$. In contrast, there exists some \underline{r} such that it is profitable to commit to a deadline for $r < \underline{r}$ if

$$\frac{1 - q_h}{q_h} \left(1 - \frac{\lambda_1 (b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right) > \frac{1 - p_0}{p_0}.$$

It can be seen from this argument that a decrease in r (a more stable relationship) in general raises the value of up-or-out because the agent can be more forward-looking and tends to procrastinate more, which in turn makes the principal unable to stop at the right time.

Job similarity across ranks: Another immediate consequence of stable job descriptions is that jobs necessarily become similar across different ranks.³³ In the current setup, this aspect may be captured by y , which can be interpreted as the expected payoff of promoting a good-type agent. If entry- and senior-level jobs are similar, y tends to be larger because a success at the entry level is a reliable predictor of productivity at the senior level. This is not necessarily the case in typical firm organizations where jobs across ranks can differ to a considerable extent, as often discussed in the context of the Peter Principle.³⁴

The effect of a change in y is fairly straightforward, as it only affects the break-even belief with no impact on the agent's behavior. A large value of y implies a small value of q_h which gives the principal a stronger incentive to wait for a breakthrough. Given this, the agent also has a stronger incentive to procrastinate, hoping to achieve a breakthrough with low effort. The principal can then unambiguously benefit from committing to a deadline when y is sufficiently large, implying that the value of up-or-out is higher in industries where jobs are similar and there is a strong correlation between performances at different ranks.

³²Here, we allow the principal to terminate the project when the shock arrives; see section 6 for the case where the principal must abide by the deadline under any circumstance.

³³Ghosh and Waldman (2010) also raises job similarity as one of the distinctive characteristics of academia. Job similarity is also crucial in Kahn and Huberman's (1988) classic argument because if jobs are sufficiently different, we may use promotions to solve the double moral-hazard problem as suggested by Prendergast (1993). However, our argument here may be more applicable to academia (at least in humanities and social sciences) but less so to other professional service industries such as law where the role of a partner is more about bringing in business (or "rainmaking").

³⁴The Peter Principle states that every post tends to be occupied by an employee who is incompetent to carry out its duties, because employees are promoted through positions where they have excelled until they reach a level of incompetence (Peter and Hull, 1969). A premise of this argument is that jobs are inherently different and become progressively harder as one climbs through ranks.

6 Conclusion

This paper presents a simple stylized model of dynamic performance evaluation with particular emphasis on the role of commitment to an evaluation deadline. We consider an environment where the principal chooses whether to commit to a deadline and if so, at what time. Within this framework, we obtain a complete characterization of equilibrium and derive a necessary and sufficient condition for committing to a deadline to be optimal. The simple framework also allows us to conduct various comparative statics exercises to give insight into when it is beneficial to set an evaluation deadline at the outset.

Since our model is deliberately stylized to obtain sharp analytical characterizations, there are naturally several avenues to extend the current analysis. Before we conclude, we briefly discuss two of those possible extensions (see Appendices C and D for more formal analyses of these cases).

Uncertainty: An important cost of making commitment arises from the fact that the principal loses flexibility to adjust to future stochastic shocks. The baseline model does not capture this cost because the only feasible history at each continuation game consists only of failures up to that point; as a consequence, it is weakly optimal to commit to a deadline in the baseline model. This is apparently unsatisfactory, provided that we live in a world filled with uncertainty where there is clearly value in being flexible.

One way to cope with this possibility is to consider a setup where a permanent productivity shock may strike with some probability, which totally changes the nature of the task and subsequently makes the agent unproductive, i.e., $\lambda_h = \lambda_l = 0$.³⁵ For simplicity, suppose that (i) the shock is permanent and arrives at most once, with a Poisson arrival rate β , and (ii) the arrival of the shock is publicly observable. In this situation, the principal can retain the flexibility to adjust to a negative shock by immediately terminating the project whenever it strikes. We can then show that due to this benefit, there arises a situation where the principal strictly benefits from not committing to a deadline.

Exogenous wages: In the baseline model, we have assumed that the contractual environment is highly incomplete in that no contingent wage contracts can be written. Even in such an environment, it may still be possible to agree on the flow wage contingent on the continuation of the project. Of course, since the flow wage in this context is a pure transfer

³⁵Once the principal chooses to commit to a deadline, the principal must abide by it and is not allowed to terminate the project, unlike in the discussion of stable job descriptions in section 5.3 where we allow the principal to terminate the project whenever a stochastic shock arrives. If the principal is allowed to terminate the project at any time before the committed deadline, the commitment solution then yields a weakly higher payoff than the no-commitment solution.

payment with no incentive effect on the agent's side, there is no reason for the principal to offer any wage beyond the minimum level if she can stop at the right time. However, as we have seen, this is not always true when the principal makes no commitment to a deadline, in which case the flow wage can be used as a substitute for a deadline. This can be potentially useful if the commitment cost is prohibitively large for some reason.

To explore whether there are circumstances in which she offers any strictly positive w to this effect, suppose that the principal can deliberately choose any $w \geq 0$ at time 0.³⁶ The benefit of raising w above zero is clear if we look at the case where $w = 0$: in this case, the instantaneous payoff is always strictly positive (as in the cooperative case), no matter how unlikely the agent is to succeed; as a consequence, the principal can never terminate the project, and given this, the agent never exerts high effort. The principal can alter this structure by raising w above zero because the flow cost of employment is now positive, rendering the instantaneous payoff negative at some point. This can be profit-enhancing for the principal because with the credible threat of termination, she can induce the agent to exert high effort, which is especially beneficial when the success probability under low effort is low: formally, we can show that the optimal flow wage is strictly positive if λ_l is sufficiently small. This reasoning is similar to Buisseret (2016) in that a higher flow wage is used as a commitment device to stop at the right time and reduce future rents for the agent.³⁷

References

- Bergemann, D. and Hege, U., 1998, Venture Capital Financing, Learning, and Moral Hazard, *Journal of Banking and Finance*, 22, 703-35.
- Bergemann, D. and Hege, U., 2005, The Financing of Innovation: Learning and Stopping, *RAND Journal of Economics*, 36, 719-52.
- Bergemann, D. and Välimäki, J., 1996, Learning and Strategic Pricing, *Econometrica*, 64, 1125-49.
- Bergemann, D. and Välimäki, J., 2000, Experimentation in Markets, *Review of Economic Studies*, 67, 213-34.

³⁶In other words, we assume that the minimum wage level is zero. On the other hand, if the principal can set a negative w , she can immediately screen out the mediocre type by offering $w < 0$. This means that we focus on the situation where the minimum wage is so high that even the mediocre type can earn a positive rent by being employed.

³⁷Buisseret (2016) considers a two-period model where the principal can raise the success probability by investing in the project and shows that the principal may over-invest, so that she can credibly terminate the project upon the first failure. This can be profit-enhancing as it denies the agent the continuation payoff.

- Bergemann, D. and Välimäki, J., 2008, Bandit Problems, in Durlauf, S. and Blume, L., eds, *The New Palgrave Dictionary of Economics*, 2nd edition, Macmillan Press.
- Bolton, P. and Harris, C., 1999, Strategic Experimentation, *Econometrica*, 67, 349-74.
- Bonatti, A. and Hörner, J., 2011, Collaborating, *American Economic Review*, 101, 632-63.
- Bonatti, A. and Hörner, J., 2017, Career Concerns with Exponential Learning, *Theoretical Economics*, 12, 425-75.
- Buisseret, P., 2016, Investing in Failure: Belief Manipulation and Strategic Experimentation, mimeo.
- Campbell, A., Ederer, F. and Spinnewijn, J., 2014, Delay and Deadlines: Freeriding and Information Revelation in Partnerships, *American Economic Journal: Microeconomics*, 6, 163-204.
- Chen, C.-H., 2012, Name Your Own Price at Priceline.com: Strategic Bidding and Lockout Periods, *Review of Economic Studies*, 79, 1341-69.
- Chen, C.-H. and Ishida, J., 2015, Hierarchical Experimentation, ISER Discussion Paper No. 949.
- Damiano, E., Li, H. and Suen, W., 2012, Optimal Deadlines for Agreements, *Theoretical Economics*, 7, 357-93.
- Eriksson, T., 2016, Inter- and Intra-Firm Mobility of Workers, in Saridakis, G. and Cooper, C.L., eds, *Research Handbook on Employee Turnover*, Edward Elgar.
- Fershtman, C. and Seidmann, D.J., 1993, Deadline Effects and Inefficient Delay in Bargaining with Endogenous Commitment, *Journal of Economic Theory*, 60, 306-21.
- Fuchs, W. and Skrzypacz, A., 2013, Bargaining with Deadlines and Private Information, *American Economic Journal: Microeconomics*, 5, 219-43.
- Fudenberg, D. and Tirole, J., 1991, Perfect Bayesian Equilibrium and Sequential Equilibrium, *Journal of Economic Theory*, 53, 236-260.
- Fudenberg, D., Levine, D. and Tirole, J., 1985, Infinite Horizon Models of Bargaining with One-Sided Incomplete Information, in Roth, A., ed, *Game Theoretic Models of Bargaining*, Cambridge University Press.
- Garfagnini, U., 2011, Delegated Experimentation, mimeo.
- Georgiadis, G., Lippman, S.A. and Tang, C.S., 2014, Project Design with Limited Commitment and Teams, *RAND Journal of Economics*, 45, 598-623.

- Gerardi, D. and Maestri, L., 2012, A Principal-Agent Model of Sequential Testing, *Theoretical Economics*, 7, 425-63.
- Ghosh, S. and Waldman, M., 2010, Standard Promotion Practices Versus Up-or-Out Contracts, *RAND Journal of Economics*, 41, 301-25.
- Guo, Y., 2016, Dynamic Delegation of Experimentation, *American Economic Review*, 106, 1969-2008.
- Halac, M., Kartik, N. and Liu, Q., 2016, Optimal Contracts for Experimentation, *Review of Economic Studies*, 83, 1040-91.
- Hendricks, K., Weiss, A. and Wilson, C., 1988, The War of Attrition in Continuous Time with Complete Information, *International Economic Review*, 29, 663-80.
- Holmström, B., 1999, Managerial Incentive Problems: A Dynamic Perspective, *Review of Economic Studies*, 66, 169-82.
- Hörner, J. and Samuelson, L., 2011, Managing Strategic Buyers, *Journal of Political Economy*, 119, 379-425.
- Hörner, J. and Samuelson, L., 2013, Incentives for Experimenting Agents, *RAND Journal of Economics*, 44, 632-63.
- Hörner, J. and Samuelson, L., 2016, Dynamic Moral Hazard without Commitment, *International Journal of Game Theory*, 45, 89-136.
- Kahn, C. and Huberman, G., 1988, Two-Sided Uncertainty and “Up-or-Out” Contracts, *Journal of Labor Economics*, 6, 423-44.
- Keller G., Rady, S. and Cripps, M., 2005, Strategic Experimentation with Exponential Bandits, *Econometrica*, 73, 39-68.
- Klein, N.A. and Rady, S., 2011, Negatively Correlated Bandits, *Review of Economic Studies*, 78, 693-732.
- Lazear, E.P., and Gibbs, M., 2014. *Personnel Economics in Practice*. 3rd edition, Wiley.
- Manso, G., 2011, Motivating Innovation, *Journal of Finance*, 66, 1823-69.
- Mason, R. and Välimäki, J., 2015, Getting It Done: Dynamic Incentives to Complete a Project, *Journal of the European Economic Association*, 13, 62-97.
- O’Flaherty, B. and Siow, A., 1992, On the Job Screening, Up or Out Rules, and Firm Growth, *Canadian Journal of Economics*, 25, 346-68.

- Peter, L.J. and Hull, R., 1969, *The Peter Principle: Why Things Always Go Wrong*, Morrow.
- Prendergast, C., 1993, The Role of Promotion in Inducing Specific Human Capital Acquisition, *Quarterly Journal of Economics*, 108, 523-34.
- Rothschild, M., 1997, A Two-Armed Bandit Theory of Market Pricing, *Journal of Economic Theory*, 9, 185-202.
- Spier, K.E., 1992, The Dynamics of Pretrial Negotiation, *Review of Economic Studies*, 59, 93-108.
- Starbuck, W., 1992, Learning by Knowledge-Intensive Firms, *Journal of Management Studies*, 29, 713-40.
- Torres, D., 1991, What, If Anthing, is Professionalism? Institutions and the Problem of Change, *Research in the Sociology of Organizations: Organizations and Occupations*, 8, 43-68.
- von Bordenflycht, A., 2010, What is a Professional Service Firm? Towards a Theory and Taxonomy of Knowledge Intensive Firms, *Academy of Management Review*, 35, 155-74.
- Waldman, M., 1990, Up-or-Out Contracts: A Signaling Perspective, *Journal of Labor Economics*, 8, 230-50.
- Winch, G. and Schneider, E., 1993, Managing the Knowledge-Based Organization: The Case of Architectural Practice, *Journal of Management Studies*, 30, 923-37.

Appendix A: Proofs

Proof of Proposition 1: It is intuitively clear that the value function is strictly increasing in k . Formally, define

$$rU_a(k) = \left((\lambda_a b - d_a + w) - \lambda_a U_a(k) - \dot{U}_a(k) \right) \quad (14)$$

as the value function when the effort level is fixed at $a = l, h$. Solving this differential equation and imposing the boundary condition $U(0) = 0$, we derive

$$U_h(k) = \frac{\lambda_h b - d + w}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k}).$$

If $\Delta_\lambda b > d$, it is optimal for the agent to exert high effort as long as $U_h(k)$ is sufficiently small. Moreover, since $U_h(k)$ is strictly increasing in k , the agent would never switch back

to low effort, once $U(k)$ reaches this critical point. Then, k^* is obtained as a solution to $U_h(k^*) = b - \frac{d}{\Delta_\lambda}$, which can be written as

$$\frac{\lambda_h b - d + w}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k^*}) = b - \frac{d}{\Delta_\lambda}. \quad (15)$$

Note that the agent always exerts high effort if

$$b - \frac{d}{\Delta_\lambda} \geq \frac{\lambda_h b - d + w}{\lambda_h + r} \Leftrightarrow \frac{rb - w}{\lambda_l + r} \geq \frac{d}{\Delta_\lambda},$$

while he never does so if $\frac{d}{\Delta_\lambda} \geq b$. For in-between cases, solving (15) yields

$$k^* = -\frac{1}{\lambda_h + r} \ln \left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right),$$

so that the agent exerts high effort when the remaining time is less than or equal to k^* . For $k > k^*$, the agent exerts low effort because $U(k) \geq U_h(k) > b - \frac{d}{\Delta_\lambda}$. ■

Proof of Proposition 2: The agent always exerts high effort in the noncooperative case if (2) fails to hold. An analogous condition for the cooperative case is given by

$$\frac{\lambda_h(y + b) - d}{\lambda_h + r} > y + b - \frac{d}{\Delta_\lambda} \Leftrightarrow \frac{d}{\Delta_\lambda} > \frac{r}{\lambda_l + r}(y + b).$$

Applying the same argument as in Proposition 1, we obtain

$$\tilde{k}^* = -\frac{1}{\lambda_h + r} \ln \left(1 - \frac{(\lambda_h + r)(y + b - \frac{d}{\Delta_\lambda})}{\lambda_h(y + b) - d} \right).$$

To see that the project is never terminated, we first show that if it is ever optimal to set a deadline in some finite time, the only possibility is to set it at time \tilde{k}^* , so that the agent exerts high effort from the beginning. To see this, the principal extends the deadline beyond time \tilde{k}^* if and only if

$$\lambda_l(y + b) > \frac{\lambda_l + r}{\lambda_h + r} (\lambda_h(y + b) - d) (1 - e^{-(\lambda_h + r)\tilde{k}^*}). \quad (16)$$

Note that this condition is independent of the deadline. As such, if this condition holds, it is then optimal to continue the project indefinitely until the project succeeds. If it fails to hold, on the other hand, it is clearly optimal to set a deadline at time \tilde{k}^* .

Since, by definition,

$$y + b - \frac{d}{\Delta_\lambda} = \tilde{U}(\tilde{k}^*) = \frac{(\lambda_h(y + b) - d)}{\lambda_h + r} (1 - e^{-(\lambda_h + r)\tilde{k}^*}).$$

we can modify (16) as

$$\frac{\lambda_l(y+b)}{\lambda_l+r} > y+b - \frac{d}{\Delta_\lambda}.$$

The left-hand side is the expected payoff of exerting low effort indefinitely while the right-hand side is that of exerting high effort from time 0 to time \tilde{k}^* which must equal the static gain.

This condition must hold for any finite \tilde{k}^* when (6) is satisfied, meaning that it is strictly better to let the project continue indefinitely. If (6) fails to hold, on the other hand, $\tilde{k}^* \rightarrow \infty$, and again it is better to let the project continue indefinitely. This shows that no finite deadline can improve the social surplus in the cooperative case. \blacksquare

Proof of Proposition 3: First, if $\phi(q_h) \geq p_0 > q_h$ (the latter inequality holds by Assumption 2), there exists a unique deadline $\tau^C \in (0, k^*]$ that satisfies

$$\frac{p_0 e^{-\lambda_h \tau^C}}{1 - p_0 + p_0 e^{-\lambda_h \tau^C}} = q_h.$$

There is clearly no reason to wait beyond this point because the instantaneous payoff is strictly negative for any effort choice. It is also not optimal to stop before τ^C because the instantaneous payoff is still strictly positive. The optimal termination date is then τ^C and it is unique in this range.

Second, suppose that $p_0 > \phi(q_h)$, in which case the first-order condition is given by $M(\tau^C, p_0) = 0$. If $M(k^*, p_0) \leq 0$, the expected payoff is strictly decreasing in τ^C for $\tau^C > k^*$, and the optimal termination date is $\tau^C = k^*$. Alternatively, define $\hat{p} \in (0, 1)$ such that $M(k^*, \hat{p}) = 0$ if (10) holds and $\hat{p} = 1$ otherwise. Then, the optimal termination date is k^* for $p_0 \leq \hat{p}$. If $p_0 > \hat{p}$, on the other hand, there must exist a unique $\hat{\tau}(p_0)$ such that $M(\hat{\tau}(p_0), p_0) = 0$. Since the instantaneous payoff is strictly positive for $\hat{\tau}(p_0) > \tau^C$ and negative for $\tau^C > \hat{\tau}(p_0)$, the optimal termination date in this case is $\hat{\tau}(p_0)$.

Finally, to show that $\hat{p} > q_l$ if $1 > q_l$, there are two cases. If (10) holds (but $1 > q_l$), then $\hat{p} = 1 > q_l$. If (10) fails to hold, on the other hand, $M(k^*, p_0)$ is increasing in p_0 for any k^* . Given that $1 > q_l$, it suffices to show that $M(k^*, q_l) < 0$, i.e.,

$$(1 - q_l)w e^{-rk^*} > q_l \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h (1 - e^{-(\lambda_h + r)k^*}) \right).$$

This can be written as

$$-e^{rk^*} \left(q_l \frac{\lambda_l + r}{\lambda_h + r} \pi_h e^{-\lambda_h k^*} - (1 - q_l)w \right) > q_l \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h \right).$$

which is further reduced to

$$-e^{rk^*} (q_l \pi_l e^{-\lambda_h k^*} - (1 - q_l)w) > q_l (1 - e^{-(\lambda_h + r)k^*}) \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h \right).$$

Note that the left-hand side is positive by definition while the right-hand side is negative, so that this condition holds for any k^* . From the preceding argument, we can see that $\hat{p} = 1$ if and only if (10) holds. It is also evident that $\tau^C(p_0)$ is weakly increasing in p_0 . ■

Proof of Proposition 4: We show the existence and uniqueness of the equilibrium by construction. We start with a candidate strategy $T^1 = \{\tau^1\}$ and the corresponding best response such that

$$a_t = \begin{cases} l & \text{for } t \in [0, \max\{\tau^1 - k^*, 0\}), \\ h & \text{for } t \in [\max\{\tau^1 - k^*, 0\}, \tau^1]. \end{cases} \quad (17)$$

This pair of strategies constitutes an equilibrium if $V^*(K(p, q_h), p) > 0$ for all $p \in (q_h, p_0]$. Alternatively, we check for what interval of p_0 this condition is satisfied. To this end, the following result is useful.

Lemma 2 *Let $q_l > q$. Then, $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \in (q, 1)$ if and only if*

$$\lim_{p \uparrow q_l} V^*(K(p, q), p) > 0.$$

Proof: Since the necessity is evident by definition, we only prove the sufficiency part. Since the case with $q_l = 1$ can be proved analogously (by taking the limit), we suppose for now that $1 > q_l > q$.

First, if $p_0 \geq q_l$, $V^*(K(q_l, q), q_l) > 0$ directly implies $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \geq q_l$ because the instantaneous payoff is strictly positive for any effort choice when the belief is above q_l . Note also that $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \in (q, \phi(q_h)]$ since the high type always exerts high effort in this case. This implies that $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \in (q, 1)$ if $\phi(q) \geq q_l$. In what follows, therefore, we assume $q_l > \phi(q)$.

Given this, it suffices to show that

$$V^*(K(q_l, q), q_l) > 0 \Rightarrow V^*(K(p_0, q), p_0) > 0 \text{ for all } p_0 \in (\phi(q), q_l).$$

Suppose on the contrary that $V^*(K(q_l, q), q_l) > 0$ but $V^*(K(p, q), p) \leq 0$ for some $p \in (\phi(q), q_l)$. If this is the case, there must exist some $z \in (\phi(q), q_l)$ such that $V^*(K(z, q), z) = 0$. Since

$$\begin{aligned} V^*(K(p, q), p) &= \frac{p(\lambda_l y - w)}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(K(p, q) - K(p', q))}) - \frac{(1 - p)w}{r} (1 - e^{-r(K(p, q) - K(p', q))}) \\ &\quad + (1 - p + p e^{-\lambda_l(K(p, q) - K(p', q))}) e^{-r(K(p, q) - K(p', q))} V^*(K(p', q), p'), \end{aligned}$$

for any $p > p' > \phi(q)$, due to the recursive structure of the expected payoff, we have

$$\begin{aligned} V^*(K(q_l, q), q_l) &= \frac{q_l(\lambda_l y - w)}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\Delta K}) - \frac{(1 - q_l)w}{r} (1 - e^{-r\Delta K}) \\ &\quad + (1 - q_l + q_l e^{-\lambda_l \Delta K}) e^{-r\Delta K} V^*(K(z, q), z) \\ &= \frac{q_l(\lambda_l y - w)}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\Delta K}) - \frac{(1 - q_l)w}{r} (1 - e^{-r\Delta K}), \end{aligned}$$

where $\Delta K := K(q_l, q) - K(z, q)$. This is strictly negative by definition, and therefore a contradiction. \blacksquare

The lemma shows that we only need to check the continuation payoff at q_l to see whether a given strategy is sequentially rational. An immediate corollary of the lemma is that $V^*(K(p_0, q_h), p_0) > 0$ for all p_0 if $\phi(q_h) \geq q_l$. In this case, $m(\Theta) = 1$, $P^1 = q_h$ and $T = \{\tau^1\}$ constitutes an equilibrium for any p_0 .

We now suppose that $q_l > \phi(q_h)$ and check if the candidate strategy can satisfy the equilibrium conditions. From Lemma 2, this is the case if and only if $V^*(K(q_l, q_h), q_l) > 0$, i.e.,

$$\frac{q_l \pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(K - k^*)}) + \frac{q_l \pi_h}{\lambda_h + r} e^{-(\lambda_l + r)(K - k^*)} (1 - e^{-(\lambda_h + r)k^*}) > \frac{(1 - q_l)w}{r} (1 - e^{-rK}). \quad (18)$$

where $K = K(q_l, q_h)$. If (18) holds, the expected payoff is always positive for any p_0 , so we can apply the same argument as in Case 2 to show the existence and uniqueness of the equilibrium.

The situation becomes more complicated when (18) fails to hold, in which case the principal has an incentive to deviate and stop at some point, and the candidate strategy no longer constitutes an equilibrium. In this case, there instead exists a critical belief that satisfies (18) with equality, because $V^*(K(p, q_h), p)$ is continuous in p with $V^*(K(\phi(q_h), q_h), \phi(q_h)) > 0 > V^*(K(q_l, q_h), q_l)$. We now redefine $P^1 = \min\{q : V^*(K(q, q_h), q) = 0\}$ and $P^2 = q_h$. Given the critical belief, we can then consider another candidate strategy $T^2 = \{\tau^1, \tau^2\}$ for $p_0 > P^1$ where $p_{\tau^1} = q^1$, and the best response is given by

$$a_t = \begin{cases} l & \text{for } t \in [0, \max\{\tau^1 - k^*, 0\}), \\ h & \text{for } t \in [\max\{\tau^1 - k^*, 0\}, \tau^1), \\ l & \text{for } t \in (\tau^1, \tau^2 - k^*), \\ h & \text{for } t \in [\tau^2 - k^*, \tau^2]. \end{cases}$$

Note that $q^1 > \phi(q^2) > q^2 = q_h$ or $\tau^2 > \tau^1 + k^*$.

The game is now divided into two segments. First, we can show that the equilibrium is unique after time τ^1 (as done in Case 2). Before time τ^1 , we can apply the same procedure to

the truncated interval $[0, \tau^1]$ to see whether T^2 can be an equilibrium strategy. More precisely, we need to check if $V^*(K(p, P^1), P^1) > 0$ for all $p \in [p_0, \phi(P^1)]$. Applying Lemma 2, T^2 is an equilibrium strategy for all $p_0 \in (P^1, 1)$ if $V(K(q_l, P^1), q_l) > 0$. For $p_0 \in (P^1, 1)$, the principal's equilibrium strategy consists of two termination dates. Given that the continuation equilibrium after time τ^1 is unique, we can establish the uniqueness by applying the same argument.

If $V(K(q_l, P^1), q_l) \leq 0$, on the other hand, we can find yet another critical belief in $(\phi(P^1), q_l]$. We again redefine P^1 analogously and repeat the same process until we find P^1 such that $V(K(q_l, P^1), q_l) > 0$. In the end, this process would give us $m(\Theta)$ intervals $\{(P^j, P^{j-1})\}_{j=1}^m$ where the principal's equilibrium strategy consists of $n = m(\Theta) - j + 1$ termination dates if p_0 falls into (P^j, P^{j-1}) . Note also that since $V^*(K(P^1, P^2), P^1) = 0$ by definition, it must be that $q_l > P^1$ if $m(\Theta) > 1$. ■

Proof of Proposition 5: (i) For $p_0 \in (q_h, \phi(q_h))$, $\tau^C(p_0) = \tau^{NC}(p_0) = K(p_0, q_h)$. Since the agent's strategy depends only on the remaining time, the two cases yield the same expected payoff for the principal.

(ii) To show this, we first examine how V^C and V^{NC} vary with respect to p_0 for $p_0 \in (\phi(q_h), 1)$. It is evident that V^C must monotonically increase with the initial prior. To see this, note that

$$\frac{dV^C}{dp_0} = \frac{\partial V^*}{\partial p_0} + \frac{\partial V^*}{\partial \tau^C} \frac{d\tau^C}{dp_0}.$$

The last term is the indirect effect though an increase in τ^C which is invariably zero because either $\frac{\partial V^*}{\partial \tau^C} = 0$ or $\frac{d\tau^C}{dp_0} = 0$ must hold by the optimality condition. Since the direct effect $\frac{\partial V^*}{\partial p_0}$ is unambiguously positive, V^C is strictly increasing in p_0 . More precisely, we obtain

$$\begin{aligned} \frac{dV^C}{dp_0} &= \frac{\pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k^*}) + \frac{w}{r} (1 - e^{-rk^*}) \\ &= p_0 \pi_h - (1 - p_0)w - (p_0 \lambda_h + r)V^C(p_0) - (p_0 e^{-\lambda_h k^*} \pi_h - (1 - p_0)w)e^{-rk^*}, \end{aligned} \quad (19)$$

for $p_0 \in (\phi(q_h), \hat{p})$, and

$$\begin{aligned} \frac{dV^C}{dp_0} &= \frac{\pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(\tau^C - k^*)}) + \frac{e^{-(\lambda_l + r)(\tau^C - k^*)} \pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k^*}) + \frac{w}{r} (1 - e^{-r\tau^C}) \\ &= p_0 \pi_l \frac{\lambda_h + r}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(\tau^C - k^*)}) + p_0 \pi_h e^{-(\lambda_l + r)(\tau^C - k^*)} - (1 - p_0)w \\ &\quad - (p_0 \lambda_h + r)V^C(p_0) - (p_0 \pi_h e^{-\lambda_l(\tau^C - k^*) - \lambda_h k^*} - (1 - p_0)w)e^{-r\tau^C}, \end{aligned} \quad (20)$$

for $p_0 \in (\hat{p}, 1)$.

In contrast, V^{NC} changes with p_0 in a non-monotonic way if $m > 1$. Consider $p_0 \in (\phi(P^j), \phi(P^{j-1}))$ for $j = 2, 3, \dots, m$. Since $V^{NC}(p_0) = V^*(K(p_0, q), p_0)$ where q is the closest termination belief, it follows that

$$\frac{dV^{NC}}{dp_0} = \frac{\partial V^*}{\partial p_0} + \frac{\partial V^*}{\partial k} \frac{\partial K}{\partial p_0} = \frac{\partial V^*}{\partial p_0} + \frac{1}{\lambda_l p_0 (1 - p_0)} \frac{\partial V^*}{\partial k}.$$

First, for $p_0 \in (\phi(P^j), P^{j-1})$, we obtain

$$\begin{aligned} \frac{dV^{NC}}{dp_0} &= \frac{\pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(K - k^*)}) + \frac{\pi_h e^{-(\lambda_l + r)(K - k^*)}}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k^*}) + \frac{w}{r} (1 - e^{-rK}) \\ &\quad + \frac{e^{-(\lambda_l + r)(K - k^*)}}{\lambda_l (1 - p_0)} \left(\frac{\pi_l}{\lambda_l + r} - \frac{\pi_h}{\lambda_h + r} (1 - e^{-(\lambda_l + r)k^*}) \right) - \frac{w e^{-rK}}{\lambda_l p_0} \\ &= p_0 \pi_l - (1 - p_0) w - (p_0 \lambda_l + r) V^{NC}(p_0), \end{aligned} \quad (21)$$

which is negative because $q_l > P^1 > p_0$. For $p_0 \in (P^{j-1}, \phi(P^{j-1}))$, we obtain

$$\begin{aligned} \frac{dV^{NC}}{dp_0} &= \frac{\pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)K}) + \frac{w}{r} (1 - e^{-rK}) + \frac{e^{-rK}}{\lambda_h p_0 (1 - p_0)} (p_0 \pi_h e^{-\lambda_h K} - (1 - p_0) w) \\ &= p_0 \pi_h - (1 - p_0) w - (p_0 \lambda_h + r) V^{NC}(p_0), \end{aligned} \quad (22)$$

which is positive for any $p_0 > q_h$.

We are now ready to see how ΔV changes with respect to p_0 . First, it is clear that ΔV is strictly increasing in p_0 for $p_0 \in (\phi(P^j), P^{j-1})$ because V^C is increasing while V^{NC} is decreasing. In contrast, for $p_0 \in (P^{j-1}, \phi(P^{j-1}))$, both V^C and V^{NC} are increasing in p_0 . For $\hat{p} > p_0$, it follows from (19) and (22) that

$$\frac{d\Delta V}{dp_0} < 0 \Leftrightarrow (p_0 \lambda_h + r)(V^C(p_0) - V^{NC}(p_0)) > -(p_0 e^{-\lambda_k^*} - (1 - p_0) w) e^{-r k^*}.$$

This condition always holds because $p_0 e^{-\lambda_h k^*} > (1 - p_0) w$ and $V^C(p_0) \geq V^{NC}(p_0)$ by definition. Similarly, for $p_0 > \hat{p}$, comparing (20) and (21) yields

$$\begin{aligned} &p_0 \left(\pi_h - \frac{\lambda_h + r}{\lambda_l + r} \pi_l \right) (1 - e^{-(\lambda_l + r)(\tau^C - k^*)}) + (p_0 \lambda_h + r)(V^C(p_0) - V^{NC}(p_0)) \\ &> -(p_0 e^{-\lambda_l(\tau^C - k^*) - \lambda_k^*} - (1 - p_0) w) e^{-r \tau^C}, \end{aligned}$$

which again always holds.

(iii) If $\hat{p} \geq \phi(P^1)$, then $\tau^C(\phi(P^1)) = k^*$. Since $\tau^{NC}(\phi(P^1)) = k^*$ by definition, $\Delta V(\phi(P^1)) = 0$. To see the necessity, note that $\Delta V(p_0) = 0$ implies $\tau^C(p_0) = \tau^{NC}(p_0)$. Since $\tau^{NC}(\phi(P^1)) = k^*$, we must have $\tau^C(\phi(P^1)) = k^*$ which holds if and only if $\hat{p} \geq \phi(P^1)$.

(iv) Suppose first that $1 > \hat{p}$, in which case τ^C and τ^{NC} solve

$$e^{-\lambda\tau^C(p_0)} = \frac{1-p_0}{p_0} \left(\frac{(\lambda_h+r)we^{-(\lambda+r)k^*}}{(\lambda_h+r)\pi_l - (\lambda_l+r)\pi_h(1-e^{-(\lambda_h+r)k^*})} \right) =: \frac{1-p_0}{p_0} \Phi^C,$$

$$e^{-\lambda_l\tau^{NC}(p_0)} = \frac{1-p_0}{p_0} \left(\frac{q^1}{(1-q^1)e^{-\Delta\lambda k^*}} \right) =: \frac{1-p_0}{p_0} \Phi^{NC},$$

respectively, for $p_0 > \hat{p}$. Note that both Φ^C and Φ^{NC} are independent of p_0 . From this, since $\lim_{p_0 \rightarrow 1} \tau^C(p_0) = \infty$ and $\lim_{p_0 \rightarrow 1} \tau^{NC}(p_0) = \infty$, we obtain

$$\lim_{p_0 \rightarrow 1} V^C(p_0) = \lim_{p_0 \rightarrow 1} V^{NC}(p_0) = \frac{p_0\pi_l}{\lambda_l+r} - \frac{(1-p_0)w}{r}.$$

If $\hat{p} = 1$, on the other hand, $\tau^C(p_0) = k^*$ for all p_0 , but as we have seen, $\lim_{p_0 \rightarrow 1} \tau^{NC}(p_0) = \infty$.

■

Appendix B: the value of commitment in $(\phi(P^1), 1)$

As we have seen in Proposition 5, the value of commitment moves with some regularities for $p_0 \in (q_h, \phi(P^1))$. The exception is the last segment $(\phi(P^1), 1)$. The reason is that for this (and only this) segment, we could have $\phi(P^1) > \hat{p} > q_l$ (see Propositions 3 and 4), so that the instantaneous payoff under low effort can be positive and $\tau^C(p_0) > k^*$ for $p_0 \in (\phi(P^1), 1)$. There are three cases we need to consider, depending on the value of \hat{p} .

Case 1: $\phi(P^1) > \hat{p}$

In this case, $\tau^C(\phi(P^1)) > \tau^{NC}(\phi(P^1)) = k^*$ and hence $\Delta V(\phi(P^1)) > 0$ (as we show in (iii) of Proposition 4). For $p_0 > \phi(P^1)$, $\tau^C(p_0)$ and $\tau^{NC}(p_0)$ increase proportionally (see the proof of Proposition 4), with both diverging to infinity as $p_0 \rightarrow 1$. As we show in (iv) of Proposition 4, $\Delta V(p_0)$ converges to zero as $p_0 \rightarrow 1$.

Case 2: $\hat{p} > \phi(P^1)$

In this case, $\tau^C(\phi(P^1)) = \tau^{NC}(\phi(P^1)) = k^*$ and hence $\Delta V(\phi(P^1)) = 0$. For $p_0 \in (\phi(P^1), \hat{p})$, $\tau^{NC}(p_0)$ continues to go up while $\tau^C(p_0)$ stays at k^* ; as a consequence, $\Delta V(p_0)$ gradually increases in this range. Once p_0 goes above \hat{p} , $\tau^C(p_0)$ also gradually increase, but again proportionally to $\tau^{NC}(p_0)$. Since $\tau^{NC}(p_0) > \tau^C(p_0)$ for $p_0 \in (\hat{p}, 1)$ in this case, $\Delta V(p_0)$ gradually decreases and converges to zero as $p_0 \rightarrow 1$.³⁸

³⁸If $\hat{p} = \phi(P^1)$, then $\tau^C(p_0) = \tau^{NC}(p_0)$ and hence $\Delta V(p_0) = 0$ for $p_0 \in (\hat{p}, 1)$. We ignore this non-generic case.

Case 3: $\hat{p} = 1$

This is the most straightforward case. Given that $\hat{p} = 1$, $\tau^C(p_0) = k^*$ for $p_0 \in (\hat{p}, 1)$ and hence $\tau^C(p_0)$ and $\tau^{NC}(p_0)$ diverge away from each other in this range. As a consequence, $\Delta V(p_0)$ increases as in all the preceding intervals.

Appendix C: The case with uncertainty

We analyze an extended version of the baseline model where a permanent productivity shock may strike with some probability. For simplicity, we assume that a shock arrives with a Poisson arrival rate β ; once it arrives, the agent becomes totally unproductive thereafter, i.e., $\lambda_h = \lambda_l = 0$. We also assume that the arrival of a shock is publicly observable. Other more complicated cases, e.g., those with transitory shocks, can in principle be analyzed in an analogous manner.

The agent: The agent's problem requires a slight modification. The augmented Bellman equation is obtained as

$$rU^j(k) = \max_{a \in \{l, h\}} \left((b\lambda_a - d_a + w) - (\lambda_a + \beta)U(k) + \beta\tilde{U}^j(k) - \dot{U}(k) \right), \quad j = C, NC$$

where \tilde{U}^j denotes the value function after the shock has arrived. Since the principal's reaction to the shock differs, \tilde{U}^j , and hence $U^j(k)$, depend on whether or not the principal makes commitment. As in the baseline model, the incentive compatibility constraint (before the shock strikes) can be written as

$$\Delta_\lambda(b - U^j(k)) \geq d,$$

although the value function $U^j(k)$ now differs. As such, the agent exerts high effort only when the project is sufficiently close to termination. We let k^C and k^{NC} denote the threshold under commitment and no commitment, respectively.

Proposition 7 *For any Θ and β , $k^{NC} > k^C$. Both k^{NC} and k^C are strictly increasing in β if $\frac{b\lambda_h - d + w}{\lambda_h + \beta + r} > b - \frac{d}{\Delta_\lambda} > 0$.*

Proof: We first consider the case where the principal commits to a deadline. Note that the principal cannot terminate the project even if the shock arrives before the deadline. Since the agent has no incentive to exert high effort, we have $\tilde{U}^C(k) = \frac{w(1 - e^{-rk})}{r}$. Let $U_a^C(k)$ denote the value function under commitment when the effort level is fixed at a . We then obtain

$$U_h^C(k) = \frac{b\lambda_h - d + w}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}) + \frac{\beta w}{r} \left(\frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} - \frac{e^{-rk}(1 - e^{-(\lambda_h + \beta)k})}{\lambda_h + \beta} \right).$$

The threshold under commitment is a solution to $U_h^C(k^C) = b - \frac{d}{\Delta_\lambda}$.

The problem is much more straightforward when the principal makes no commitment. In this case, the optimal choice for the principal is to terminate the project as soon as the shock arrives, which implies $\tilde{U}^{NC}(k) = 0$ for all k . As above, define $U_a^{NC}(k)$ as the value function under no commitment with fixed effort a , which is given by

$$U_h^{NC}(k) = \frac{b\lambda_h - d + w}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}).$$

If $\frac{b\lambda_h - d + w}{\lambda_h + \beta + r} > b - \frac{d}{\Delta_\lambda} > 0$, there exists an interior threshold k^{NC} given by

$$k^{NC} = -\frac{1}{\lambda_h + \beta + r} \ln \left(1 - \frac{(\lambda_h + \beta + r)(b - \frac{d}{\Delta_\lambda})}{b\lambda_h - d + w} \right).$$

It is evident that $k^{NC} > k^C$ because the expected payoff must be higher under commitment for a given horizon. Formally, we need to show that

$$\frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} > \frac{e^{-rk}(1 - e^{-(\lambda_h + \beta)k})}{\lambda_h + \beta},$$

which can also be written as

$$\int_0^k e^{-(\lambda_h + \beta + r)s} ds > \int_0^k e^{-(\lambda_h + \beta)s - rk} ds.$$

This condition holds because

$$e^{-(\lambda_h + \beta + r)s} > e^{-(\lambda_h + \beta)s - rk} \Leftrightarrow e^{-rs} > e^{-rk},$$

for any $s < k$.

Finally, note that $k^C, k^{NC} \in (0, \infty)$ when $\frac{b\lambda_h - d + w}{\lambda_h + \beta + r} > b - \frac{d}{\Delta_\lambda} > 0$. We can then show that k^j , $j = C, NC$, is increasing in β if $U_a^j(k)$ is decreasing in β for $k \leq k^j$. Straightforward computation shows that $U_h^{NC}(k)$ is strictly decreasing in β . In contrast, it is a little more involved to show the same for $U_h^C(k)$. To see this, with some computation, we obtain

$$\begin{aligned} U_h^C(k) &= \frac{b\lambda_h - d}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}) - \frac{\lambda_h w}{r} \left(\frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} - \frac{e^{-rk}(1 - e^{-(\lambda_h + \beta)k})}{\lambda_h + \beta} \right) \\ &\quad + \frac{w(1 - e^{-rk})}{r} \\ &= \frac{b\lambda_h - d - \lambda_h \frac{w(1 - e^{-rk})}{r}}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}) + \frac{\lambda_h w e^{-rk}}{r} \left(\frac{1 - e^{-(\lambda_h + \beta)k}}{\lambda_h + \beta} - \frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} \right) \\ &\quad + \frac{w(1 - e^{-rk})}{r}. \end{aligned}$$

First, it is easy to verify that $\frac{1-e^{-\alpha k}}{\alpha}$ is strictly decreasing for any $\alpha > 0$. Further, observe that the second term is strictly decreasing in β . These two facts imply that it suffices to show that

$$b\lambda_h - d > \lambda_h \frac{w(1 - e^{-rk})}{r},$$

for $k \leq k^C$. To show this, since $b\lambda_h - d > b\lambda_h - \frac{\lambda_h}{\Delta_\lambda}d \geq \lambda_h U_h^C(k)$ for $k \leq k^C$, we have

$$U_h^C(k) > \frac{\lambda_h(U_h^C(k) - \frac{w(1-e^{-rk})}{r})}{\lambda_h + \beta + r}(1 - e^{-(\lambda_h + \beta + r)k}) + \frac{\lambda_h w e^{-rk}}{r} \left(\frac{1 - e^{-(\lambda_h + \beta)k}}{\lambda_h + \beta} - \frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} \right) + \frac{w(1 - e^{-rk})}{r},$$

for $k \leq k^C$. This is further reduced to

$$\left(1 - \frac{\lambda_h(1 - e^{-(\lambda_h + \beta + r)k})}{\lambda_h + \beta + r} \right) \left(U_h^C(k) - \frac{w(1 - e^{-rk})}{r} \right) > \frac{\lambda_h w e^{-rk}}{r} \left(\frac{1 - e^{-(\lambda_h + \beta)k}}{\lambda_h + \beta} - \frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} \right),$$

which gives $U_h^C(k) > \frac{w(1 - e^{-rk})}{r}$. ■

It is interesting to note that the possibility of a negative shock generally lowers the expected future payoff and hence induces the agent to exert high effort for longer, either with or without commitment ($k^{NC} > k^C > k^*$). First, with commitment, the best the agent can do after the shock strikes is to exert low effort and earn the flow payoff w , which is the lowest possible payoff as long as the project survives. The situation is even worse without commitment, however, because the principal can immediately terminate the project, thereby leaving no rent for the agent. The extent of the loss is therefore larger when the principal makes no commitment, which amounts to the fact that $k^{NC} > k^C$ and in turn favors the no-commitment solution.

The principal: We start with the case where the principal commits to a deadline at the outset. Let $V^C(k, p_0)$ denote the principal's expected payoff at time 0 under commitment, which can be written as

$$\begin{aligned} V^C(k, p_0) &= \int_0^{\nu^C} (\lambda_l p_s y - w) e^{-\int_0^s (\lambda_l p_u + \beta + r) du} ds \\ &\quad + e^{-\int_0^{\nu^C} (\lambda_l p_u + \beta + r) du} \int_{\nu^C}^k (\lambda_h p_s y - w) e^{-\int_{\nu^C}^s (\lambda_h p_u + \beta + r) du} ds \\ &\quad - \frac{\beta}{r} \int_0^{\nu^C} w(1 - e^{-r(k-s)}) e^{-\int_0^s (\lambda_l p_u + \beta + r) du} ds \\ &\quad - \frac{\beta}{r} e^{-\int_0^{\nu^C} (\lambda_l p_u + \beta + r) du} \int_{\nu^C}^k w(1 - e^{-r(k-s)}) e^{-\int_{\nu^C}^s (\lambda_h p_u + \beta + r) du} ds. \end{aligned} \quad (23)$$

where $\nu^C := \max\{\tau^C - k^C, 0\}$. The difference from the previous case is that a negative shock may now strike with some probability, in which case the principal earns zero from that point on. The commitment solution is obtained as

$$\tau^C = \operatorname{argmax}_\tau V_0^C(\tau).$$

As can be seen from (23), the marginal benefit of extending the deadline decreases with β , which prompts the principal to terminate earlier than in the case with no uncertainty.

In contrast, the problem is roughly the same as in the baseline model when the principal makes no commitment because the principal only pays attention to the instantaneous payoff. As in the baseline model, the principal stops when the instantaneous payoff becomes sufficiently low. To be more precise, define

$$\phi^{NC}(p) := \frac{p}{p + (1-p)e^{-\lambda_h k^{NC}}}.$$

Given that no shock has arrived, if $\phi^{NC}(q_h) > q_l$, the principal stops when the belief reaches q_h . The only difference is that when the shock strikes, the instantaneous payoff falls to $-w$, which prompts her to stop immediately. The principal's expected payoff is given by

$$\begin{aligned} V^{NC}(k, p) = & \int_0^{\nu^{NC}} (\lambda_l p_s y - w) e^{-\int_0^s (\lambda_l p_u + \beta + r) du} ds \\ & + e^{-\int_0^{\nu^{NC}} (\lambda_l p_u + \beta + r) du} \int_{\nu^{NC}(q)}^k (\lambda_h p_s y - w) e^{-\int_{\nu^{NC}(q)}^s (\lambda_h p_u + \beta + r) du} ds, \end{aligned} \quad (24)$$

where ν^{NC} is defined analogously. Note that for the same time horizon, $V^{NC}(k, p_0) > V_0^C(k, p_0)$ for any p_0 .

The optimal incentive scheme: In the baseline model with no uncertainty, any allocation which can be realized without commitment can be realized with commitment as well. This is no longer the case with uncertainty because the principal cannot foresee if and when the shock strikes, in which case she needs to immediately terminate the project. We can then make the following statement.

Proposition 8 *The principal strictly benefits from not committing to a deadline for any $\beta > 0$ if $\phi^{NC}(q_h) \geq p_0$.*

Proof: If $\phi^{NC}(q_h) \geq p_0$, the agent exerts high effort from the beginning under no commitment. The principal's payoff is maximized, and the principal strictly benefits from making no commitment in this case. To see this, observe that $k^{NC} > k^C$ implies $\phi^{NC}(q_h) > \phi^C(q_h)$. If $\phi^{NC}(q_h) \geq p_0 > \phi^C(q_h)$, the agent starts off with low effort, and the principal is forced to

set a deadline before the belief reaches q_h . The profit is clearly lower than $V^{NC}(\tau^{NC}, p_0)$. On the other hand, if $\phi^{NC}(q_h) > \phi^C(q_h) \geq p_0$, the agent exerts high effort from the beginning even under commitment. The profit is still lower, however, because the principal must incur a flow loss when the shock strikes whereas she can immediately terminate the project if she makes no commitment. ■

The presence of uncertainty lowers the value of commitment through two channels. First, by making no commitment, the principal can retain the flexibility to adjust to the negative shock by immediately terminating the project whenever it strikes. Aside from this conventional effect of flexibility, there also arises another force that favors the no-commitment solution: in the face of uncertainty, the agent is better motivated by the constant threat of project termination, and thus, he starts exerting high effort earlier. Since $\phi^{NC}(q_h) > \phi(q_h)$ for any $\beta > 0$ by Proposition 7, we could have $\phi^{NC}(q_h) \geq p_0 > \phi(q_h)$, in which case the value of commitment is strictly negative with uncertainty whereas it is strictly positive without.

7 Appendix D: The case with endogenous wages

We consider a setup where the principal can deliberately choose $w \geq 0$ at time 0. In general, an increase in w above zero yields two opposing effects. On one hand, it raises the critical value q_h , which may give the principal the incentive to stop earlier than otherwise. On the other hand, it also raises the expected future payoff which in turn lowers k^* and gives the agent the incentive to procrastinate even more. A necessary condition for $w > 0$ is that the first effect dominates the second, or alternatively that $\tau^{NC} - k^*$ decreases with w .³⁹ We can then make the following statement.

Proposition 9 *It is optimal to set $w > 0$ if λ_l is sufficiently small.*

Proof: As $w \rightarrow 0$, we have $q_h \rightarrow 0$ and $q_l \rightarrow 0$, which implies $\tau^{NC}(p_0) \rightarrow \infty$ for any p_0 . The expected payoff then converges to

$$\lim_{w \rightarrow 0} V^*(\tau^{NC}, p_0) = \frac{p_0 \lambda_l y}{\lambda_l + r}, \quad (25)$$

for any p_0 . Define $\mu(w)$ such that $\phi(\frac{w}{\lambda_h y}) = \mu(w)$, i.e.,

$$\frac{\mu(w)^{-\lambda_h k^*}}{1 - \mu(w) + \mu(w)^{-\lambda_h k^*}} = q_h = \frac{w}{\lambda_h y}.$$

³⁹Note that $\tau^{NC} - k^*$ is the interval during which the agent exerts low effort. If it is longer with a smaller k^* , the principal's expected payoff is necessarily lower.

Note that

$$\frac{1 - \mu(w)}{\mu(w)} = \frac{1 - q_h}{q_h} e^{-\lambda_h k^*} = \frac{\lambda_h y - w}{w} \left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right)^{\frac{\lambda_h}{\lambda_h + r}}.$$

Since $\lim_{w \rightarrow 0} \mu(w) = 0$ and $\mu(\lambda_h y) = 1$, there must exist at least one \tilde{w} such that $\mu(\tilde{w}) = p_0$. If we set $w = \tilde{w}$, then $\tau^{NC} = k^* > 0$ by definition, and the expected payoff is

$$V(k^*, p_0) = \int_0^{k^*} (\lambda_h y p_s y - \tilde{w}) e^{-\int_0^s (\lambda_h p_u + r) du} ds,$$

which is always strictly positive and larger than (25) if λ_l is sufficiently small. ■

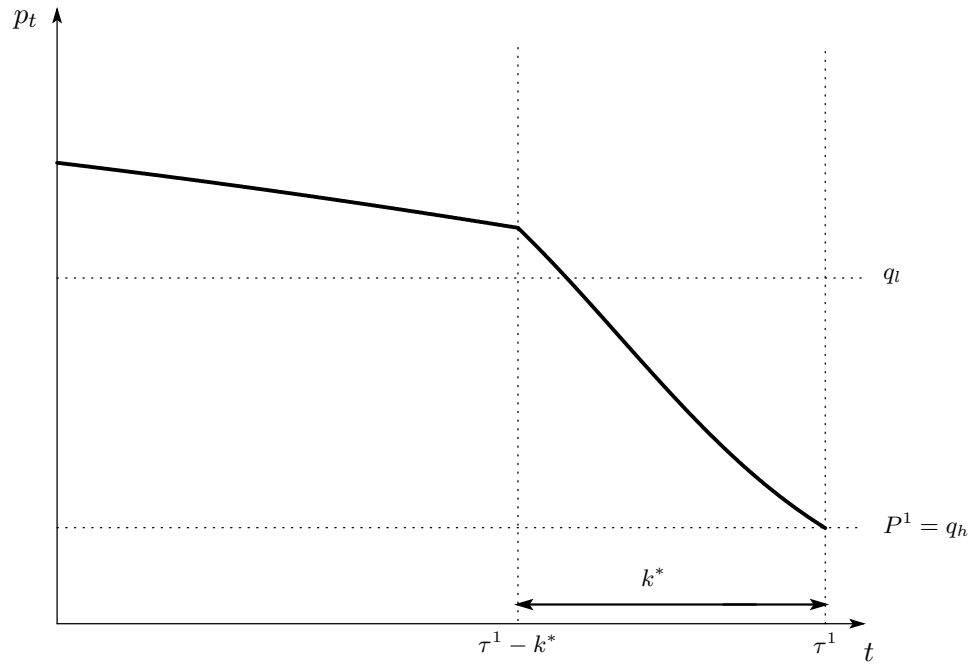


Figure 1: Evolution of the belief on the equilibrium path ($m(\Theta) = 1$)

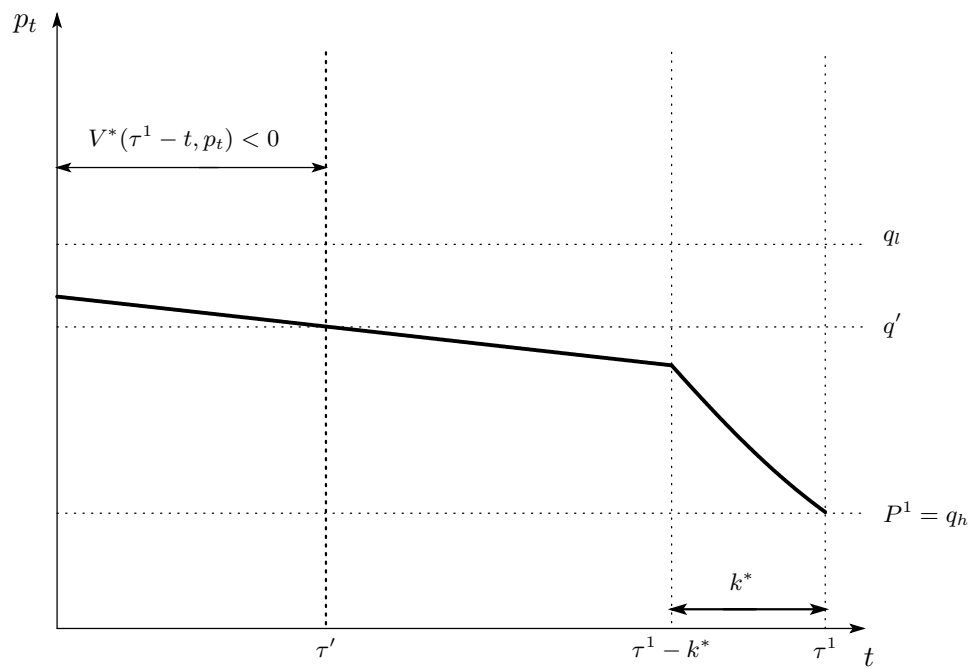


Figure 2: A violation of Condition C: the belief ($m(\Theta) = 2$)

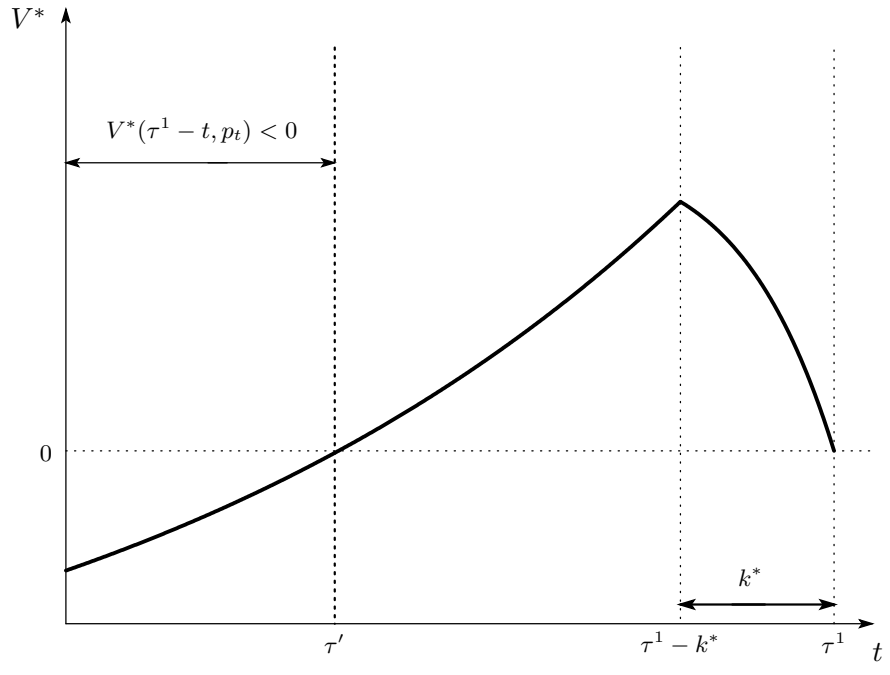


Figure 3: A violation of Condition C: the continuation payoff

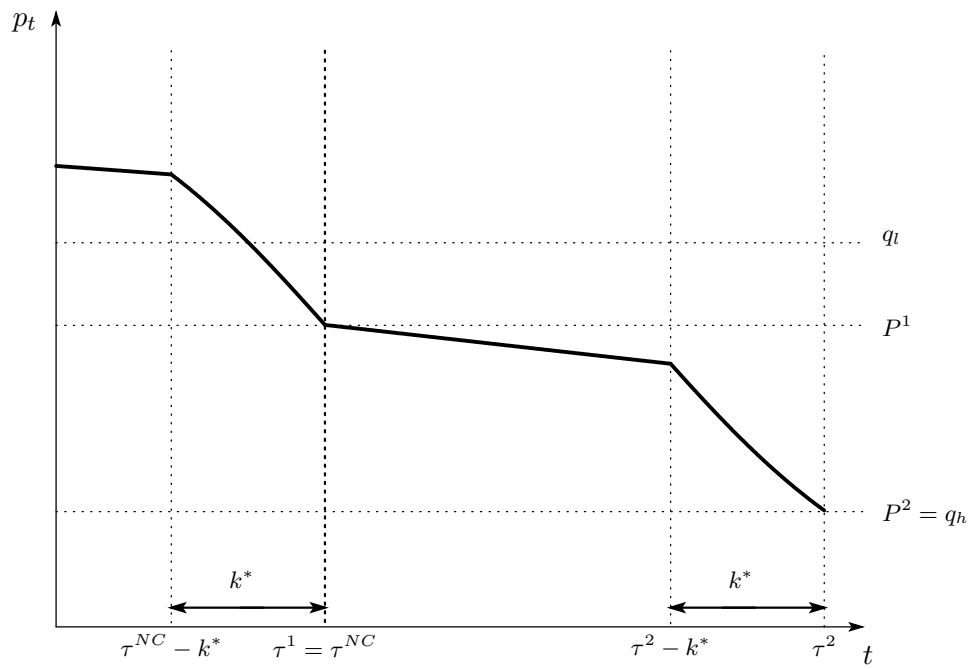


Figure 4: Evolution of the belief, on and off the equilibrium path

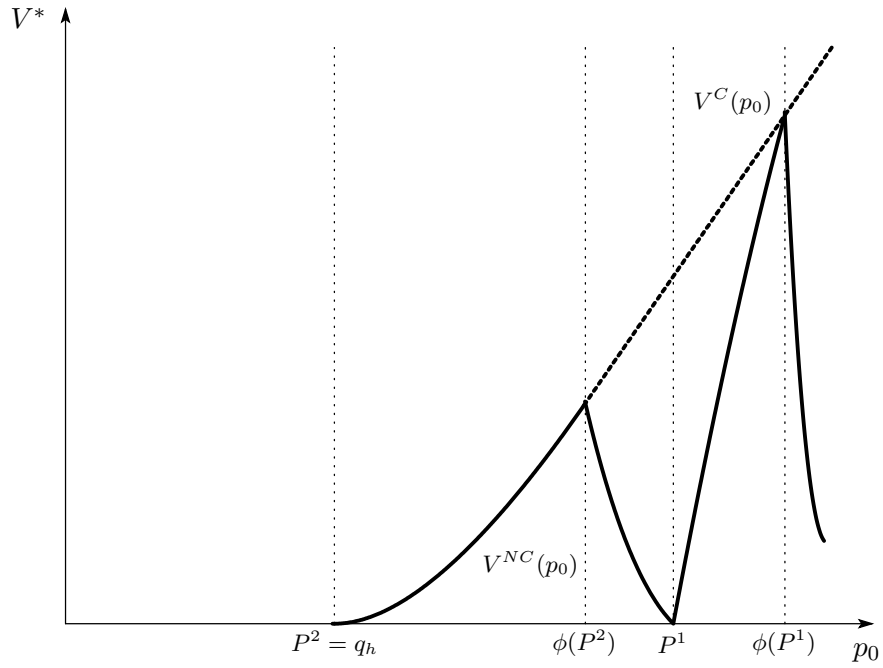


Figure 5: The expected payoffs with and without commitment

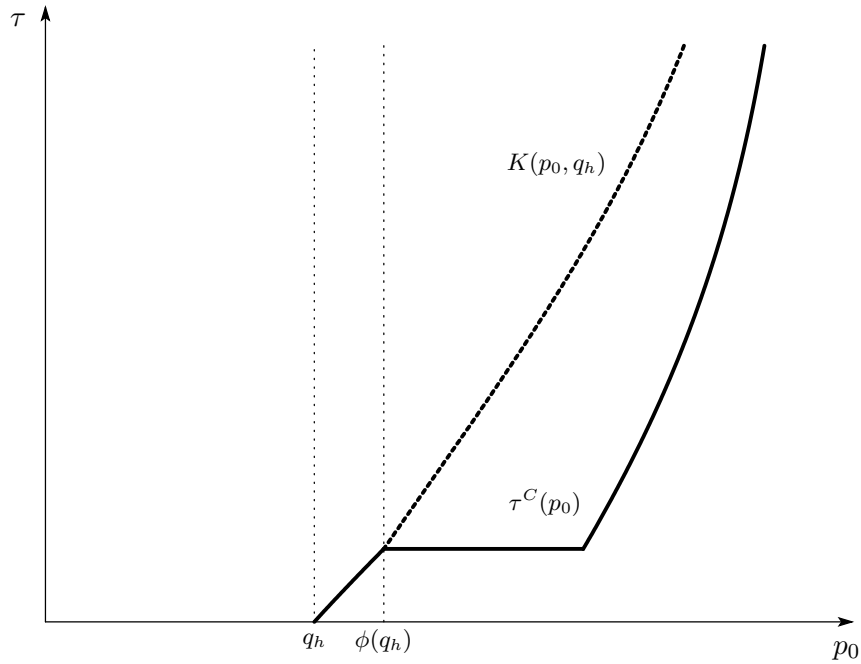


Figure 6a: The commitment solution when λ_l is relatively high ($\lambda_l = 0.07$)

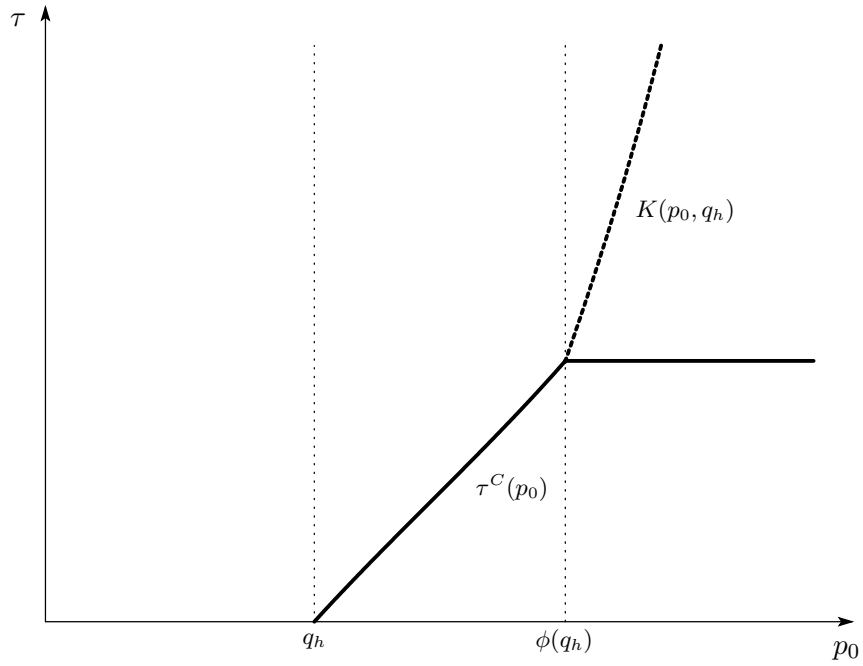


Figure 6b: The commitment solution when λ_l is relatively low ($\lambda_l = 0.04$)

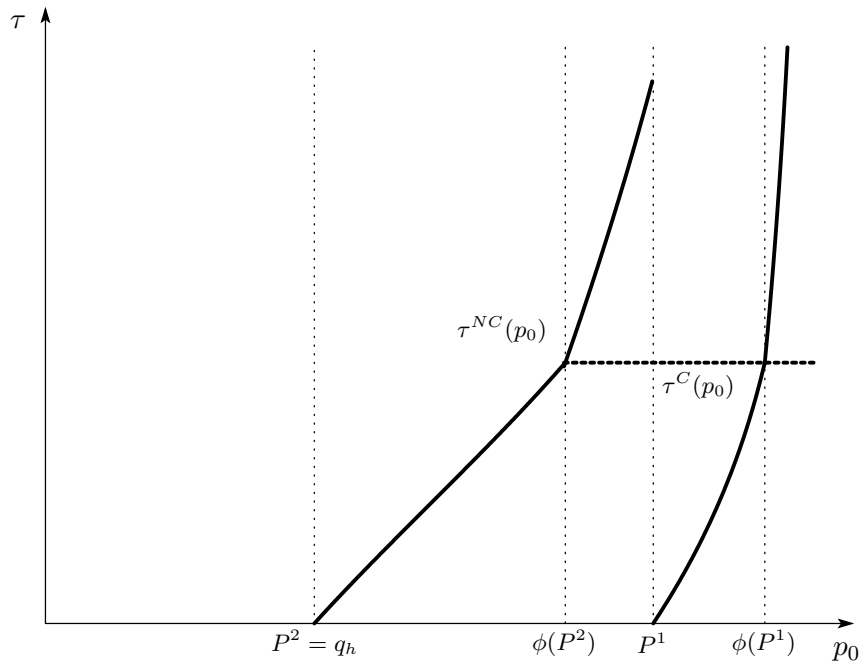


Figure 7: The no-commitment solution

Dynamic Performance Evaluation with Deadlines: The Role of Commitment*

Chia-Hui Chen[†] and Junichiro Ishida[‡]

October 20, 2017

Abstract

We consider an environment in which a principal hires an agent and evaluates his productivity over time in an ongoing relationship. The problem is embedded in a continuous-time model with both hidden action and hidden information, where the principal must induce the agent to exert effort to facilitate her learning process. The value of committing to a deadline is examined in this environment, and factors which make the deadline more profitable are identified. Our framework generates a unique recursive equilibrium structure under no commitment which can be exploited to obtain a full characterization of equilibrium. The analysis allows us to evaluate the exact value of commitment for any given set of parameters and provides insight into when it is beneficial to commit to an evaluation deadline at the outset of a relationship.

JEL Classification Number: D82, D83

Keywords: dynamic agency, deadlines, experimentation, commitment, up-or-out contracts.

*The first author acknowledges financial support from JSPS KAKENHI Grant-in-Aid for Scientific Research (B) 16H03615 and Young Scientists (B) 17K13699. The second author acknowledges financial support from JSPS KAKENHI Grant-in-Aid for Scientific Research (S) 15H05728, (A) 20245031, (B) 16H03615 and (C) 24530196 as well as the program of the Joint Usage/Research Center for Behavioral Economics at ISER, Osaka University.

[†]Institute of Economic Reserach, Kyoto University. Email: chchen@kier.kyoto-u.ac.jp

[‡]Institute of Social and Economic Research, Osaka University. Email: jishida@iser.osaka-u.ac.jp

1 Introduction

Suppose that an employer needs to hire a worker to carry out a project over time. The project is ability-intensive in that the worker can successfully complete the project only if he is sufficiently productive. As is often the case, however, the worker's sheer productivity is not directly observable to the employer, who must instead make an inference from a sequence of observed outputs. Since success cannot be achieved overnight, the employer must exercise some patience even when things do not appear to be in good shape. Excessive tolerance for failure, however, diminishes the worker's incentive to take costly actions that are indispensable to be successful. A number of issues arise in this dynamic environment. How much time should the employer give the worker before she terminates the project? Should the employer commit to a deadline at the outset? If so, under what conditions?

In this paper, we attempt to address these issues by analyzing a situation where a principal hires an agent and evaluates his productivity over time in an ongoing relationship. The problem is embedded in a continuous-time framework with both hidden action and hidden information. At each instant, the agent privately chooses how much effort to supply. The outcome is either a success or a failure, depending on his effort choice as well as his productivity type. The game ends immediately when the agent achieves a success (or a "breakthrough"). The principal's task in this environment is to determine when to terminate the project, conditional on no success having occurred. Within this setup, we analyze two distinct cases to illustrate the role of commitment: in one case, the principal sets a deadline and commits to it at the outset, and the project is terminated automatically when the deadline is reached without attaining a success; in the other, the principal makes no such commitment, thereby retaining discretion to terminate the project at any instant, and simply terminates the project when the continuation payoff is not high enough to justify further experimentation. By directly comparing these two cases, we evaluate the extent to which the principal benefits from committing to an evaluation deadline in this dynamic environment.

The driving force of our analysis is a dynamic strategic interaction between the agent's effort choice and the principal's termination strategy. On one hand, the agent's effort choice depends clearly on how much time is left until the project is terminated: since the net value of achieving a success is low when the project is still far from termination, the agent tends to start off with low effort and gradually shift to higher effort as the expected termination date approaches. On the other hand, the principal's willingness to terminate the project depends also on the agent's effort choice: when the agent is less motivated and exerts low effort, less information is revealed about his type, which makes the principal more reluctant

to terminate the project. This strategic interaction can generate a vicious cycle where the principal's reluctance to terminate the project diminishes the agent's motivation, which in turn makes the principal even more reluctant. It is in general profitable to commit to a deadline when each player's incentive to "procrastinate" is sufficiently strong.

The main contribution of the paper is that we devise an analytical framework that is tractable enough to admit a complete characterization of pure-strategy equilibria, both with and without commitment, while capturing this dynamic interaction. Our framework thus allows us to evaluate the exact value of commitment to an evaluation deadline for any given set of parameters. To this end, our main analytical focus is on the no-commitment case, which is generally far more challenging than the commitment case when dynamic interactions are considered. In the no-commitment case, the principal's termination strategy must be sequentially rational along the way, and the belief off the equilibrium path plays a crucial role. Even in this case, we can show that our framework generates a unique recursive equilibrium structure which can be exploited to establish the existence and uniqueness of (pure-strategy) perfect Bayesian equilibrium. We then build on this result to derive a necessary and sufficient condition under which the principal can strictly benefit from committing to an evaluation deadline.

As can be expected, the lack of commitment to a deadline entirely alters the dynamic allocation of effort as well as the timing of project termination. When the principal sets a deadline at the outset, she does so by taking into account how it affects the agent's entire effort sequence. In particular, since extending a deadline in general relaxes the agent's incentive compatibility constraint and reduces his total effort supply, the deadline must be set at a point where the expected benefit (of achieving a success at the next instant) equals the cost of a decrease in total effort. The situation changes rather drastically when she makes no such commitment, because her termination decision would have no influence on the agent's past behavior. As such, she terminates the project whenever the continuation payoff is about to turn negative while taking the agent's effort strategy as given. This fact implies that the cost of extending a deadline is smaller under no commitment and tends to give her an excessive incentive to wait for a success.

Given this incentive structure, one may expect that the principal would always wait longer for a success in equilibrium under no commitment. As it turns out, though, this conjecture does not always hold true in the current setup. Among other things, an interesting, and somewhat counterintuitive, property of our model is that the average duration of the project may not be monotonic with respect to the initial prior belief under no commitment, whereas it is always weakly increasing under commitment. As a direct consequence of this, the average

duration can be either shorter or longer with commitment than without: in other words, there exists an equilibrium in which the principal prematurely terminates the project when she does not commit to a deadline.

This result is somewhat surprising, provided that the inefficiency of the problem stems from the principal's reluctance to terminate the project in the first place. The principal tends to terminate the project too early when the agent's productivity under low effort is sufficiently small. For the sake of argument, suppose that the agent fails almost surely when he exerts low effort, in which case the expected instantaneous payoff is negative while the principal can learn almost nothing from failures (during the phase where the agent is supposed to exert low effort). The principal's belief then declines very slowly over time, forcing her to incur a large amount of loss if she is to wait until she totally loses her confidence in the agent's ability. If this expected loss is prohibitively large, the principal may find it optimal to terminate the project even when the belief is still relatively high.

Of course, in equilibrium, the agent correctly anticipates this reaction and adjusts his effort allocation accordingly. We show that the equilibrium timing of project termination can be pinned down by backward induction, where we start from the final critical time (to be derived) and solve backward. This process gives rise to the aforementioned recursive equilibrium structure which allows us to establish the uniqueness of equilibrium and also obtain a diverse set of equilibrium dynamics under no commitment.

The implications of our analysis can be applied broadly to a range of circumstances in which a principal (an evaluator) must assess an agent's upside potential that is only gradually revealed in an ongoing relationship, e.g., a manager who must evaluate subordinates, a professor who must evaluate graduate students, a head coach in professional sports who must evaluate players, and so on. Among those possibilities, the most prominent example of evaluation schemes with deadlines is perhaps the "up-or-out system," which is widely observed in academia and professional service industries such as law, accounting, and consulting. As a specific application, our framework offers some insight for when up-or-out contracts are more valuable by identifying several key factors – such as high ability intensity, stable job descriptions, and similar jobs across ranks – which favor the use of an evaluation deadline from a previously unexplored channel. Each of these factors intensifies either the agent's incentive to delay exerting high effort for a given deadline or the principal's incentive to delay terminating the project for a given effort sequence (or both), thereby rendering it more profitable to set a deadline at the outset.

The paper is organized as follows. The literature review is provided in the remainder of this section. The model is presented in section 2 and analyzed in sections 3 and 4, where we

characterize both the commitment and no-commitment solutions. These solutions are then compared in section 5, in order to analyze the value of commitment and derive a necessary and sufficient condition under which it is strictly optimal to commit to a deadline. Some extensions of the baseline model and concluding remarks are offered in section 6.

Related Literature: The current analysis is most closely related to the experimentation literature in that the principal here attempts to uncover the agent’s type through a sequence of experiments.¹ From the principal’s point of view, our model can be seen as a variant of the canonical two-armed bandit problem with one safe arm (terminating the project) and one risky arm (continuing the project). Our model is particularly related to the literature on strategic experimentation which analyzes a situation where a group of individuals, rather than a single individual, face bandit problems (Bolton and Harris, 1999; Bergemann and Välimäki, 1996, 2000; Keller et al., 2005; Klein and Rady, 2011; Bonatti and Hörner, 2011). A crucial difference from this strand of literature is that experiments in our context are “intermediated,” i.e., experiments are conducted not by the principal herself but by an informed intermediary, the agent, who faces no uncertainty about the project.²

Recently, there have been increasingly many works that explore the optimal provision of incentives in bandit problems. Manso (2011) considers the classic two-armed bandit problem and shows that the optimal contract in this context must tolerate, or even reward, early failures in order to encourage exploratory activities. Bergemann and Hege (1997, 2005) and Hörner and Samuelson (2013) analyze a financing problem of a venture capitalist where the principal provides funding to the agent who conducts experiments on a project of unknown quality.³ Gerardi and Maestri (2012) consider a similar environment where an agent conducts experiments but assume that the outcome of each experiment can only be observed by the agent. The principal must hence devise a contract not only to induce costly effort but also to truthfully reveal the information. Halac et al. (2013) analyze a model of long-term contracting for experimentation with hidden information about the agent’s ability and dynamic moral hazard and obtain an explicit characterization of optimal contracts. Aside from

¹An early economic application of the bandit problem can be found in Rothschild (1974). See Bergemann and Välimäki (2008) for a survey.

²Several recent works analyze models of “delegated experimentation” where a principal delegates experimentation to an agent. Guo (2014) analyzes a situation where the principal can specify, with full commitment power, how the agent should allocate the resource in all future contingencies and solves for the optimal delegation rule. Garfagnini (2011) considers a similar setting to ours but assumes that: (i) the principal and the agent are symmetrically informed about the state of nature; and (ii) the agent’s payoff is independent of the state. Chen and Ishida (2015) consider the opposite case in which the principal, with the termination right, may be privately informed about the project quality while the agent focuses on implementing the project assigned to him.

³Also, see Buisseret (2016) who considers a two-period model of this setting with a more general, convex cost function.

some technical differences,⁴ these previous works are primarily concerned with characterizing optimal contracts. In contrast, the aim of this paper is to compare the allocations under full commitment and no commitment, while considering a less complete contractual environment, in order to evaluate the extent to which the principal can gain from committing to a deadline at the outset. Although it is hard to tell *a priori* which specification is more plausible, as it depends on the details of the underlying situation, our simple and tractable framework allows us to identify and illuminate the counterintuitive role of commitment in dynamic performance evaluation.

Several recent works examine the role of commitment in dynamic moral hazard setting. Mason and Välimäki (2015) consider a dynamic moral hazard problem and derive optimal wage contracts, both with and without commitment on wage payments. Hörner and Samuelson (2016) consider a repeated-game setting in which the principal chooses the scale of the project in addition to contingent payments and characterize the set of equilibrium payoff vectors that can be achieved without commitment.⁵ A key difference from our model is that these two works do not consider the principal’s learning, which is our main focus, with no hidden information about the agent’s type; as a consequence, the project is never terminated in their models.⁶ This stands in sharp contrast to our setting in which the principal accumulates information about the agent over time, and the project must be terminated in finite time.⁷

It is well known that players often wait until the deadline to reach an agreement in finite-horizon models. This behavior, which is referred to as the deadline effect, is a topic of utmost concern in many bargaining and war of attrition models (Hendricks et al., 1988; Spier, 1992; Fershtman and Seidman, 1993; Hörner and Samuelson, 2011; Chen, 2012; Damiano et al., 2012; Fuchs and Skrzypacz, 2013). Some recent works also explore the role of deadlines in dynamic problems with multiple agents. Bonatti and Hörner (2011) analyze a dynamic

⁴As a key technical difference, we consider a case where the agent knows his own productivity (or the project quality), so that our model belongs to the class of dynamic signaling (with stochastic signals), rather than of experimentation, from the agent’s point of view.

⁵In their model, outcomes are privately observed by the agent, and the moral-hazard problem regards the truthful disclosure of this private information.

⁶Georgiadis et al. (2014) analyze the role of commitment in a dynamic contribution games where the manager has the decision right over the project size. With multiple agent types and dynamic learning, Bonatti and Hörner (2017) analyze a (symmetric-information) experimentation model in which wages are determined competitively a la Holmström (1999) and characterize effort and wage dynamics with an exogenous termination date.

⁷A subtle technical difference which directly arises from this fact lies in the ways in which to construct an equilibrium. Mason and Välimäki (2016) and Hörner and Samuelson (2017) construct an equilibrium via arguments found in the analysis of infinitely repeated games (reversion to the worst continuation equilibrium), and as such, their analyses yield a non-degenerate set of equilibria. In our analysis, an equilibrium is obtained via backward induction which gives rise to a unique equilibrium.

moral-hazard problem with a team in which multiple agents work on a project of unknown quality and briefly discuss the optimal deadline in this context. Campbell et al. (2013) consider a similar environment where two agents work jointly on a project. They assume that there is only one project type but assume that one’s own outcomes are his private information. Each agent can exert effort to produce a breakthrough individually, and a successful agent can reveal that he has been successful. The focus of these works is placed on the interaction between the agents, especially the freerider problem, whereas ours is on the dynamic interaction between the principal’s termination decisions and the agent’s effort choices.

2 A dynamic model of performance evaluation

Environment: We employ a continuous-time model because of its greater tractability. Consider a situation where a principal (female) hires an agent (male) to complete a project. The game ends either when the agent attains a success (a “breakthrough”) or when the principal terminates the project. The agent is either good with prior probability $p_0 \in (0, 1)$ or mediocre with probability $1 - p_0$. The ability type is the agent’s private information and is not directly observable to the principal who must instead evaluate it from a sequence of observed outcomes.

Production: The agent makes unobservable effort $a_t \in \{l, h\}$ at each instant t .⁸ We interpret that low effort ($a_t = l$) refers to the minimum level of effort that can be induced via input monitoring while high effort ($a_t = h$) refers to any part of effort that cannot be directly monitored by any means. The instantaneous cost of effort a is denoted by d_a where $d_h = d > 0$ and $d_l = 0$. A success arrives stochastically, depending on the effort choice as well as the agent’s type. If the good type chooses $a_t = a$ over time $[t, t + dt)$, he attains a success with probability $\lambda_a dt$ where $\lambda_h > \lambda_l > 0$. In contrast, the mediocre type can never succeed with any effort level.⁹ Define $\Delta_\lambda := \lambda_h - \lambda_l$.

Payoff: We consider an incomplete-contracting environment where contingent rewards on the arrival of a success cannot be enforced.¹⁰ A success yields a net present value of $y > 0$

⁸Our focus on binary effort reflects our implicit presumption that the effort cost and success probability are linear, as usually assumed in this literature (e.g., Keller et al, 2005; Bonatti and Horner, 2011, for most of their analysis). As long as this structure is maintained, an extension to continuous effort, say $a_t \in [l, h]$, yields exactly the same allocation and is hence irrelevant.

⁹Our model specification is thus the “breakthrough” type in which one success can resolve all the uncertainty regarding the agent’s type – an assumption that is predominant in the experimentation literature. See, for instance, Keller et al. (2005) and Bonatti and Hörner (2011, 2017).

¹⁰In reality, we rarely observe complete wage contracts in industries comprised of professionals. For instance, few academic institutions, if any, offer rewards specifically contingent on verifiable measures of output: salaries

to the principal and $b > 0$ to the agent;¹¹ otherwise, they both receive zero. Aside from this, the principal must also pay a flow wage $w > 0$ to the agent as long as the project continues.¹² We assume that w is exogenously given for most part; the case with an endogenous wage is briefly discussed in the concluding section. The reservation payoff is assumed to be zero for both players. The common discount rate is denoted by $r \in (0, \infty)$.

Contract: The only contractible decision for the principal in this environment is whether to set a deadline, and if so, at what point. If the principal commits to a deadline, she terminates the project at the deadline (but never before) if the agent has not attained a success up to that point. If the principal chooses not to commit to any specific deadline, on the other hand, she retains discretion to terminate the project at any instant. As mentioned earlier, our analytical focus is on the latter case which requires that both players' strategies be sequentially rational.

3 Analysis

3.1 Agent's effort decision

The agent decides whether or not to exert high effort at each instant. To analyze this problem, it is important to note that the principal's belief affects the agent's payoff only through her termination decision. The agent's optimal effort choice thus depends only on the remaining time to the termination date (hereafter, simply the remaining time), i.e., the maximum length of time for which the principal continues the project without attaining a success. The remaining time is obvious when the principal sets a deadline at time τ , in which case the remaining time at time t is simply given by $\tau - t$. Even without such an explicit commitment, however, the remaining time can be computed from the principal's equilibrium strategy in essentially the same manner as we shall discuss below. For now, we proceed with the presumption that the remaining time exists and is well-defined at any given point in time.

are determined through bilateral negotiations, often dictated by market forces, in some countries whereas they are subject to bureaucratic regulations in others. One possible reason for the lack of complete wage contracts in those industries is that it is often difficult, and perhaps prohibitively costly, to measure the exact value of a "success" in a verifiable manner. Additionally, this assumption makes our analytical framework applicable to a wider range of circumstances. One such possibility is that the benefit of achieving a success accrues from non-transferrable (psychological) gains such as prestige, authority, and the sense of achievement, and extrinsic rewards are hence of secondary importance. There are also many cases where contingent monetary transfers are neither feasible nor desirable, as in a professor-student relationship.

¹¹An obvious interpretation is that y and b represent the continuation payoffs of achieving a success, including not only the intrinsic value of a successful outcome but also other benefits of identifying/signaling talent.

¹²Alternatively, w can be regarded as a flow cost of production (e.g., hiring an agent). Of course, under this interpretation, w is no longer a transfer payment to the agent and does not appear in his payoff. It is straightforward to make the setup consistent with this interpretation without having any qualitative impact on our results.

The agent's problem is rather straightforward since the mediocre type, knowing that his marginal value of effort is zero, never exerts high effort. As such, we can focus on the good type whom we refer to simply as the agent in what follows. Denote by $U(k)$ the agent's value function when the remaining time is k . Taking $k > 0$ as given, the value function can be written as

$$U(k) = \max_{a \in \{l, h\}} ((\lambda_a b - d_a + w)dt + e^{-rdt}(1 - \lambda_a dt)U(k - dt)).$$

Taking the limit $dt \rightarrow 0$, we obtain the Bellman equation:

$$rU(k) = \max_{a \in \{l, h\}} (\lambda_a b - d_a + w - \lambda_a U(k) - \dot{U}(k)), \quad (1)$$

with $\lim_{k \downarrow 0} U(k) = 0$. It is clear from this that the agent chooses high effort if and only if

$$\Delta_\lambda(b - U(k)) \geq d.$$

As usual in this type of setup (see, e.g., Bergemann and Hege, 2005), the continuation payoff $U(k)$ captures the agent's reservation payoff which he receives in case of a failure. The cost of not succeeding today is obviously small when he has a high reservation payoff. Since the continuation payoff is higher when the agent has more time to prove himself, the agent has a stronger incentive to work hard as the project approaches the termination date. This is a manifestation of the deadline effect that lies at the core of our entire analysis.

Proposition 1 *If*

$$\frac{d}{\Delta_\lambda} > \frac{rb - w}{\lambda_l + r}, \quad (2)$$

there exists k^ such that the good type exerts high effort if the remaining time is less than or equal to k^* and low effort otherwise. The threshold k^* is given by*

$$k^* = \begin{cases} -\frac{1}{\lambda_h + r} \ln\left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w}\right) & \text{if } b > \frac{d}{\Delta_\lambda}, \\ 0 & \text{if } \frac{d}{\Delta_\lambda} \geq b. \end{cases} \quad (3)$$

If (2) does not hold, the agent always exerts high effort.

Proof: See Appendix.

When $\Delta_\lambda b \leq d$, the static incentive is too weak for the agent to exert high effort for any remaining time. In contrast, when (2) fails to hold, the static incentive is strong enough to overcome the dynamic agency cost, and the agent is willing to supply high effort under any circumstance. As these cases only result in trivial solutions and are clearly of less interest for the question we pose here, we restrict our attention to the case where the strength of the static incentive lies in some intermediate range by making the following assumption.

Assumption 1 $\frac{\lambda_h b - d + w}{\lambda_h + r} > b - \frac{d}{\Delta_\lambda} > 0$.

Among other things, the assumption implies that the optimal threshold k^* is bounded from above and away from zero, i.e., $k^* \in (0, \infty)$.

3.2 Principal's termination decision

Let p_t denote the principal's belief that the agent is good at time t , conditional on no success having occurred. Given some effort sequence $\{a_s\}_{s=0}^t$, the updated belief is then given by

$$p_t = \frac{p_0 e^{-\int_0^t \lambda_{a_s} ds}}{1 - p_0 + p_0 e^{-\int_0^t \lambda_{a_s} ds}}. \quad (4)$$

Alternatively, taking the time derivative, we obtain

$$\dot{p}_t = -\lambda_{a_t} p_t (1 - p_t). \quad (5)$$

It is clear that the belief is strictly decreasing over time for any effort choice (due to the fact that $\lambda_l > 0$) until the agent attains a success, in which case the belief immediately jumps up to one.

The principal's problem is to determine when to terminate the project, conditional on no success having occurred. Now suppose that the principal intends to terminate the project at time $\tau > t$. Taking the effort sequence as given, the principal's continuation payoff can also be written as a function of the remaining time k and the current belief p_t where

$$V(k, p_t) = \int_t^{t+k} (\lambda_{a_s} p_s y - w) e^{-\int_t^s (\lambda_{a_u} p_u + r) du} ds,$$

subject to (5).

It is immediate to see that for a given effort level, there exists a threshold belief below which the principal's instantaneous payoff is strictly negative. We denote by $q_a := \min\{\frac{w}{\lambda_a y}, 1\}$ the "break-even" belief at which the instantaneous payoff equals zero under effort a . Combined with the fact that the belief is strictly decreasing over time for any effort sequence, q_h represents the absolute lower bound of the belief, as the principal clearly has no incentive to continue the project once her belief dips below this level. Since the belief must reach this level sooner or later (due to the fact that $\lambda_l > 0$ and $w > 0$), the presence of such a lower bound suggests that the game must end in some finite time. This allows us to solve the game by backward induction.

Finally, if the value of a success is too small, the model only admits a trivial solution where the principal chooses to stop immediately (or not to hire the agent in the first place). In what follows, therefore, we assume that the value of a success is large enough for the principal to hire the agent at least for some positive duration.

Assumption 2 $p_0 > q_h = \frac{w}{\lambda_h y}$.

Since the principal's continuation payoff depends on her belief, it is often convenient for the subsequent analysis to explicitly relate the current and termination beliefs to the remaining time. Suppose that the current belief is p and the principal intends to terminate the project when the belief reaches q . If the agent adopts the best response, the remaining time can be written as $K(p, q)$ such that

$$\frac{pe^{-\lambda_l \max\{K(p,q)-k^*, 0\}} - \lambda_h \min\{K(p,q), k^*\}}{1 - p + pe^{-\lambda_l \max\{K(p,q)-k^*, 0\}} - \lambda_h \min\{K(p,q), k^*\}} = q,$$

which reduces to

$$\frac{pe^{-\lambda_l K(p,q) - \Delta_\lambda k^*}}{1 - p + pe^{-\lambda_l K(p,q) - \Delta_\lambda k^*}} = q,$$

if $K(p, q) > k^*$. In addition, it is also convenient to define a backward operator $\phi(q)$ such that

$$\frac{\phi(q)e^{-\lambda_h k^*}}{1 - \phi(q) + \phi(q)e^{-\lambda_h k^*}} = q.$$

The backward operator suggests that if $p_t = \phi(q)$ and the agent exerts high effort for $t \in [t, t + k^*]$, then $p_{t+k^*} = q$; alternatively, if the agent expects the project to be terminated at $p_t = q$, $\phi(q)$ indicates the belief at which he must switch to high effort. Note that these two notions are closely related in that $K(p, q) > k^*$ if and only if $p > \phi(q)$.

4 Equilibrium

4.1 The cooperative solution: a benchmark

Before we move on to analyze our model, we first examine as a benchmark the case where both the principal and the agent attempt to maximize the sum of their individual payoffs to derive the first-best allocation. More precisely, the instantaneous payoff in the cooperative case, common to both players, is the difference between the expected benefit from a success minus the effort cost, i.e., $\lambda_a p(y + b) - d_a$ for a given effort a .¹³

The agent's problem is essentially the same and only needs a slight modification. Letting $\tilde{U}(k)$ denote the agent's continuation payoff, the agent exerts high effort if and only if

$$\Delta_\lambda((y + b) - \tilde{U}(k)) \geq d.$$

¹³If the agent shared the same objective as we assume here, his private information could in principle be induced at no cost: given that the agent's reservation payoff is zero, the mediocre type is indifferent between participating and dropping out immediately (before time 0), as he would obtain zero payoff in either case for any given deadline. As it turns out, however, the optimal deadline is independent of the initial belief p_0 and the analysis holds irrespective of whether or not the principal can immediately screen out the mediocre type.

As in the noncooperative case, there exists a threshold \tilde{k}^* such that the good type exerts high effort if and only if the remaining time is less than or equal to \tilde{k}^* . Under the maintained assumptions, one can easily verify that the agent is generally more motivated in the cooperative case than in the noncooperative case, i.e., $\tilde{k}^* > k^*$.

Given this, it is fairly straightforward to derive the cooperative solution. Since there is no social cost of hiring the mediocre type, the joint surplus from hiring the agent is strictly positive for any belief level, and as such, it is efficient for the principal to continue the project indefinitely until a success is attained. This implies that the good type will eventually succeed at some point, although he only exerts low effort all the way if \tilde{k}^* is finite.

Proposition 2 *In the cooperative case, the principal never terminates the project, and the game continues indefinitely. The agent always exerts low effort if*

$$\frac{d}{\Delta_\lambda} > \frac{r(b+y)}{\lambda_l + r}, \quad (6)$$

and high effort otherwise.

Proof: See Appendix.

Note the difference between (2) and (6). In the cooperative case, the joint expected payoff of exerting low effort indefinitely is $\frac{r(b+y)}{\lambda_l+r}$ because the value of a success is now $b+y$, instead of just b , and there is no flow cost w .

4.2 The commitment solution

The principal's problem under commitment is to choose a termination date τ which maximizes her expected payoff at time 0 while restricting attention to deterministic deadlines.¹⁴ We focus exclusively on deterministic deadlines primarily because it is generally difficult to commit to a lottery, even when it is possible to commit to a particular termination date. To see this point, consider a contract where the principal terminates the project at τ^1 or τ^2 , $\tau^1 < \tau^2$, each with probability strictly less than one. At time τ^1 , however, the principal's preferences are generically strict, i.e., it is strictly better either to stop (the continuation payoff is negative) or to continue (positive), so that there is no incentive to randomize at that point. If it is strictly better to stop at time τ^1 , for instance, the principal stops at time τ^1 with probability one but

¹⁴If stochastic deadlines are feasible, the principal may screen out the low type by offering a menu of contracts, one with a deterministic deadline and the other with stochastic ones. We do not pursue this possibility both because we do not think that it is realistic (given the enforcement problem) and because it is outside the scope of our analysis (given that our focus is on comparing the commitment and no-commitment solutions). Note that this screening cannot be done if deadlines are deterministic, because both types always strictly prefer a longer deadline.

knowing that, the contract is effectively reduced to the one with a deterministic deadline. The lack of credible enforcement is perhaps the reason why we almost never observe a contract with stochastic deadlines.

As we have already seen, the agent exerts high effort if and only if the remaining time is less than or equal to k^* . We denote by $V^*(k, p_t)$ the continuation payoff when the agent adopts this best response, where k is the remaining time and p_t is the principal's belief. This can be written as

$$\begin{aligned} V^*(k, p_t) &= \int_t^{t+\nu} (\lambda_l p_s y - w) e^{-\int_0^s (\lambda_l p_u + r) du} ds \\ &\quad + e^{-\int_t^{t+\nu} (\lambda_l p_u + r) du} \int_{t+\nu}^{t+k} (\lambda_h p_s y - w) e^{-\int_\nu^s (\lambda_h p_u + r) du} ds \\ &= \frac{p_t \pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\nu}) + \frac{p_t e^{-(\lambda_l + r)\nu} \pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)(k-\nu)}) - \frac{(1-p_t)w}{r} (1 - e^{-rk}), \end{aligned} \quad (7)$$

where $\nu := \max\{k - k^*, 0\}$ and $\pi_a := \lambda_a y - w$. The first two terms represent the expected gain from the good type whereas the last term represents the expected loss from the mediocre type.

Since the termination date is equivalent to the remaining time at time 0, the commitment solution, denoted by τ^C , is given by

$$\tau^C(p_0) := \operatorname{argmax}_\tau V^*(\tau, p_0).$$

The continuation payoff is continuous but kinked at $\tau = k^*$. For $\tau \in (0, k^*)$, the first-order condition is given by

$$p_0 \pi_h e^{-\lambda_h \tau} - (1 - p_0)w = 0. \quad (8)$$

For $\tau \in (k^*, \infty)$, it is given by

$$M(\tau, p_0) := p_0 e^{-(\lambda_l + r)(\tau - k^*)} \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h (1 - e^{-(\lambda_h + r)k^*}) \right) - (1 - p_0)w e^{-r\tau} = 0. \quad (9)$$

To better understand this condition, it is instructive to rewrite this as

$$(p_0 \pi_h e^{-(\lambda_l(\tau - k^*) + \lambda_h k^*)} - (1 - p_0)w) e^{-r\tau} = \frac{p_0 \Delta \lambda e^{-(\lambda_l + r)(\tau - k^*)}}{\lambda_h + r} (w + ry + \pi_h e^{-(\lambda_h + r)k^*}).$$

Note that the left-hand side is the expected instantaneous payoff for the principal at the time of termination, whereas the right-hand side is the marginal cost of extending the deadline. Since the right-hand side is strictly positive, the principal must set the deadline at a point where the expected instantaneous payoff is positive if $\tau > k^*$.

Since the instantaneous payoff is negative for any effort level once the belief dips below q_h , there is clearly no incentive to set $\tau > K(p_0, q_h)$. When the initial belief is sufficiently close to q_h ($\phi(q_h) \geq p_0$), it is optimal to let the agent work until the belief reaches this lower bound. When it is far from it ($p_0 > \phi(q_h)$), on the other hand, the principal must terminate the project at a point where the belief is still above the lower bound. There are two cases, depending on the value of $M(k^*, p_0)$.¹⁵ If

$$\frac{\lambda_l + r}{\lambda_h + r} \pi_h (1 - e^{-(\lambda_h + r)k^*}) > \pi_l, \quad (10)$$

$M(k^*, p_0) < 0$ for any $p_0 \in (q_h, 1)$, and we let $\hat{p} = 1$ in this case. If (10) fails to hold, on the other hand, there exists a unique \hat{p} such that $M(k^*, \hat{p}) = 0$. Then, for $p_0 > \hat{p}$, there exists a unique interior solution $\hat{\tau}(p_0)$ such that $M(\hat{\tau}(p_0), p_0) = 0$. When the initial prior is in this range, it is not so costly to have a phase where the agent exerts low effort, and as such, the principal would wait beyond time k^* . Of course, extending the deadline beyond time k^* means that the agent would slack off at the beginning, thereby pushing back the realization of the high-effort phase. The principal trades off the gain of inducing low effort for an additional instant against the cost of realizing the high-effort phase an instant later and sets the termination date at a point where they are equalized.

For clarity, we summarize the set of parameters by $\Theta := (b, y, w, d, \lambda_h, \lambda_l, r)$. We can then make the following statement.

Proposition 3 *For any given set of parameters (Θ, p_0) satisfying Assumptions 1 and 2, there exists a unique commitment solution $\tau^C(p_0) \in (0, K(p_0, q_h)]$. For $\phi(q_h) \geq p_0$, the commitment solution is given by $\tau^C(p_0) = K(p_0, q_h)$. For $p_0 > \phi(q_h)$, there exists some $\hat{p} \in (q_l, 1]$ such that*

$$\tau^C(p_0) = \begin{cases} k^* & \text{if } \hat{p} \geq p_0 > \phi(q_h), \\ \hat{\tau}(p_0) & \text{if } p_0 > \hat{p}, \end{cases}$$

where: (i) $\hat{p} > q_l$ if $1 > q_l$; and (ii) $\hat{p} = 1$ if and only if (10) holds. The commitment solution is weakly increasing in p_0 .

Proof: See Appendix.

Note that $K(p_0, q_h)$ constitutes the upper bound of $\tau^C(p_0)$. Since $K(p_0, q_h)$ is bounded from above when $q_h > 0$, the proposition suggests that the project must be terminated in finite time, meaning that some good projects are bound to be terminated prematurely. This draws clear contrast with the cooperative solution where the project is never terminated.

¹⁵Note that $M(k, p_0) < 0$ for any $k > k^*$ and p_0 if and only if $M(k^*, p_0) < 0$.

4.3 The no-commitment solution

The situation becomes more complicated, and perhaps more intriguing, when the principal makes no commitment at the outset and hence her termination decision must be sequentially rational. Because of this, the principal's termination strategy now needs to be specified in a different way, as her strategy off the equilibrium path (after the project is supposed to be terminated) may matter. More precisely, the principal's strategy is given by a set of termination dates T which consists of a first termination date and further dates along off-equilibrium continuation paths. Formally, for this no-commitment case, we solve for a pure-strategy perfect Bayesian equilibrium (hereafter, simply an equilibrium) in which the principal terminates the project with probability one at each termination date (a formal definition is given below).¹⁶

To illustrate how we specify the principal's strategy, consider a simple example where the principal's strategy consists of two distinct termination dates, i.e., $T = \{\tau^1, \tau^2\}$ where $\tau^1 < \tau^2$. Under this strategy, the principal terminates the project with probability one at each τ^i , $i = 1, 2$, conditional on the continuation of the game.¹⁷ This means that the project is surely terminated at time τ^1 on the equilibrium path. To show that this constitutes an equilibrium, however, one must assure that the principal cannot profitably deviate from this strategy. To see this, suppose that the principal unexpectedly continues the project at time τ^1 . The game then enters into a new phase where the agent now expects the project to be terminated at the next termination date τ^2 and chooses his effort accordingly ($a_t = h$ if $k^* \geq \tau^2 - t$ and $a_t = l$ otherwise). Note also that the principal's belief about the agent's type is not affected by the deviation: given T and the agent's best response, we can still apply Bayes' rule to compute how the belief evolves both on and off the equilibrium path.¹⁸ This gives the continuation payoff when the principal deviates which must be non-positive.

How long does this process continue? It generally goes on until the principal has no

¹⁶We restrict our attention to pure-strategy equilibria by assuming that the principal terminates the project with probability one when the continuation payoff is zero. Since the principal is actually indifferent at this point, however, there may exist other mixed-strategy equilibria in which the principal terminates the project with probability less than one at each termination date. We do not pursue this possibility for two reasons. First, we do not believe that such an equilibrium is particularly realistic and hence appealing. Second, even with randomization, the game still must end with probability one by time τ^1 , and the continuation equilibrium is exactly the same once the game enters the final interval (τ^1, τ^2) ; solving backwards, one can see that the structure of equilibrium is essentially the same where the agent starts exerting high effort as t approaches each termination date and switches back to low effort if the project is not terminated at the termination date.

¹⁷If the game continues beyond time τ^2 , the principal terminates the project immediately as we will see below.

¹⁸Technically, we assume that the belief depends only on the agent's past effort and is given by (4) even off the equilibrium path. This is actually the "no signaling what you don't know" condition (Fudenberg and Tirole, 1991) although it is a rather obvious restriction in games with two players (it only implies that the principal's belief should not be affected by her own deviation).

incentive to continue the project regardless of the agent's effort choice or, in other words, until the belief reaches the lower bound q_h . With slight abuse of notation, we denote by p_t the equilibrium belief at time t , conditional on the continuation of the game (including the off-equilibrium path). Let τ^n denote the time at which p_t reaches q_h for some n (which is unknown at this point). We can then restrict our attention to $t \in (0, \tau^n]$, for it is a dominant strategy for the principal to terminate the project once the belief reaches this lower bound.¹⁹

Lemma 1 *In any equilibrium, T is a finite set.*

Proof: The lemma is directly implied by the fact that any two adjacent termination dates must satisfy $\tau^i - \tau^{i-1} \geq k^* > 0$. To see this, suppose otherwise, i.e., $k^* > \tau^i - \tau^{i-1}$. Suppose further that the principal deviates and continues the project at time τ^{i-1} . Then, it is strictly better for the agent to exert high effort because $k^* > \tau^i - \tau^{i-1}$. Note, however, that since $p_t > q_h$ for any $t \in (0, \tau^n)$ by definition, the instantaneous payoff is strictly positive until the belief reaches the next termination belief. This is a contradiction because the principal can strictly benefit from deviating and not stopping at time τ^{i-1} . Given this, it is clear that we can only have finitely many termination dates in $(0, \tau^n]$. ■

Given this result, we denote each element of T by τ^i where $\tau^1 < \tau^2 < \dots < \tau^n$ for some n . Accordingly, the game is divided into n distinct segments T^i , $i = 1, 2, \dots, n$ by termination dates, where $T^i = (\tau^{i-1}, \tau^i)$ for $i = 1, 2, \dots, n$ with $\tau^0 = 0$. The remaining time at time t is now given by $\tau^i - t$ for $t \in T^i$ while it is zero for all $t > \tau^n$ (see footnote 19). Let $q^i := p_{\tau^i}$ denote the corresponding termination belief. Only the first segment T^1 is actually played on the equilibrium path, as the game ends with probability one by the time the game reaches time τ^1 . For expositional clarity, we call τ^1 the no-commitment solution and denote it by $\tau^{NC}(p_0)$ to indicate its dependence on p_0 . The formal definition of our equilibrium is given below.

Definition 1 *A perfect Bayesian equilibrium in this game is a pair of strategies $\{a_s\}_{s=0}^\infty$ and $T := \{\tau^i\}_{i=1}^n$ and a belief system such that:*

- *given T , the (good-type) agent chooses a_t at each t to maximize his continuation payoff, i.e., chooses high effort if and only if the remaining time at time t is less than or equal to k^* ;*

¹⁹ Suppose that the principal deviates and continues beyond time τ^n . In this case, the principal's instantaneous payoff is negative even if the agent chooses $a_t = h$, so that it is a dominant strategy to terminate immediately for all $t > \tau^n$. Given this, since the remaining time is invariably zero, the agent always chooses $a_t = h$. This is the unique continuation equilibrium after time τ^n (although, to describe this process formally, we need to consider a discrete-time counterpart of our model where there is a minimum time unit, and take the limit as the time unit goes to zero).

- given $\{a_s\}_{s=t}^{\infty}$ and the current belief, the principal terminates the project at time t if and only if the continuation payoff is non-positive;
- The belief at time t depends only on the effort sequence $\{a_s\}_{s=0}^t$ and is given by (4), both on and off the equilibrium path.

To pin down the equilibrium termination dates, we need to pay closer attention to the principal's continuation payoff. Since the agent's best response is the same, the principal's continuation payoff can still be written as $V^*(k, p_t)$. The only difference is that the principal's strategy must now be sequentially rational, which amount to the following two equilibrium conditions which the principal's termination strategy T must satisfy.

Condition T: (i) $q^n = q_h$ and (ii) if $n > 1$, for each $i = 2, \dots, n$, $V^*(K(q^{i-1}, q^i), q^{i-1}) = 0$.

Condition C: For each $i = 1, 2, \dots, n$, $V^*(\tau^i - t, p_t) > 0$ for all $t \in T^i$.

Condition T is the usual indifference condition that requires the principal to terminate the project when the continuation payoff (when she deviates) is non-positive at each τ^i . This indifference condition alone is in general not sufficient because the instantaneous payoff may not be monotonically decreasing over time: in any segment T^i , the agent may start off with low effort, during which the instantaneous payoff could be so small that the principal is tempted to stop prematurely. Condition C assures that the principal does not stop before the intended termination date τ^i is reached.

The next result establishes that there exists a perfect Bayesian equilibrium that is always unique even under no commitment.

Proposition 4 *For any given set of parameters (Θ, p_0) satisfying Assumptions 1 and 2, there exists a (generically) unique pure-strategy equilibrium. Given Θ , the belief space $(q_h, 1)$ is partitioned into $m(\Theta) \geq 1$ distinct intervals $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ where $P^{m(\Theta)} = q_h$ and $P^0 = 1$. For $p_0 \in (P^j, P^{j-1})$, $j = 1, 2, \dots, m(\Theta)$, the equilibrium strategies are as follows:*

- The principal's equilibrium strategy is characterized by a set of $n = m(\Theta) - j + 1$ distinct termination dates $\{\tau^i\}_{i=1}^n$ where $q^i = P^{i+j-1}$ and $\tau^i - \tau^{i-1} > k^*$,²⁰
- The agent exerts high effort if the remaining time is less than or equal to k^* and low effort otherwise.

Moreover, if $m(\Theta) > 1$, then $q_l > P^1$.

²⁰As discussed in footnote 19, $a_t = 0$ for all $t \geq \tau^n$.

Proof: See Appendix A.

As the proposition suggests, the structure of equilibrium is thoroughly characterized by how the belief space is partitioned into intervals. A crucial determinant of the partition $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ is the profitability under low effort which is captured by π_l .²¹ Below, we first briefly illustrate how we pin down $m(\Theta)$ and $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ from a given Θ (in what follows, we simply denote $m = m(\Theta)$ to save notation). Once they are obtained, it is quite straightforward to derive the equilibrium strategy T for a given p_0 .

Suppose first that the profitability under low effort is sufficiently high, so that the principal's instantaneous payoff is positive even under low effort. In this case, the principal is not tempted to stop prematurely, and only Condition T is sufficient to pin down the equilibrium. Figure 1 depicts this situation where $m = 1$: the agent starts off with low effort and switches to high effort when the remaining time is k^* ; the principal stops when the belief reaches the lower bound $P^1 = q_h$ at time τ^1 . Formally, $m = 1$ and hence $n = 1$ for all $p_0 \in (q_h, 1)$ if and only if $V^*(K(q_l, q_h), q_l) > 0$, i.e.,

$$\frac{q_l \pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\nu}) + \frac{q_l e^{-(\lambda_l + r)\nu} \pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)(k - \nu)}) > \frac{(1 - p_t)w}{r} (1 - e^{-rk}), \quad (11)$$

where $\mu = \max\{K(q_l, q_h) - k^*, 0\}$ (see the proof of Proposition 4, especially Lemma 2, for more detail). Note that (11) depends only on Θ . The no-commitment solution $\tau^{NC}(p_0)$ is monotonic in p_0 if and only if this condition holds.

[Figure 1 about here]

In contrast, the instantaneous payoff may become negative under low effort when the success probability under low effort is relatively low, in which case we may have a situation where the principal's continuation payoff also becomes negative before the belief reaches q_h . Figures 2 and 3 show the evolution of the belief and the expected payoff when $m = 2$ while fixing the principal's strategy at $T = \{\tau^1\}$. In the figures, q_l is so high that the instantaneous payoff is negative for the entire interval during which the agent exerts low effort, and there exists a point τ' (in Figure 3) such that the continuation payoff is negative for $t \in (0, \tau')$. This implies that $T = \{\tau^1\}$ does not constitute an equilibrium as it violates Condition C. In this case, the game is divided into two segments, T^1 and T^2 , as illustrated in Figure 4. The pair of strategies now satisfies the equilibrium conditions since the instantaneous payoff is always positive in T^1 . Formally, if $V^*(K(q_l, q_h), q_l) < 0$, there must exist $p' \in (q_h, q_l)$ such

²¹Note that the partition $\{(P^j, P^{j-1})\}_{j=1}^{m(\Theta)}$ is determined solely by Θ while q^i depends also on the initial prior p_0 .

that $V^*(K(p', q_h), p') = 0$, in which case we redefine $P^1 = p'$ and $P^2 = q_h$. As above, $m = 2$ if and only if $V^*(K(q_l, P^1), q_l) > 0$.

[Figures 2-4 about here]

We can continue this process until we find P^1 such that $V^*(K(q_l, P^1), q_l) > 0$.²² Once the partition is pinned down, we can then easily derive the equilibrium strategies: for $p_0 \in (P^j, P^{j-1})$, the principal's equilibrium strategy consists of $n = m - i + 1$ termination dates where $q^i = P^{i+j-1}$ and

$$\tau^{NC} = \tau^1 = K(p_0, q^1), \quad \tau^i = \sum_{i'=1}^{i-1} \tau^{i'} + K(q^{i-1}, q^i) \text{ for } i = 2, 3, \dots, n.$$

Note also that $q^{i-1} > \phi(q^i) > q^i$ because the continuation payoff is always strictly positive when the belief is in $(q^i, \phi(q^i))$, the range where the remaining time is less than k^* and the agent exerts high effort. This alternatively means $\tau^i - \tau^{i-1} > k^*$, i.e., the length between any two adjacent termination dates must be larger than k^* .

Finally, we would also like to note that the constructed equilibrium is generically unique for any given set of parameters. This uniqueness result stems crucially from the fact that there is a lower bound of the belief q_h below which the principal would never continue the project. As stated above, together with the fact that $\lambda_l > 0$, this implies that the game must end in some finite time, which allows us to solve the game via backward induction analogously to the ‘‘gap case’’ of the durable-good monopoly problem (Fudenberg et al., 1985). More precisely, since the agent's strategy depends only on the remaining time, we know exactly how the game must end as the belief approaches the lower bound. Applying this reasoning backward, we can identify a unique continuation equilibrium for each $p_t > q_h$ and all the way back to the initial prior p_0 .

5 Discussion

5.1 The value of commitment

Our framework yields unique commitment and no-commitment solutions which enable us to directly assess the value of commitment to an evaluation deadline. Given that both τ^C and τ^{NC} can be written as functions of p_0 , the expected equilibrium payoffs can also be written as functions of p_0 . Let

$$V^C(p_0) := V^*(\tau^C(p_0), p_0), \quad V^{NC}(p_0) := V^*(\tau^{NC}(p_0), p_0).$$

²²When $q_l = 1$, there is no such q^1 and m will go to infinity. If $q_l < 1$, on the other hand, this process must converge after a finite number of rounds.

The value of commitment is then defined as

$$\Delta V(p_0) := V^C(p_0) - V^{NC}(p_0).$$

In principle, we can compute the exact value of $\Delta V(p_0)$ for any given p_0 . The following statement is the main result of the paper which analytically characterizes how the value of commitment varies with respect to p_0 .

Proposition 5 (i) $\Delta V(p_0) = 0$ for $p_0 \in (P^m, \phi(P^m))$ where $P^m = q_h$. (ii) If $m > 1$, for each $j = 2, 3, \dots, m$, ΔV is strictly increasing in p_0 for $p_0 \in (\phi(P^j), P^{j-1})$ and strictly decreasing for $p_0 \in (P^{j-1}, \phi(P^{j-1}))$ with $\Delta V(\phi(P^j)) = 0$. (iii) $\Delta V(\phi(P^1)) = 0$ if and only if $\hat{p} \geq \phi(P^1)$ where \hat{p} is the threshold defined in Proposition 3. (iv) $\lim_{p_0 \rightarrow 1} \Delta V(p_0) = 0$ if and only if $1 > \hat{p}$.

Proof: See Appendix.

The proposition suggests that the value of commitment changes in a non-monotonic way as p_0 increases, precisely due to the non-monotonic nature of V^{NC} . Despite this, we can still find some regularities except for the last segment $(\phi(P^1), 1)$.²³ Take an interval $(\phi(P^j), \phi(P^{j-1}))$ for $j = 2, 3, \dots, m$ where $\phi(P^{j-1}) > P^{j-1} > \phi(P^j)$. In each of these intervals, (i) the value of commitment is single-peaked and maximized at P^{j-1} (at which point the no-commitment solution yields a zero payoff); (ii) the value of commitment is zero at both ends of the interval (at which point the two solutions are identical). Important features of Proposition 5 are captured by Figure 5 which shows a typical path of the value of commitment when $m = 2$.²⁴

[Figure 5 about here]

To see how ΔV changes with respect to p_0 , observe that the only feasible history at any continuation game is the one consisting only of failures up to that point, and the principal hence has no additional information other than the fact that the agent has achieved no success. This immediately implies that the no-commitment solution can yield no higher payoff than the commitment solution, i.e., $\Delta V(p_0) \geq 0$ for all p_0 . Clearly, the no-commitment solution

²³In the last segment, the profitability under low effort can be high enough (if $p_0 > q_l$), and the expected payoff can go up even during the low-effort phase. Moreover, in this range, $\tau^C(\phi(P^1)) = \tau^{NC}(\phi(P^1))$ may not even be satisfied since $\tau^C(\phi(P^1))$ can be larger than k^* . As a consequence, the value of commitment can either go up or down (and can even be non-monotonic). See Appendix B for more detail.

²⁴For the figure, the parameters are set as follows: $\lambda_h = 0.1$, $\lambda_l = 0.04$, $b = y = 1$, $w = 0.035$, $d = 0.02$, $r = 0.05$.

gives the principal a different, and necessarily lower, payoff when $\tau^C \neq \tau^{NC}$. Still, the no-commitment solution can replicate the same allocation as the commitment solution when $\tau^C = \tau^{NC}$, because Proposition 1 ensures the unique equilibrium given the same horizon. It follows from Proposition 5 that the commitment solution generically yields a strictly higher expected payoff for the principal than the no-commitment solution if and only if $p_0 > \phi(q_h)$,²⁵ i.e.,

$$p_0 > \frac{q_h}{q_h + (1 - q_h)e^{-\lambda_h k^*}} \Leftrightarrow \frac{1 - q_h}{q_h} \left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right)^{\frac{\lambda_h}{\lambda_h + r}} > \frac{1 - p_0}{p_0}. \quad (12)$$

Conversely, when this condition fails to hold, the absence of commitment entails no additional cost, and the no-commitment solution becomes strictly profitable in the presence of a (possibly very small) commitment cost.²⁶

The proposition states that the principal benefits from committing to a deadline if the initial prior belief p_0 exceeds a certain threshold $\phi(q_h)$, i.e., if the initial prior is so high that the belief at time k^* is above the lower bound q_h even if the agent exerts high effort for $t \in [0, k^*]$. The value of a deadline depends crucially on each player's propensity to "procrastinate." In short, the threshold is lower and hence favors the commitment solution for a given initial belief either when the agent delays exerting high effort for a given deadline or when the principal delays terminating the project for a given effort sequence. Below, we briefly summarize how each player's propensity to procrastinate is determined:

1. The agent's propensity to procrastinate is determined by the tradeoff between the current gain of attaining a success and the potential loss of future payoffs. For a given deadline, the agent tends to procrastinate more (a small k^*) when the potential loss of future payoffs is large relative to the current gain.
2. The principal's propensity to procrastinate depends on the break-even belief and the likelihood of a success under high effort. The break-even belief is evidently the major force determining how patient the principal can be, where the principal is more tempted to wait for an eventual success when q_h is low. Further, for a given q_h , the principal tends to wait longer when λ_h is small, because the information about the agent's type is revealed only slowly in that case.

²⁵To be more precise, as Proposition 5 suggests, there could be some non-generic cases where the expected payoffs are identical.

²⁶An obvious commitment cost is that the principal must give up flexibility to adjust *ex post* to any uncertainty that may resolve during the course of play. Although we have thus far assumed away this aspect for clarity, we extend the analysis to incorporate uncertainty into the current setup in section 6: there, we show that the commitment solution is not always weakly optimal. The commitment cost may also arise from the cost of writing and enforcing a formal contract.

Before we move on, it is important to note that our characterization result owes largely to the fact that we consider a binary effort choice or, more importantly, that there is an upperbound for effort. In contrast, if the effort domain is unbounded with a convex effort cost, the commitment solution in general yields a strictly higher payoff than the no-commitment solution.²⁷ Even in this case, however, the nature of the problem remains largely intact. The intuition provided above carries over to this more general case as well because the value of a deadline is still determined in essentially the same way, although the analysis would be substantially more complicated.

5.2 Too early or too late?

In the absence of commitment, the principal often fails to stop at the right time because she has a strong incentive to wait for a success. This reasoning seems to indicate that the principal always terminates the project earlier with commitment than without, or equivalently, $\tau^{NC}(p_0) \geq \tau^C(p_0)$ for all p_0 . As indicated in Figures 3 and 4, however, this does not necessarily hold in the current setup. When the principal knows that she cannot stop soon enough, and it is too costly for her, she may terminate the project even when the instantaneous payoff is still strictly positive.

From Proposition 3, it is conceptually straightforward to compute the commitment solution τ^C . Figures 6a and 6b illustrate τ^C as a function of p_0 , using the same parameters as in Figure 5 except for λ_l . The qualitative nature of τ^C depends largely on the productivity under low effort as captured by λ_l . Figure 6a is the case where λ_l is relatively high. In this case, the principal does not lose much during the low-effort phase, and is willing to extend beyond k^* when the initial prior is sufficiently high. In contrast, when λ_l is relatively low, as depicted in Figure 6b, the principal sets $\tau^C \leq k^*$ to assure that the agent always exerts high effort. In either case, for $p_0 > \phi(q_h)$, τ^C is set strictly below the upper bound $K(p_0, q_h)$ so that the project is terminated before the belief reaches q_h .

[Figures 6a and b about here]

As we show in the proof of Proposition 4, we can also explicitly derive the no-commitment solution $\tau^{NC}(p_0)$ although it may take a much more complicated form. Figure 7 depicts $\tau^{NC}(p_0)$ along with $\tau^C(p_0)$, which confirms that the average duration of the project can be longer with commitment than without, using the same parameters as in Figure 5. The figure also reveals that the no-commitment may not be monotonically increasing in p_0 , in

²⁷On the other hand, we can obtain essentially the same result even if the effort level is continuous as long as the feasible effort level is bounded from above.

clear contrast to the commitment solution which is always weakly increasing. When the initial belief is relatively high, it takes long for the belief to reach the lowerbound q_h , and the principal may have to incur large losses if she is to wait until the moment at which the instantaneous payoff equals zero. To avoid this situation and discipline the agent, the principal must act early, i.e., she may need to pull the trigger earlier for *ex ante* more promising agents in some cases. This is more likely to be the case when λ_l is low, in which case the principal must incur losses in the low-effort phase. Formally, we can make the following statement.

Proposition 6 $\tau^{NC}(p_0)$ can be larger or smaller than $\tau^C(p_0)$. Moreover, $\tau^C(p_0) > \tau^{NC}(p_0)$ for almost all $p_0 \in (\phi(q_h), 1)$ as λ_l tends to zero.

Proof: It is clear that $\tau^{NC}(p_0)$ can be larger than $\tau^C(p_0)$ for some p_0 , because $\tau^{NC}(p_0) > k^*$ for p_0 slightly larger than $\phi(q_h)$ while $\tau^C(p_0) = k^*$ for $p_0 \in (\phi(q_h), \hat{p})$. We can also show that $\tau^{NC}(p_0)$ can be smaller if $m \geq 2$, because $\lim_{p_0 \downarrow P^1} \tau^{NC}(p_0) = 0 < k^* \leq \tau^C(P^1)$.

As for the second statement, It follows from Proposition 3 that $\tau^C(p_0) \geq k^*$ for $p_0 \in (\phi(q_h), 1)$. To prove the proposition, it thus suffices to show that $k^* > \tau^{NC}(p_0)$ in this range. Suppose otherwise, i.e., $\tau^{NC}(p_0) \geq k^*$ for some $p_0 \in (\phi(q_h), 1)$. This happens if and only if $p_0 \in (\phi(P^j), P^{j-1})$ for $j = 1, 2, \dots, m$ or, in other words, P^{j-1} must be bounded away from $\phi(P^j)$. This is a contradiction, however, as $\lim_{\lambda_l \rightarrow 0} V^*(K(P^{j-1}, P^j), P^{j-1}) < 0$ for any $P^{j-1} > \phi(P^j)$. ■

[Figure 7 about here]

Our results yield some empirical implications. An obvious one is that the initial prior p_0 is not necessarily a reliable predictor of future success under no commitment. Aside from this, it is also worth emphasizing that the equilibrium allocation depends crucially on what happens off the equilibrium path. As a consequence, two organizations that are observationally similar may exhibit drastically different equilibrium outcomes. To put this idea in context, consider two distinct organizations, A and B , each characterized by λ_l^i , $i = A, B$, and the corresponding no-commitment solution $\tau^{NC,i}$ (while assuming that the two organizations are equivalent in every other dimension). Suppose further that $\lambda_l^A > \lambda_l^B$ and $\lambda_l^B \rightarrow 0$. In this case, as Proposition 6 suggests, $\tau^{NC,B}(p_0) < k^*$ for almost all p_0 , meaning that we almost never observe the low-effort phase in organization B , and hence λ_l^B cannot be estimated from actual data. The two organizations are thus observationally indistinguishable even though their respective no-commitment solutions can be totally different from each other.

5.3 An application: when is up-or-out optimal?

The most prominent example of evaluation schemes with deadlines is arguably what is called the “up-or-out system” – a promotion scheme that is widely observed in academia and professional service industries such as law, accounting, and consulting (Lazear and Gibbs, 2014; Eriksson, 2016).²⁸ Under a typical up-or-out contract, the employer sets a deadline by which a promotion must occur, and the worker must leave the firm if he is not promoted by the deadline. In academia, it is often the case that incoming assistant professors must be promoted within a certain period of time (typically around six years). Outside of academia, a leading example of industries with up-or-out is the legal service industry. Many law firms traditionally adopt a set of managerial practices, known as the Cravath system, which include up-or-out (or “partnership track”) where incoming associates must make partner within a certain period of time (roughly seven to ten years). This contrasts sharply with other typical firm organizations which almost never specify deadlines by which workers must be promoted.

Here, we attempt to shed some light on this issue by examining under what conditions (12) is likely to be satisfied. There are of course some caveats. First and foremost, our theoretical framework is perhaps too simplistic to capture all the important details of this diverse set of industries; some of our arguments may thus be applied only to a subset of those industries.²⁹ Second, there is also a danger in comparing allocations under different incentives structures. One particular concern is the potential endogeneity of parameter values that are taken as exogenous here.³⁰ With those qualifications in mind, our framework can still provide some useful, though preliminary, insight for when it pays to commit to a promotion deadline as in the case of up-or-out.

Ability intensity: It is often argued that knowledge intensity is one of the most fundamental characteristics of professional service industries (von Nordenflycht, 2010). Combined with the fact that they are also less capital intensive,³¹ the productivity of an organization depends

²⁸For brevity, we refer to those industries characterized by up-or-out, including academia, broadly as professional service industries. The military is another example which is characterized by up-or-out, but we do not consider this case because the underlying structure appears to be quite different. See Kahn and Huberman (1988), Waldman (1990), O’Flaherty and Siow (1992), and Ghosh and Waldman (2010) for formal analyses of up-or-out contracts.

²⁹For instance, our exponential specification implies that success is very rare and information is coarse, as assumed in Bonatti and Hörner (2017), which might be applicable to some industries but not to some others. Also, team production may play a bigger role in some industries than in others.

³⁰It is likely that y and b may vary with the contract duration when we interpret them as the net continuation payoffs of achieving a success. In particular, we normalize the continuation payoff at zero when the project is terminated, but this could well be a function of the contract duration (or the market belief at the time of termination).

³¹High knowledge intensity and low capital intensity are not equivalent in the strict sense of the word because, as argued by von Nordenflycht (2010), an industry can be both knowledge- and capital-intensive at the same time. For the purpose of this study, however, we do not make any distinction between them.

crucially on the extent of knowledge embodied in individuals (Starbuck, 1992; Winch and Schneider, 1993). A likely consequence of this fact is that innate ability matters and creates value that cannot be easily substituted by sheer effort, at least in the short run – a feature which we call ability intensity for expositional clarity.

Since the mediocre type can never attain a success in the current setup, one way to measure the extent of ability intensity is by the (good type’s) success probabilities. We say that the project is more ability-intensive when λ_l and λ_h are larger with a fixed Δ_λ . Note that an increase in λ_h , while fixing Δ_λ , yields three conflicting effects. First, it raises the agent’s productivity at the margin, which in turn lowers the threshold q_h . Second, it gives the agent a stronger incentive to procrastinate (a lower k^*) as it raises his expected future payoff under high effort. Finally, it also facilitates the principal’s learning as each failure reveals more information. The first two effects raises the value of up-or-out while the last lowers it, such that the overall impact is not immediately clear.

One can readily show, however, that there exists some threshold $\bar{\lambda}$ such that it is strictly optimal to set a deadline, thereby favoring the use of up-or-out, for $\lambda_h > \bar{\lambda}$. To see this, we rewrite (12) as

$$\frac{p_0}{1 - p_0} > \frac{q_h}{(1 - q_h)e^{-\lambda_h k^*}} = \frac{w}{(\lambda_h y - w)e^{-\lambda_h k^*}}. \quad (13)$$

Under the maintained assumptions, $\lambda_h > \underline{\lambda} := \max\{\frac{w}{p_0 y}, \frac{\Delta_\lambda}{d}(r(b - \frac{d}{\Delta_\lambda}) + d - w)\}$. If $\frac{\Delta_\lambda}{d}(r(b - \frac{d}{\Delta_\lambda}) + d - w) > \frac{w}{p_0 y}$, we have $\lim_{\lambda_h \rightarrow \underline{\lambda}} k^* = \infty$, and hence, (13) is never satisfied if λ_h is small. In contrast, we have

$$\lim_{\lambda_h \rightarrow \infty} \frac{q_h}{(1 - q_h)e^{-\lambda_h k^*}} = 0,$$

implying that (13) must hold if λ_h is sufficiently large.

Stable job descriptions: Another distinctive feature of professional service industries is a professionalized workforce that builds on a particular knowledge base (Torres, 1991; von Nordenflycht, 2010). This in turn creates well-defined job boundaries and stable job descriptions. Due to various professional requirements and accreditation processes, workers in professional service industries are typically responsible only for a narrow and clear set of tasks, compared to workers in other industries. This implies that the nature of tasks that they are expected to carry out is very stable over time, and the production environment is relatively immune to stochastic shocks. For instance, a demand or technology shock that entirely changes the job description of a lawyer or a college professor is highly unlikely.

We can interpret r as the rate at which the project is terminated for exogenous reasons,

e.g., the arrival of a stochastic shock that makes the project completely worthless.³² Define \bar{r} such that

$$\frac{\lambda_h b - d + w}{\lambda_h + \bar{r}} = b - \frac{d}{\Delta_\lambda}.$$

Assumption 1 then implies that r must be bounded between $(0, \bar{r})$. It is clearly not profitable to commit to a deadline if r is sufficiently close to \bar{r} because $\lim_{r \rightarrow \bar{r}} k^* = \infty$. In contrast, there exists some \underline{r} such that it is profitable to commit to a deadline for $r < \underline{r}$ if

$$\frac{1 - q_h}{q_h} \left(1 - \frac{\lambda_1 (b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right) > \frac{1 - p_0}{p_0}.$$

It can be seen from this argument that a decrease in r (a more stable relationship) in general raises the value of up-or-out because the agent can be more forward-looking and tends to procrastinate more, which in turn makes the principal unable to stop at the right time.

Job similarity across ranks: Another immediate consequence of stable job descriptions is that jobs necessarily become similar across different ranks.³³ In the current setup, this aspect may be captured by y , which can be interpreted as the expected payoff of promoting a good-type agent. If entry- and senior-level jobs are similar, y tends to be larger because a success at the entry level is a reliable predictor of productivity at the senior level. This is not necessarily the case in typical firm organizations where jobs across ranks can differ to a considerable extent, as often discussed in the context of the Peter Principle.³⁴

The effect of a change in y is fairly straightforward, as it only affects the break-even belief with no impact on the agent's behavior. A large value of y implies a small value of q_h which gives the principal a stronger incentive to wait for a breakthrough. Given this, the agent also has a stronger incentive to procrastinate, hoping to achieve a breakthrough with low effort. The principal can then unambiguously benefit from committing to a deadline when y is sufficiently large, implying that the value of up-or-out is higher in industries where jobs are similar and there is a strong correlation between performances at different ranks.

³²Here, we allow the principal to terminate the project when the shock arrives; see section 6 for the case where the principal must abide by the deadline under any circumstance.

³³Ghosh and Waldman (2010) also raises job similarity as one of the distinctive characteristics of academia. Job similarity is also crucial in Kahn and Huberman's (1988) classic argument because if jobs are sufficiently different, we may use promotions to solve the double moral-hazard problem as suggested by Prendergast (1993). However, our argument here may be more applicable to academia (at least in humanities and social sciences) but less so to other professional service industries such as law where the role of a partner is more about bringing in business (or "rainmaking").

³⁴The Peter Principle states that every post tends to be occupied by an employee who is incompetent to carry out its duties, because employees are promoted through positions where they have excelled until they reach a level of incompetence (Peter and Hull, 1969). A premise of this argument is that jobs are inherently different and become progressively harder as one climbs through ranks.

6 Conclusion

This paper presents a simple stylized model of dynamic performance evaluation with particular emphasis on the role of commitment to an evaluation deadline. We consider an environment where the principal chooses whether to commit to a deadline and if so, at what time. Within this framework, we obtain a complete characterization of equilibrium and derive a necessary and sufficient condition for committing to a deadline to be optimal. The simple framework also allows us to conduct various comparative statics exercises to give insight into when it is beneficial to set an evaluation deadline at the outset.

Since our model is deliberately stylized to obtain sharp analytical characterizations, there are naturally several avenues to extend the current analysis. Before we conclude, we briefly discuss two of those possible extensions (see Appendices C and D for more formal analyses of these cases).

Uncertainty: An important cost of making commitment arises from the fact that the principal loses flexibility to adjust to future stochastic shocks. The baseline model does not capture this cost because the only feasible history at each continuation game consists only of failures up to that point; as a consequence, it is weakly optimal to commit to a deadline in the baseline model. This is apparently unsatisfactory, provided that we live in a world filled with uncertainty where there is clearly value in being flexible.

One way to cope with this possibility is to consider a setup where a permanent productivity shock may strike with some probability, which totally changes the nature of the task and subsequently makes the agent unproductive, i.e., $\lambda_h = \lambda_l = 0$.³⁵ For simplicity, suppose that (i) the shock is permanent and arrives at most once, with a Poisson arrival rate β , and (ii) the arrival of the shock is publicly observable. In this situation, the principal can retain the flexibility to adjust to a negative shock by immediately terminating the project whenever it strikes. We can then show that due to this benefit, there arises a situation where the principal strictly benefits from not committing to a deadline.

Exogenous wages: In the baseline model, we have assumed that the contractual environment is highly incomplete in that no contingent wage contracts can be written. Even in such an environment, it may still be possible to agree on the flow wage contingent on the continuation of the project. Of course, since the flow wage in this context is a pure transfer

³⁵Once the principal chooses to commit to a deadline, the principal must abide by it and is not allowed to terminate the project, unlike in the discussion of stable job descriptions in section 5.3 where we allow the principal to terminate the project whenever a stochastic shock arrives. If the principal is allowed to terminate the project at any time before the committed deadline, the commitment solution then yields a weakly higher payoff than the no-commitment solution.

payment with no incentive effect on the agent's side, there is no reason for the principal to offer any wage beyond the minimum level if she can stop at the right time. However, as we have seen, this is not always true when the principal makes no commitment to a deadline, in which case the flow wage can be used as a substitute for a deadline. This can be potentially useful if the commitment cost is prohibitively large for some reason.

To explore whether there are circumstances in which she offers any strictly positive w to this effect, suppose that the principal can deliberately choose any $w \geq 0$ at time 0.³⁶ The benefit of raising w above zero is clear if we look at the case where $w = 0$: in this case, the instantaneous payoff is always strictly positive (as in the cooperative case), no matter how unlikely the agent is to succeed; as a consequence, the principal can never terminate the project, and given this, the agent never exerts high effort. The principal can alter this structure by raising w above zero because the flow cost of employment is now positive, rendering the instantaneous payoff negative at some point. This can be profit-enhancing for the principal because with the credible threat of termination, she can induce the agent to exert high effort, which is especially beneficial when the success probability under low effort is low: formally, we can show that the optimal flow wage is strictly positive if λ_l is sufficiently small. This reasoning is similar to Buisseret (2016) in that a higher flow wage is used as a commitment device to stop at the right time and reduce future rents for the agent.³⁷

References

- Bergemann, D. and Hege, U., 1998, Venture Capital Financing, Learning, and Moral Hazard, *Journal of Banking and Finance*, 22, 703-35.
- Bergemann, D. and Hege, U., 2005, The Financing of Innovation: Learning and Stopping, *RAND Journal of Economics*, 36, 719-52.
- Bergemann, D. and Välimäki, J., 1996, Learning and Strategic Pricing, *Econometrica*, 64, 1125-49.
- Bergemann, D. and Välimäki, J., 2000, Experimentation in Markets, *Review of Economic Studies*, 67, 213-34.

³⁶In other words, we assume that the minimum wage level is zero. On the other hand, if the principal can set a negative w , she can immediately screen out the mediocre type by offering $w < 0$. This means that we focus on the situation where the minimum wage is so high that even the mediocre type can earn a positive rent by being employed.

³⁷Buisseret (2016) considers a two-period model where the principal can raise the success probability by investing in the project and shows that the principal may over-invest, so that she can credibly terminate the project upon the first failure. This can be profit-enhancing as it denies the agent the continuation payoff.

- Bergemann, D. and Välimäki, J., 2008, Bandit Problems, in Durlauf, S. and Blume, L., eds, *The New Palgrave Dictionary of Economics*, 2nd edition, Macmillan Press.
- Bolton, P. and Harris, C., 1999, Strategic Experimentation, *Econometrica*, 67, 349-74.
- Bonatti, A. and Hörner, J., 2011, Collaborating, *American Economic Review*, 101, 632-63.
- Bonatti, A. and Hörner, J., 2017, Career Concerns with Exponential Learning, *Theoretical Economics*, 12, 425-75.
- Buisseret, P., 2016, Investing in Failure: Belief Manipulation and Strategic Experimentation, mimeo.
- Campbell, A., Ederer, F. and Spinnewijn, J., 2014, Delay and Deadlines: Freeriding and Information Revelation in Partnerships, *American Economic Journal: Microeconomics*, 6, 163-204.
- Chen, C.-H., 2012, Name Your Own Price at Priceline.com: Strategic Bidding and Lockout Periods, *Review of Economic Studies*, 79, 1341-69.
- Chen, C.-H. and Ishida, J., 2015, Hierarchical Experimentation, ISER Discussion Paper No. 949.
- Damiano, E., Li, H. and Suen, W., 2012, Optimal Deadlines for Agreements, *Theoretical Economics*, 7, 357-93.
- Eriksson, T., 2016, Inter- and Intra-Firm Mobility of Workers, in Saridakis, G. and Cooper, C.L., eds, *Research Handbook on Employee Turnover*, Edward Elgar.
- Fershtman, C. and Seidmann, D.J., 1993, Deadline Effects and Inefficient Delay in Bargaining with Endogenous Commitment, *Journal of Economic Theory*, 60, 306-21.
- Fuchs, W. and Skrzypacz, A., 2013, Bargaining with Deadlines and Private Information, *American Economic Journal: Microeconomics*, 5, 219-43.
- Fudenberg, D. and Tirole, J., 1991, Perfect Bayesian Equilibrium and Sequential Equilibrium, *Journal of Economic Theory*, 53, 236-260.
- Fudenberg, D., Levine, D. and Tirole, J., 1985, Infinite Horizon Models of Bargaining with One-Sided Incomplete Information, in Roth, A., ed, *Game Theoretic Models of Bargaining*, Cambridge University Press.
- Garfagnini, U., 2011, Delegated Experimentation, mimeo.
- Georgiadis, G., Lippman, S.A. and Tang, C.S., 2014, Project Design with Limited Commitment and Teams, *RAND Journal of Economics*, 45, 598-623.

- Gerardi, D. and Maestri, L., 2012, A Principal-Agent Model of Sequential Testing, *Theoretical Economics*, 7, 425-63.
- Ghosh, S. and Waldman, M., 2010, Standard Promotion Practices Versus Up-or-Out Contracts, *RAND Journal of Economics*, 41, 301-25.
- Guo, Y., 2016, Dynamic Delegation of Experimentation, *American Economic Review*, 106, 1969-2008.
- Halac, M., Kartik, N. and Liu, Q., 2016, Optimal Contracts for Experimentation, *Review of Economic Studies*, 83, 1040-91.
- Hendricks, K., Weiss, A. and Wilson, C., 1988, The War of Attrition in Continuous Time with Complete Information, *International Economic Review*, 29, 663-80.
- Holmström, B., 1999, Managerial Incentive Problems: A Dynamic Perspective, *Review of Economic Studies*, 66, 169-82.
- Hörner, J. and Samuelson, L., 2011, Managing Strategic Buyers, *Journal of Political Economy*, 119, 379-425.
- Hörner, J. and Samuelson, L., 2013, Incentives for Experimenting Agents, *RAND Journal of Economics*, 44, 632-63.
- Hörner, J. and Samuelson, L., 2016, Dynamic Moral Hazard without Commitment, *International Journal of Game Theory*, 45, 89-136.
- Kahn, C. and Huberman, G., 1988, Two-Sided Uncertainty and “Up-or-Out” Contracts, *Journal of Labor Economics*, 6, 423-44.
- Keller G., Rady, S. and Cripps, M., 2005, Strategic Experimentation with Exponential Bandits, *Econometrica*, 73, 39-68.
- Klein, N.A. and Rady, S., 2011, Negatively Correlated Bandits, *Review of Economic Studies*, 78, 693-732.
- Lazear, E.P., and Gibbs, M., 2014. *Personnel Economics in Practice*. 3rd edition, Wiley.
- Manso, G., 2011, Motivating Innovation, *Journal of Finance*, 66, 1823-69.
- Mason, R. and Välimäki, J., 2015, Getting It Done: Dynamic Incentives to Complete a Project, *Journal of the European Economic Association*, 13, 62-97.
- O’Flaherty, B. and Siow, A., 1992, On the Job Screening, Up or Out Rules, and Firm Growth, *Canadian Journal of Economics*, 25, 346-68.

- Peter, L.J. and Hull, R., 1969, *The Peter Principle: Why Things Always Go Wrong*, Morrow.
- Prendergast, C., 1993, The Role of Promotion in Inducing Specific Human Capital Acquisition, *Quarterly Journal of Economics*, 108, 523-34.
- Rothschild, M., 1997, A Two-Armed Bandit Theory of Market Pricing, *Journal of Economic Theory*, 9, 185-202.
- Spier, K.E., 1992, The Dynamics of Pretrial Negotiation, *Review of Economic Studies*, 59, 93-108.
- Starbuck, W., 1992, Learning by Knowledge-Intensive Firms, *Journal of Management Studies*, 29, 713-40.
- Torres, D., 1991, What, If Anthing, is Professionalism? Institutions and the Problem of Change, *Research in the Sociology of Organizations: Organizations and Occupations*, 8, 43-68.
- von Bordenflycht, A., 2010, What is a Professional Service Firm? Towards a Theory and Taxonomy of Knowledge Intensive Firms, *Academy of Management Review*, 35, 155-74.
- Waldman, M., 1990, Up-or-Out Contracts: A Signaling Perspective, *Journal of Labor Economics*, 8, 230-50.
- Winch, G. and Schneider, E., 1993, Managing the Knowledge-Based Organization: The Case of Architectural Practice, *Journal of Management Studies*, 30, 923-37.

Appendix A: Proofs

Proof of Proposition 1: It is intuitively clear that the value function is strictly increasing in k . Formally, define

$$rU_a(k) = \left((\lambda_a b - d_a + w) - \lambda_a U_a(k) - \dot{U}_a(k) \right) \quad (14)$$

as the value function when the effort level is fixed at $a = l, h$. Solving this differential equation and imposing the boundary condition $U(0) = 0$, we derive

$$U_h(k) = \frac{\lambda_h b - d + w}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k}).$$

If $\Delta_\lambda b > d$, it is optimal for the agent to exert high effort as long as $U_h(k)$ is sufficiently small. Moreover, since $U_h(k)$ is strictly increasing in k , the agent would never switch back

to low effort, once $U(k)$ reaches this critical point. Then, k^* is obtained as a solution to $U_h(k^*) = b - \frac{d}{\Delta_\lambda}$, which can be written as

$$\frac{\lambda_h b - d + w}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k^*}) = b - \frac{d}{\Delta_\lambda}. \quad (15)$$

Note that the agent always exerts high effort if

$$b - \frac{d}{\Delta_\lambda} \geq \frac{\lambda_h b - d + w}{\lambda_h + r} \Leftrightarrow \frac{rb - w}{\lambda_l + r} \geq \frac{d}{\Delta_\lambda},$$

while he never does so if $\frac{d}{\Delta_\lambda} \geq b$. For in-between cases, solving (15) yields

$$k^* = -\frac{1}{\lambda_h + r} \ln \left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right),$$

so that the agent exerts high effort when the remaining time is less than or equal to k^* . For $k > k^*$, the agent exerts low effort because $U(k) \geq U_h(k) > b - \frac{d}{\Delta_\lambda}$. ■

Proof of Proposition 2: The agent always exerts high effort in the noncooperative case if (2) fails to hold. An analogous condition for the cooperative case is given by

$$\frac{\lambda_h(y + b) - d}{\lambda_h + r} > y + b - \frac{d}{\Delta_\lambda} \Leftrightarrow \frac{d}{\Delta_\lambda} > \frac{r}{\lambda_l + r}(y + b).$$

Applying the same argument as in Proposition 1, we obtain

$$\tilde{k}^* = -\frac{1}{\lambda_h + r} \ln \left(1 - \frac{(\lambda_h + r)(y + b - \frac{d}{\Delta_\lambda})}{\lambda_h(y + b) - d} \right).$$

To see that the project is never terminated, we first show that if it is ever optimal to set a deadline in some finite time, the only possibility is to set it at time \tilde{k}^* , so that the agent exerts high effort from the beginning. To see this, the principal extends the deadline beyond time \tilde{k}^* if and only if

$$\lambda_l(y + b) > \frac{\lambda_l + r}{\lambda_h + r} (\lambda_h(y + b) - d) (1 - e^{-(\lambda_h + r)\tilde{k}^*}). \quad (16)$$

Note that this condition is independent of the deadline. As such, if this condition holds, it is then optimal to continue the project indefinitely until the project succeeds. If it fails to hold, on the other hand, it is clearly optimal to set a deadline at time \tilde{k}^* .

Since, by definition,

$$y + b - \frac{d}{\Delta_\lambda} = \tilde{U}(\tilde{k}^*) = \frac{(\lambda_h(y + b) - d)}{\lambda_h + r} (1 - e^{-(\lambda_h + r)\tilde{k}^*}).$$

we can modify (16) as

$$\frac{\lambda_l(y+b)}{\lambda_l+r} > y+b - \frac{d}{\Delta_\lambda}.$$

The left-hand side is the expected payoff of exerting low effort indefinitely while the right-hand side is that of exerting high effort from time 0 to time \tilde{k}^* which must equal the static gain.

This condition must hold for any finite \tilde{k}^* when (6) is satisfied, meaning that it is strictly better to let the project continue indefinitely. If (6) fails to hold, on the other hand, $\tilde{k}^* \rightarrow \infty$, and again it is better to let the project continue indefinitely. This shows that no finite deadline can improve the social surplus in the cooperative case. \blacksquare

Proof of Proposition 3: First, if $\phi(q_h) \geq p_0 > q_h$ (the latter inequality holds by Assumption 2), there exists a unique deadline $\tau^C \in (0, k^*]$ that satisfies

$$\frac{p_0 e^{-\lambda_h \tau^C}}{1 - p_0 + p_0 e^{-\lambda_h \tau^C}} = q_h.$$

There is clearly no reason to wait beyond this point because the instantaneous payoff is strictly negative for any effort choice. It is also not optimal to stop before τ^C because the instantaneous payoff is still strictly positive. The optimal termination date is then τ^C and it is unique in this range.

Second, suppose that $p_0 > \phi(q_h)$, in which case the first-order condition is given by $M(\tau^C, p_0) = 0$. If $M(k^*, p_0) \leq 0$, the expected payoff is strictly decreasing in τ^C for $\tau^C > k^*$, and the optimal termination date is $\tau^C = k^*$. Alternatively, define $\hat{p} \in (0, 1)$ such that $M(k^*, \hat{p}) = 0$ if (10) holds and $\hat{p} = 1$ otherwise. Then, the optimal termination date is k^* for $p_0 \leq \hat{p}$. If $p_0 > \hat{p}$, on the other hand, there must exist a unique $\hat{\tau}(p_0)$ such that $M(\hat{\tau}(p_0), p_0) = 0$. Since the instantaneous payoff is strictly positive for $\hat{\tau}(p_0) > \tau^C$ and negative for $\tau^C > \hat{\tau}(p_0)$, the optimal termination date in this case is $\hat{\tau}(p_0)$.

Finally, to show that $\hat{p} > q_l$ if $1 > q_l$, there are two cases. If (10) holds (but $1 > q_l$), then $\hat{p} = 1 > q_l$. If (10) fails to hold, on the other hand, $M(k^*, p_0)$ is increasing in p_0 for any k^* . Given that $1 > q_l$, it suffices to show that $M(k^*, q_l) < 0$, i.e.,

$$(1 - q_l)w e^{-rk^*} > q_l \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h (1 - e^{-(\lambda_h + r)k^*}) \right).$$

This can be written as

$$-e^{rk^*} \left(q_l \frac{\lambda_l + r}{\lambda_h + r} \pi_h e^{-\lambda_h k^*} - (1 - q_l)w \right) > q_l \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h \right).$$

which is further reduced to

$$-e^{rk^*} (q_l \pi_l e^{-\lambda_h k^*} - (1 - q_l)w) > q_l (1 - e^{-(\lambda_h + r)k^*}) \left(\pi_l - \frac{\lambda_l + r}{\lambda_h + r} \pi_h \right).$$

Note that the left-hand side is positive by definition while the right-hand side is negative, so that this condition holds for any k^* . From the preceding argument, we can see that $\hat{p} = 1$ if and only if (10) holds. It is also evident that $\tau^C(p_0)$ is weakly increasing in p_0 . \blacksquare

Proof of Proposition 4: We show the existence and uniqueness of the equilibrium by construction. We start with a candidate strategy $T^1 = \{\tau^1\}$ and the corresponding best response such that

$$a_t = \begin{cases} l & \text{for } t \in [0, \max\{\tau^1 - k^*, 0\}), \\ h & \text{for } t \in [\max\{\tau^1 - k^*, 0\}, \tau^1]. \end{cases} \quad (17)$$

This pair of strategies constitutes an equilibrium if $V^*(K(p, q_h), p) > 0$ for all $p \in (q_h, p_0]$. Alternatively, we check for what interval of p_0 this condition is satisfied. To this end, the following result is useful.

Lemma 2 *Let $q_l > q$. Then, $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \in (q, 1)$ if and only if*

$$\lim_{p \uparrow q_l} V^*(K(p, q), p) > 0.$$

Proof: Since the necessity is evident by definition, we only prove the sufficiency part. Since the case with $q_l = 1$ can be proved analogously (by taking the limit), we suppose for now that $1 > q_l > q$.

First, if $p_0 \geq q_l$, $V^*(K(q_l, q), q_l) > 0$ directly implies $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \geq q_l$ because the instantaneous payoff is strictly positive for any effort choice when the belief is above q_l . Note also that $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \in (q, \phi(q_h)]$ since the high type always exerts high effort in this case. This implies that $V^*(K(p_0, q), p_0) > 0$ for all $p_0 \in (q, 1)$ if $\phi(q) \geq q_l$. In what follows, therefore, we assume $q_l > \phi(q)$.

Given this, it suffices to show that

$$V^*(K(q_l, q), q_l) > 0 \Rightarrow V^*(K(p_0, q), p_0) > 0 \text{ for all } p_0 \in (\phi(q), q_l).$$

Suppose on the contrary that $V^*(K(q_l, q), q_l) > 0$ but $V^*(K(p, q), p) \leq 0$ for some $p \in (\phi(q), q_l)$. If this is the case, there must exist some $z \in (\phi(q), q_l)$ such that $V^*(K(z, q), z) = 0$. Since

$$\begin{aligned} V^*(K(p, q), p) &= \frac{p(\lambda_l y - w)}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(K(p, q) - K(p', q))}) - \frac{(1 - p)w}{r} (1 - e^{-r(K(p, q) - K(p', q))}) \\ &\quad + (1 - p + p e^{-\lambda_l(K(p, q) - K(p', q))}) e^{-r(K(p, q) - K(p', q))} V^*(K(p', q), p'), \end{aligned}$$

for any $p > p' > \phi(q)$, due to the recursive structure of the expected payoff, we have

$$\begin{aligned} V^*(K(q_l, q), q_l) &= \frac{q_l(\lambda_l y - w)}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\Delta K}) - \frac{(1 - q_l)w}{r} (1 - e^{-r\Delta K}) \\ &\quad + (1 - q_l + q_l e^{-\lambda_l \Delta K}) e^{-r\Delta K} V^*(K(z, q), z) \\ &= \frac{q_l(\lambda_l y - w)}{\lambda_l + r} (1 - e^{-(\lambda_l + r)\Delta K}) - \frac{(1 - q_l)w}{r} (1 - e^{-r\Delta K}), \end{aligned}$$

where $\Delta K := K(q_l, q) - K(z, q)$. This is strictly negative by definition, and therefore a contradiction. \blacksquare

The lemma shows that we only need to check the continuation payoff at q_l to see whether a given strategy is sequentially rational. An immediate corollary of the lemma is that $V^*(K(p_0, q_h), p_0) > 0$ for all p_0 if $\phi(q_h) \geq q_l$. In this case, $m(\Theta) = 1$, $P^1 = q_h$ and $T = \{\tau^1\}$ constitutes an equilibrium for any p_0 .

We now suppose that $q_l > \phi(q_h)$ and check if the candidate strategy can satisfy the equilibrium conditions. From Lemma 2, this is the case if and only if $V^*(K(q_l, q_h), q_l) > 0$, i.e.,

$$\frac{q_l \pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(K - k^*)}) + \frac{q_l \pi_h}{\lambda_h + r} e^{-(\lambda_l + r)(K - k^*)} (1 - e^{-(\lambda_h + r)k^*}) > \frac{(1 - q_l)w}{r} (1 - e^{-rK}). \quad (18)$$

where $K = K(q_l, q_h)$. If (18) holds, the expected payoff is always positive for any p_0 , so we can apply the same argument as in Case 2 to show the existence and uniqueness of the equilibrium.

The situation becomes more complicated when (18) fails to hold, in which case the principal has an incentive to deviate and stop at some point, and the candidate strategy no longer constitutes an equilibrium. In this case, there instead exists a critical belief that satisfies (18) with equality, because $V^*(K(p, q_h), p)$ is continuous in p with $V^*(K(\phi(q_h), q_h), \phi(q_h)) > 0 > V^*(K(q_l, q_h), q_l)$. We now redefine $P^1 = \min\{q : V^*(K(q, q_h), q) = 0\}$ and $P^2 = q_h$. Given the critical belief, we can then consider another candidate strategy $T^2 = \{\tau^1, \tau^2\}$ for $p_0 > P^1$ where $p_{\tau^1} = q^1$, and the best response is given by

$$a_t = \begin{cases} l & \text{for } t \in [0, \max\{\tau^1 - k^*, 0\}), \\ h & \text{for } t \in [\max\{\tau^1 - k^*, 0\}, \tau^1), \\ l & \text{for } t \in (\tau^1, \tau^2 - k^*), \\ h & \text{for } t \in [\tau^2 - k^*, \tau^2]. \end{cases}$$

Note that $q^1 > \phi(q^2) > q^2 = q_h$ or $\tau^2 > \tau^1 + k^*$.

The game is now divided into two segments. First, we can show that the equilibrium is unique after time τ^1 (as done in Case 2). Before time τ^1 , we can apply the same procedure to

the truncated interval $[0, \tau^1]$ to see whether T^2 can be an equilibrium strategy. More precisely, we need to check if $V^*(K(p, P^1), P^1) > 0$ for all $p \in [p_0, \phi(P^1)]$. Applying Lemma 2, T^2 is an equilibrium strategy for all $p_0 \in (P^1, 1)$ if $V(K(q_l, P^1), q_l) > 0$. For $p_0 \in (P^1, 1)$, the principal's equilibrium strategy consists of two termination dates. Given that the continuation equilibrium after time τ^1 is unique, we can establish the uniqueness by applying the same argument.

If $V(K(q_l, P^1), q_l) \leq 0$, on the other hand, we can find yet another critical belief in $(\phi(P^1), q_l]$. We again redefine P^1 analogously and repeat the same process until we find P^1 such that $V(K(q_l, P^1), q_l) > 0$. In the end, this process would give us $m(\Theta)$ intervals $\{(P^j, P^{j-1})\}_{j=1}^m$ where the principal's equilibrium strategy consists of $n = m(\Theta) - j + 1$ termination dates if p_0 falls into (P^j, P^{j-1}) . Note also that since $V^*(K(P^1, P^2), P^1) = 0$ by definition, it must be that $q_l > P^1$ if $m(\Theta) > 1$. ■

Proof of Proposition 5: (i) For $p_0 \in (q_h, \phi(q_h))$, $\tau^C(p_0) = \tau^{NC}(p_0) = K(p_0, q_h)$. Since the agent's strategy depends only on the remaining time, the two cases yield the same expected payoff for the principal.

(ii) To show this, we first examine how V^C and V^{NC} vary with respect to p_0 for $p_0 \in (\phi(q_h), 1)$. It is evident that V^C must monotonically increase with the initial prior. To see this, note that

$$\frac{dV^C}{dp_0} = \frac{\partial V^*}{\partial p_0} + \frac{\partial V^*}{\partial \tau^C} \frac{d\tau^C}{dp_0}.$$

The last term is the indirect effect though an increase in τ^C which is invariably zero because either $\frac{\partial V^*}{\partial \tau^C} = 0$ or $\frac{d\tau^C}{dp_0} = 0$ must hold by the optimality condition. Since the direct effect $\frac{\partial V^*}{\partial p_0}$ is unambiguously positive, V^C is strictly increasing in p_0 . More precisely, we obtain

$$\begin{aligned} \frac{dV^C}{dp_0} &= \frac{\pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k^*}) + \frac{w}{r} (1 - e^{-rk^*}) \\ &= p_0 \pi_h - (1 - p_0)w - (p_0 \lambda_h + r)V^C(p_0) - (p_0 e^{-\lambda_h k^*} \pi_h - (1 - p_0)w)e^{-rk^*}, \end{aligned} \quad (19)$$

for $p_0 \in (\phi(q_h), \hat{p})$, and

$$\begin{aligned} \frac{dV^C}{dp_0} &= \frac{\pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(\tau^C - k^*)}) + \frac{e^{-(\lambda_l + r)(\tau^C - k^*)} \pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h + r)k^*}) + \frac{w}{r} (1 - e^{-r\tau^C}) \\ &= p_0 \pi_l \frac{\lambda_h + r}{\lambda_l + r} (1 - e^{-(\lambda_l + r)(\tau^C - k^*)}) + p_0 \pi_h e^{-(\lambda_l + r)(\tau^C - k^*)} - (1 - p_0)w \\ &\quad - (p_0 \lambda_h + r)V^C(p_0) - (p_0 \pi_h e^{-\lambda_l(\tau^C - k^*) - \lambda_h k^*} - (1 - p_0)w)e^{-r\tau^C}, \end{aligned} \quad (20)$$

for $p_0 \in (\hat{p}, 1)$.

In contrast, V^{NC} changes with p_0 in a non-monotonic way if $m > 1$. Consider $p_0 \in (\phi(P^j), \phi(P^{j-1}))$ for $j = 2, 3, \dots, m$. Since $V^{NC}(p_0) = V^*(K(p_0, q), p_0)$ where q is the closest termination belief, it follows that

$$\frac{dV^{NC}}{dp_0} = \frac{\partial V^*}{\partial p_0} + \frac{\partial V^*}{\partial k} \frac{\partial K}{\partial p_0} = \frac{\partial V^*}{\partial p_0} + \frac{1}{\lambda_l p_0 (1-p_0)} \frac{\partial V^*}{\partial k}.$$

First, for $p_0 \in (\phi(P^j), P^{j-1})$, we obtain

$$\begin{aligned} \frac{dV^{NC}}{dp_0} &= \frac{\pi_l}{\lambda_l + r} (1 - e^{-(\lambda_l+r)(K-k^*)}) + \frac{\pi_h e^{-(\lambda_l+r)(K-k^*)}}{\lambda_h + r} (1 - e^{-(\lambda_h+r)k^*}) + \frac{w}{r} (1 - e^{-rK}) \\ &\quad + \frac{e^{-(\lambda_l+r)(K-k^*)}}{\lambda_l(1-p_0)} \left(\frac{\pi_l}{\lambda_l + r} - \frac{\pi_h}{\lambda_h + r} (1 - e^{-(\lambda_l+r)k^*}) \right) - \frac{w e^{-rK}}{\lambda_l p_0} \\ &= p_0 \pi_l - (1-p_0)w - (p_0 \lambda_l + r) V^{NC}(p_0), \end{aligned} \quad (21)$$

which is negative because $q_l > P^1 > p_0$. For $p_0 \in (P^{j-1}, \phi(P^{j-1}))$, we obtain

$$\begin{aligned} \frac{dV^{NC}}{dp_0} &= \frac{\pi_h}{\lambda_h + r} (1 - e^{-(\lambda_h+r)K}) + \frac{w}{r} (1 - e^{-rK}) + \frac{e^{-rK}}{\lambda_h p_0 (1-p_0)} (p_0 \pi_h e^{-\lambda_h K} - (1-p_0)w) \\ &= p_0 \pi_h - (1-p_0)w - (p_0 \lambda_h + r) V^{NC}(p_0), \end{aligned} \quad (22)$$

which is positive for any $p_0 > q_h$.

We are now ready to see how ΔV changes with respect to p_0 . First, it is clear that ΔV is strictly increasing in p_0 for $p_0 \in (\phi(P^j), P^{j-1})$ because V^C is increasing while V^{NC} is decreasing. In contrast, for $p_0 \in (P^{j-1}, \phi(P^{j-1}))$, both V^C and V^{NC} are increasing in p_0 . For $\hat{p} > p_0$, it follows from (19) and (22) that

$$\frac{d\Delta V}{dp_0} < 0 \Leftrightarrow (p_0 \lambda_h + r)(V^C(p_0) - V^{NC}(p_0)) > -(p_0 e^{-\lambda_k^*} - (1-p_0)w) e^{-rk^*}.$$

This condition always holds because $p_0 e^{-\lambda_h k^*} > (1-p_0)w$ and $V^C(p_0) \geq V^{NC}(p_0)$ by definition. Similarly, for $p_0 > \hat{p}$, comparing (20) and (21) yields

$$\begin{aligned} &p_0 \left(\pi_h - \frac{\lambda_h + r}{\lambda_l + r} \pi_l \right) (1 - e^{-(\lambda_l+r)(\tau^C - k^*)}) + (p_0 \lambda_h + r)(V^C(p_0) - V^{NC}(p_0)) \\ &> -(p_0 e^{-\lambda_l(\tau^C - k^*) - \lambda_k^*} - (1-p_0)w) e^{-r\tau^C}, \end{aligned}$$

which again always holds.

(iii) If $\hat{p} \geq \phi(P^1)$, then $\tau^C(\phi(P^1)) = k^*$. Since $\tau^{NC}(\phi(P^1)) = k^*$ by definition, $\Delta V(\phi(P^1)) = 0$. To see the necessity, note that $\Delta V(p_0) = 0$ implies $\tau^C(p_0) = \tau^{NC}(p_0)$. Since $\tau^{NC}(\phi(P^1)) = k^*$, we must have $\tau^C(\phi(P^1)) = k^*$ which holds if and only if $\hat{p} \geq \phi(P^1)$.

(iv) Suppose first that $1 > \hat{p}$, in which case τ^C and τ^{NC} solve

$$e^{-\lambda\tau^C(p_0)} = \frac{1-p_0}{p_0} \left(\frac{(\lambda_h+r)we^{-(\lambda+r)k^*}}{(\lambda_h+r)\pi_l - (\lambda_l+r)\pi_h(1-e^{-(\lambda_h+r)k^*})} \right) =: \frac{1-p_0}{p_0} \Phi^C,$$

$$e^{-\lambda_l\tau^{NC}(p_0)} = \frac{1-p_0}{p_0} \left(\frac{q^1}{(1-q^1)e^{-\Delta\lambda k^*}} \right) =: \frac{1-p_0}{p_0} \Phi^{NC},$$

respectively, for $p_0 > \hat{p}$. Note that both Φ^C and Φ^{NC} are independent of p_0 . From this, since $\lim_{p_0 \rightarrow 1} \tau^C(p_0) = \infty$ and $\lim_{p_0 \rightarrow 1} \tau^{NC}(p_0) = \infty$, we obtain

$$\lim_{p_0 \rightarrow 1} V^C(p_0) = \lim_{p_0 \rightarrow 1} V^{NC}(p_0) = \frac{p_0\pi_l}{\lambda_l+r} - \frac{(1-p_0)w}{r}.$$

If $\hat{p} = 1$, on the other hand, $\tau^C(p_0) = k^*$ for all p_0 , but as we have seen, $\lim_{p_0 \rightarrow 1} \tau^{NC}(p_0) = \infty$.

■

Appendix B: the value of commitment in $(\phi(P^1), 1)$

As we have seen in Proposition 5, the value of commitment moves with some regularities for $p_0 \in (q_h, \phi(P^1))$. The exception is the last segment $(\phi(P^1), 1)$. The reason is that for this (and only this) segment, we could have $\phi(P^1) > \hat{p} > q_l$ (see Propositions 3 and 4), so that the instantaneous payoff under low effort can be positive and $\tau^C(p_0) > k^*$ for $p_0 \in (\phi(P^1), 1)$. There are three cases we need to consider, depending on the value of \hat{p} .

Case 1: $\phi(P^1) > \hat{p}$

In this case, $\tau^C(\phi(P^1)) > \tau^{NC}(\phi(P^1)) = k^*$ and hence $\Delta V(\phi(P^1)) > 0$ (as we show in (iii) of Proposition 4). For $p_0 > \phi(P^1)$, $\tau^C(p_0)$ and $\tau^{NC}(p_0)$ increase proportionally (see the proof of Proposition 4), with both diverging to infinity as $p_0 \rightarrow 1$. As we show in (iv) of Proposition 4, $\Delta V(p_0)$ converges to zero as $p_0 \rightarrow 1$.

Case 2: $\hat{p} > \phi(P^1)$

In this case, $\tau^C(\phi(P^1)) = \tau^{NC}(\phi(P^1)) = k^*$ and hence $\Delta V(\phi(P^1)) = 0$. For $p_0 \in (\phi(P^1), \hat{p})$, $\tau^{NC}(p_0)$ continues to go up while $\tau^C(p_0)$ stays at k^* ; as a consequence, $\Delta V(p_0)$ gradually increases in this range. Once p_0 goes above \hat{p} , $\tau^C(p_0)$ also gradually increase, but again proportionally to $\tau^{NC}(p_0)$. Since $\tau^{NC}(p_0) > \tau^C(p_0)$ for $p_0 \in (\hat{p}, 1)$ in this case, $\Delta V(p_0)$ gradually decreases and converges to zero as $p_0 \rightarrow 1$.³⁸

³⁸If $\hat{p} = \phi(P^1)$, then $\tau^C(p_0) = \tau^{NC}(p_0)$ and hence $\Delta V(p_0) = 0$ for $p_0 \in (\hat{p}, 1)$. We ignore this non-generic case.

Case 3: $\hat{p} = 1$

This is the most straightforward case. Given that $\hat{p} = 1$, $\tau^C(p_0) = k^*$ for $p_0 \in (\hat{p}, 1)$ and hence $\tau^C(p_0)$ and $\tau^{NC}(p_0)$ diverge away from each other in this range. As a consequence, $\Delta V(p_0)$ increases as in all the preceding intervals.

Appendix C: The case with uncertainty

We analyze an extended version of the baseline model where a permanent productivity shock may strike with some probability. For simplicity, we assume that a shock arrives with a Poisson arrival rate β ; once it arrives, the agent becomes totally unproductive thereafter, i.e., $\lambda_h = \lambda_l = 0$. We also assume that the arrival of a shock is publicly observable. Other more complicated cases, e.g., those with transitory shocks, can in principle be analyzed in an analogous manner.

The agent: The agent's problem requires a slight modification. The augmented Bellman equation is obtained as

$$rU^j(k) = \max_{a \in \{l, h\}} \left((b\lambda_a - d_a + w) - (\lambda_a + \beta)U(k) + \beta\tilde{U}^j(k) - \dot{U}(k) \right), \quad j = C, NC$$

where \tilde{U}^j denotes the value function after the shock has arrived. Since the principal's reaction to the shock differs, \tilde{U}^j , and hence $U^j(k)$, depend on whether or not the principal makes commitment. As in the baseline model, the incentive compatibility constraint (before the shock strikes) can be written as

$$\Delta_\lambda(b - U^j(k)) \geq d,$$

although the value function $U^j(k)$ now differs. As such, the agent exerts high effort only when the project is sufficiently close to termination. We let k^C and k^{NC} denote the threshold under commitment and no commitment, respectively.

Proposition 7 *For any Θ and β , $k^{NC} > k^C$. Both k^{NC} and k^C are strictly increasing in β if $\frac{b\lambda_h - d + w}{\lambda_h + \beta + r} > b - \frac{d}{\Delta_\lambda} > 0$.*

Proof: We first consider the case where the principal commits to a deadline. Note that the principal cannot terminate the project even if the shock arrives before the deadline. Since the agent has no incentive to exert high effort, we have $\tilde{U}^C(k) = \frac{w(1 - e^{-rk})}{r}$. Let $U_a^C(k)$ denote the value function under commitment when the effort level is fixed at a . We then obtain

$$U_h^C(k) = \frac{b\lambda_h - d + w}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}) + \frac{\beta w}{r} \left(\frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} - \frac{e^{-rk}(1 - e^{-(\lambda_h + \beta)k})}{\lambda_h + \beta} \right).$$

The threshold under commitment is a solution to $U_h^C(k^C) = b - \frac{d}{\Delta_\lambda}$.

The problem is much more straightforward when the principal makes no commitment. In this case, the optimal choice for the principal is to terminate the project as soon as the shock arrives, which implies $\tilde{U}^{NC}(k) = 0$ for all k . As above, define $U_a^{NC}(k)$ as the value function under no commitment with fixed effort a , which is given by

$$U_h^{NC}(k) = \frac{b\lambda_h - d + w}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}).$$

If $\frac{b\lambda_h - d + w}{\lambda_h + \beta + r} > b - \frac{d}{\Delta_\lambda} > 0$, there exists an interior threshold k^{NC} given by

$$k^{NC} = -\frac{1}{\lambda_h + \beta + r} \ln \left(1 - \frac{(\lambda_h + \beta + r)(b - \frac{d}{\Delta_\lambda})}{b\lambda_h - d + w} \right).$$

It is evident that $k^{NC} > k^C$ because the expected payoff must be higher under commitment for a given horizon. Formally, we need to show that

$$\frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} > \frac{e^{-rk}(1 - e^{-(\lambda_h + \beta)k})}{\lambda_h + \beta},$$

which can also be written as

$$\int_0^k e^{-(\lambda_h + \beta + r)s} ds > \int_0^k e^{-(\lambda_h + \beta)s - rk} ds.$$

This condition holds because

$$e^{-(\lambda_h + \beta + r)s} > e^{-(\lambda_h + \beta)s - rk} \Leftrightarrow e^{-rs} > e^{-rk},$$

for any $s < k$.

Finally, note that $k^C, k^{NC} \in (0, \infty)$ when $\frac{b\lambda_h - d + w}{\lambda_h + \beta + r} > b - \frac{d}{\Delta_\lambda} > 0$. We can then show that k^j , $j = C, NC$, is increasing in β if $U_a^j(k)$ is decreasing in β for $k \leq k^j$. Straightforward computation shows that $U_h^{NC}(k)$ is strictly decreasing in β . In contrast, it is a little more involved to show the same for $U_h^C(k)$. To see this, with some computation, we obtain

$$\begin{aligned} U_h^C(k) &= \frac{b\lambda_h - d}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}) - \frac{\lambda_h w}{r} \left(\frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} - \frac{e^{-rk}(1 - e^{-(\lambda_h + \beta)k})}{\lambda_h + \beta} \right) \\ &\quad + \frac{w(1 - e^{-rk})}{r} \\ &= \frac{b\lambda_h - d - \lambda_h \frac{w(1 - e^{-rk})}{r}}{\lambda_h + \beta + r} (1 - e^{-(\lambda_h + \beta + r)k}) + \frac{\lambda_h w e^{-rk}}{r} \left(\frac{1 - e^{-(\lambda_h + \beta)k}}{\lambda_h + \beta} - \frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} \right) \\ &\quad + \frac{w(1 - e^{-rk})}{r}. \end{aligned}$$

First, it is easy to verify that $\frac{1-e^{-\alpha k}}{\alpha}$ is strictly decreasing for any $\alpha > 0$. Further, observe that the second term is strictly decreasing in β . These two facts imply that it suffices to show that

$$b\lambda_h - d > \lambda_h \frac{w(1 - e^{-rk})}{r},$$

for $k \leq k^C$. To show this, since $b\lambda_h - d > b\lambda_h - \frac{\lambda_h}{\Delta_\lambda}d \geq \lambda_h U_h^C(k)$ for $k \leq k^C$, we have

$$U_h^C(k) > \frac{\lambda_h(U_h^C(k) - \frac{w(1-e^{-rk})}{r})}{\lambda_h + \beta + r}(1 - e^{-(\lambda_h + \beta + r)k}) + \frac{\lambda_h w e^{-rk}}{r} \left(\frac{1 - e^{-(\lambda_h + \beta)k}}{\lambda_h + \beta} - \frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} \right) + \frac{w(1 - e^{-rk})}{r},$$

for $k \leq k^C$. This is further reduced to

$$\left(1 - \frac{\lambda_h(1 - e^{-(\lambda_h + \beta + r)k})}{\lambda_h + \beta + r} \right) \left(U_h^C(k) - \frac{w(1 - e^{-rk})}{r} \right) > \frac{\lambda_h w e^{-rk}}{r} \left(\frac{1 - e^{-(\lambda_h + \beta)k}}{\lambda_h + \beta} - \frac{1 - e^{-(\lambda_h + \beta + r)k}}{\lambda_h + \beta + r} \right),$$

which gives $U_h^C(k) > \frac{w(1-e^{-rk})}{r}$. ■

It is interesting to note that the possibility of a negative shock generally lowers the expected future payoff and hence induces the agent to exert high effort for longer, either with or without commitment ($k^{NC} > k^C > k^*$). First, with commitment, the best the agent can do after the shock strikes is to exert low effort and earn the flow payoff w , which is the lowest possible payoff as long as the project survives. The situation is even worse without commitment, however, because the principal can immediately terminate the project, thereby leaving no rent for the agent. The extent of the loss is therefore larger when the principal makes no commitment, which amounts to the fact that $k^{NC} > k^C$ and in turn favors the no-commitment solution.

The principal: We start with the case where the principal commits to a deadline at the outset. Let $V^C(k, p_0)$ denote the principal's expected payoff at time 0 under commitment, which can be written as

$$\begin{aligned} V^C(k, p_0) &= \int_0^{\nu^C} (\lambda_l p_s y - w) e^{-\int_0^s (\lambda_l p_u + \beta + r) du} ds \\ &\quad + e^{-\int_0^{\nu^C} (\lambda_l p_u + \beta + r) du} \int_{\nu^C}^k (\lambda_h p_s y - w) e^{-\int_{\nu^C}^s (\lambda_h p_u + \beta + r) du} ds \\ &\quad - \frac{\beta}{r} \int_0^{\nu^C} w(1 - e^{-r(k-s)}) e^{-\int_0^s (\lambda_l p_u + \beta + r) du} ds \\ &\quad - \frac{\beta}{r} e^{-\int_0^{\nu^C} (\lambda_l p_u + \beta + r) du} \int_{\nu^C}^k w(1 - e^{-r(k-s)}) e^{-\int_{\nu^C}^s (\lambda_h p_u + \beta + r) du} ds. \end{aligned} \quad (23)$$

where $\nu^C := \max\{\tau^C - k^C, 0\}$. The difference from the previous case is that a negative shock may now strike with some probability, in which case the principal earns zero from that point on. The commitment solution is obtained as

$$\tau^C = \operatorname{argmax}_\tau V_0^C(\tau).$$

As can be seen from (23), the marginal benefit of extending the deadline decreases with β , which prompts the principal to terminate earlier than in the case with no uncertainty.

In contrast, the problem is roughly the same as in the baseline model when the principal makes no commitment because the principal only pays attention to the instantaneous payoff. As in the baseline model, the principal stops when the instantaneous payoff becomes sufficiently low. To be more precise, define

$$\phi^{NC}(p) := \frac{p}{p + (1-p)e^{-\lambda_h k^{NC}}}.$$

Given that no shock has arrived, if $\phi^{NC}(q_h) > q_l$, the principal stops when the belief reaches q_h . The only difference is that when the shock strikes, the instantaneous payoff falls to $-w$, which prompts her to stop immediately. The principal's expected payoff is given by

$$\begin{aligned} V^{NC}(k, p) = & \int_0^{\nu^{NC}} (\lambda_l p_s y - w) e^{-\int_0^s (\lambda_l p_u + \beta + r) du} ds \\ & + e^{-\int_0^{\nu^{NC}} (\lambda_l p_u + \beta + r) du} \int_{\nu^{NC}(q)}^k (\lambda_h p_s y - w) e^{-\int_{\nu^{NC}(q)}^s (\lambda_h p_u + \beta + r) du} ds, \end{aligned} \quad (24)$$

where ν^{NC} is defined analogously. Note that for the same time horizon, $V^{NC}(k, p_0) > V_0^C(k, p_0)$ for any p_0 .

The optimal incentive scheme: In the baseline model with no uncertainty, any allocation which can be realized without commitment can be realized with commitment as well. This is no longer the case with uncertainty because the principal cannot foresee if and when the shock strikes, in which case she needs to immediately terminate the project. We can then make the following statement.

Proposition 8 *The principal strictly benefits from not committing to a deadline for any $\beta > 0$ if $\phi^{NC}(q_h) \geq p_0$.*

Proof: If $\phi^{NC}(q_h) \geq p_0$, the agent exerts high effort from the beginning under no commitment. The principal's payoff is maximized, and the principal strictly benefits from making no commitment in this case. To see this, observe that $k^{NC} > k^C$ implies $\phi^{NC}(q_h) > \phi^C(q_h)$. If $\phi^{NC}(q_h) \geq p_0 > \phi^C(q_h)$, the agent starts off with low effort, and the principal is forced to

set a deadline before the belief reaches q_h . The profit is clearly lower than $V^{NC}(\tau^{NC}, p_0)$. On the other hand, if $\phi^{NC}(q_h) > \phi^C(q_h) \geq p_0$, the agent exerts high effort from the beginning even under commitment. The profit is still lower, however, because the principal must incur a flow loss when the shock strikes whereas she can immediately terminate the project if she makes no commitment. ■

The presence of uncertainty lowers the value of commitment through two channels. First, by making no commitment, the principal can retain the flexibility to adjust to the negative shock by immediately terminating the project whenever it strikes. Aside from this conventional effect of flexibility, there also arises another force that favors the no-commitment solution: in the face of uncertainty, the agent is better motivated by the constant threat of project termination, and thus, he starts exerting high effort earlier. Since $\phi^{NC}(q_h) > \phi(q_h)$ for any $\beta > 0$ by Proposition 7, we could have $\phi^{NC}(q_h) \geq p_0 > \phi(q_h)$, in which case the value of commitment is strictly negative with uncertainty whereas it is strictly positive without.

7 Appendix D: The case with endogenous wages

We consider a setup where the principal can deliberately choose $w \geq 0$ at time 0. In general, an increase in w above zero yields two opposing effects. On one hand, it raises the critical value q_h , which may give the principal the incentive to stop earlier than otherwise. On the other hand, it also raises the expected future payoff which in turn lowers k^* and gives the agent the incentive to procrastinate even more. A necessary condition for $w > 0$ is that the first effect dominates the second, or alternatively that $\tau^{NC} - k^*$ decreases with w .³⁹ We can then make the following statement.

Proposition 9 *It is optimal to set $w > 0$ if λ_l is sufficiently small.*

Proof: As $w \rightarrow 0$, we have $q_h \rightarrow 0$ and $q_l \rightarrow 0$, which implies $\tau^{NC}(p_0) \rightarrow \infty$ for any p_0 . The expected payoff then converges to

$$\lim_{w \rightarrow 0} V^*(\tau^{NC}, p_0) = \frac{p_0 \lambda_l y}{\lambda_l + r}, \quad (25)$$

for any p_0 . Define $\mu(w)$ such that $\phi(\frac{w}{\lambda_h y}) = \mu(w)$, i.e.,

$$\frac{\mu(w)^{-\lambda_h k^*}}{1 - \mu(w) + \mu(w)^{-\lambda_h k^*}} = q_h = \frac{w}{\lambda_h y}.$$

³⁹Note that $\tau^{NC} - k^*$ is the interval during which the agent exerts low effort. If it is longer with a smaller k^* , the principal's expected payoff is necessarily lower.

Note that

$$\frac{1 - \mu(w)}{\mu(w)} = \frac{1 - q_h}{q_h} e^{-\lambda_h k^*} = \frac{\lambda_h y - w}{w} \left(1 - \frac{(\lambda_h + r)(b - \frac{d}{\Delta_\lambda})}{\lambda_h b - d + w} \right)^{\frac{\lambda_h}{\lambda_h + r}}.$$

Since $\lim_{w \rightarrow 0} \mu(w) = 0$ and $\mu(\lambda_h y) = 1$, there must exist at least one \tilde{w} such that $\mu(\tilde{w}) = p_0$. If we set $w = \tilde{w}$, then $\tau^{NC} = k^* > 0$ by definition, and the expected payoff is

$$V(k^*, p_0) = \int_0^{k^*} (\lambda_h y p_s y - \tilde{w}) e^{-\int_0^s (\lambda_h p_u + r) du} ds,$$

which is always strictly positive and larger than (25) if λ_l is sufficiently small. ■

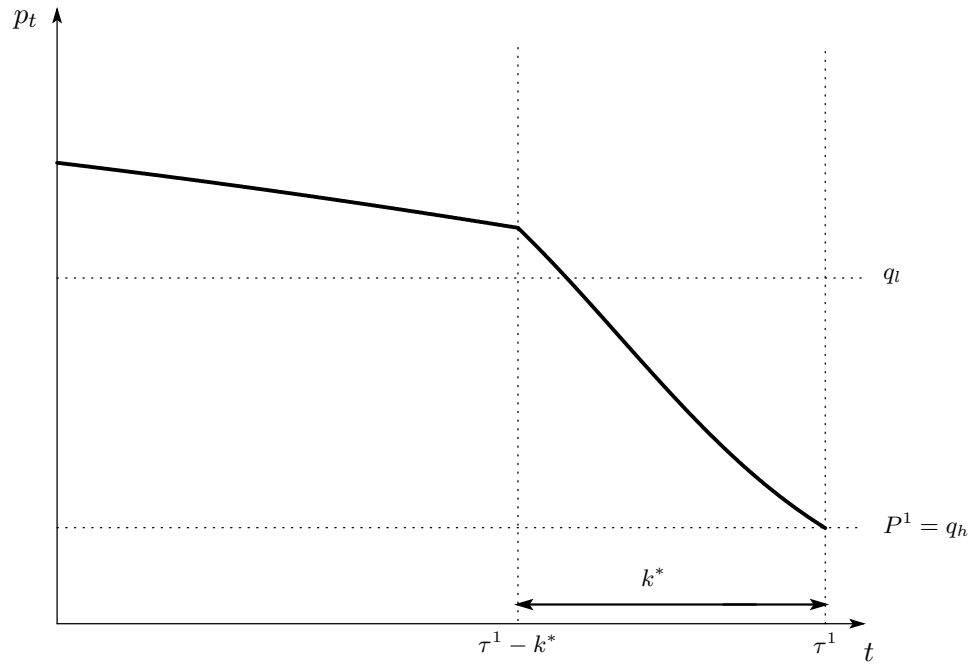


Figure 1: Evolution of the belief on the equilibrium path ($m(\Theta) = 1$)

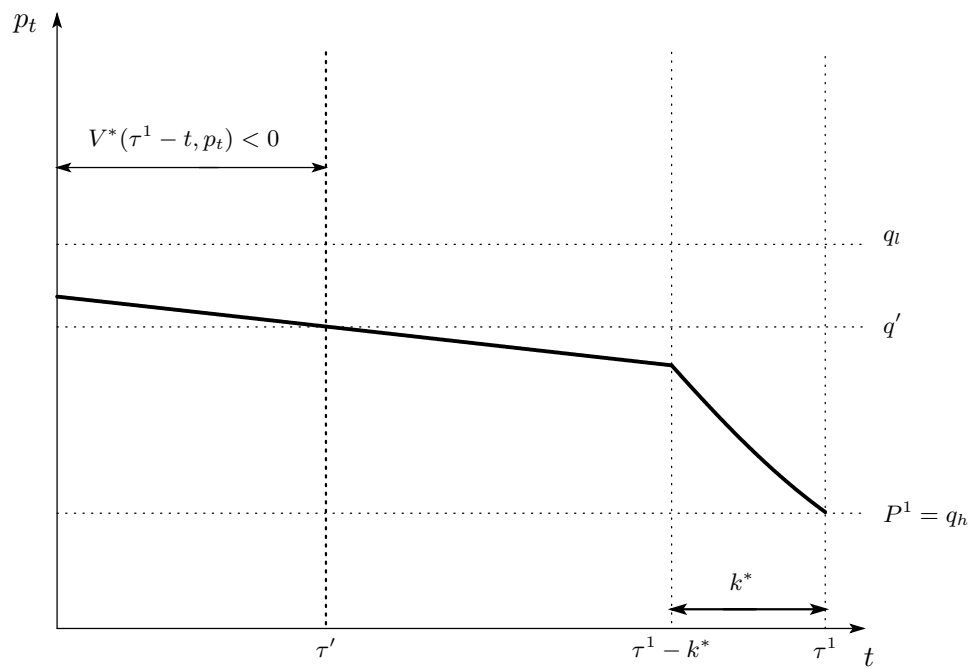


Figure 2: A violation of Condition C: the belief ($m(\Theta) = 2$)

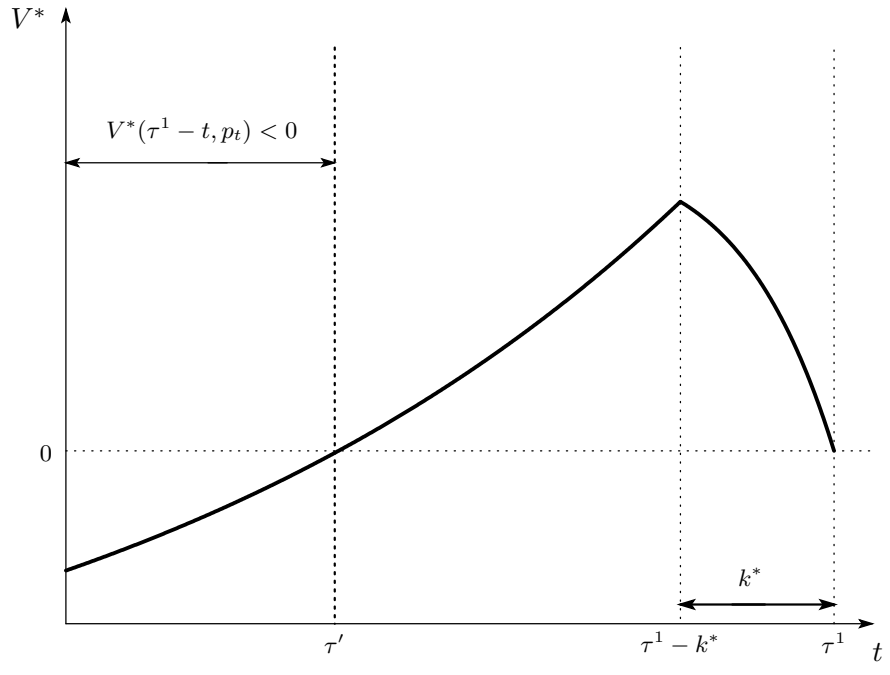


Figure 3: A violation of Condition C: the continuation payoff

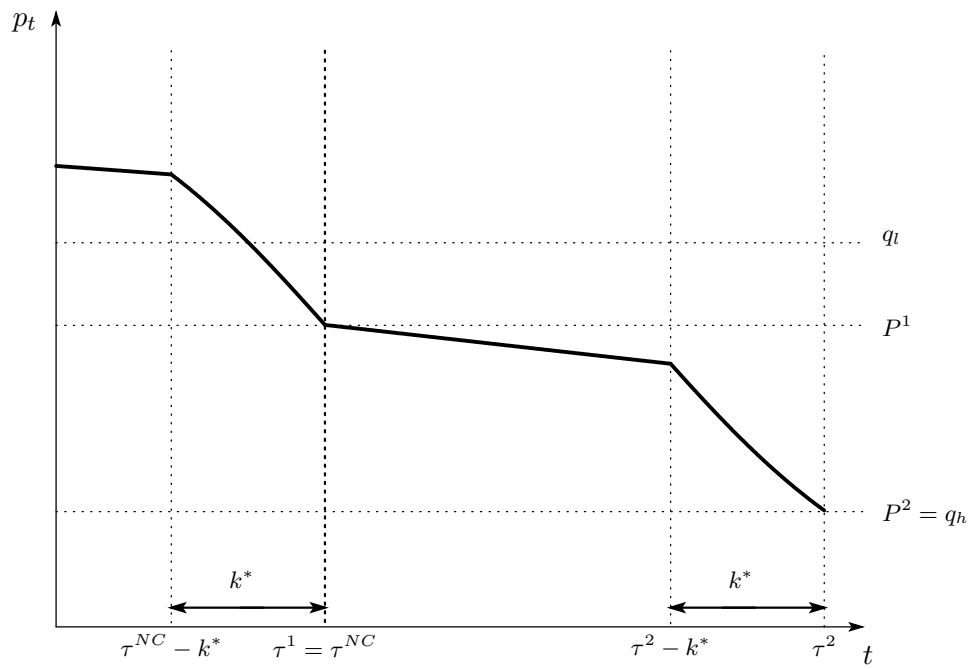


Figure 4: Evolution of the belief, on and off the equilibrium path

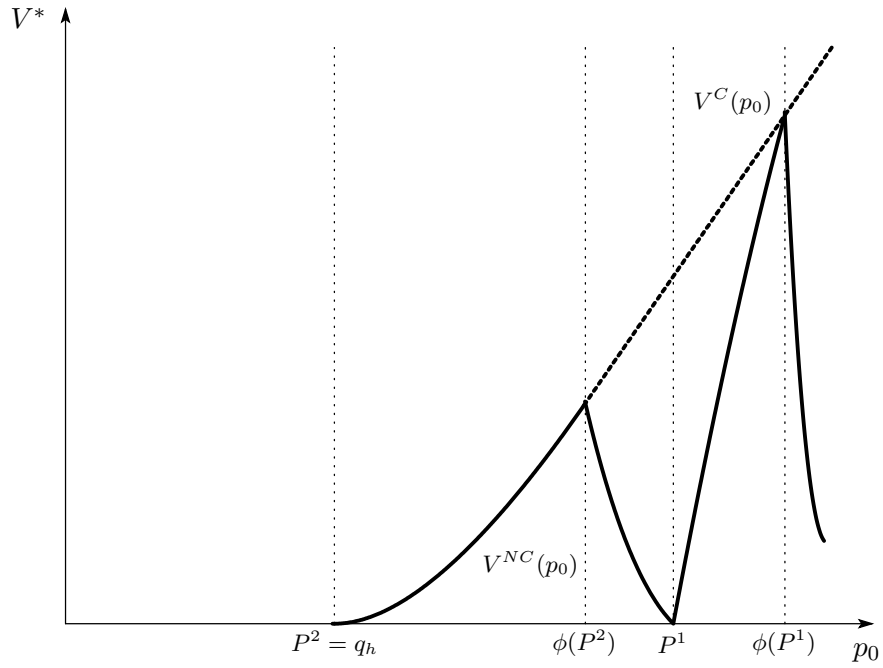


Figure 5: The expected payoffs with and without commitment

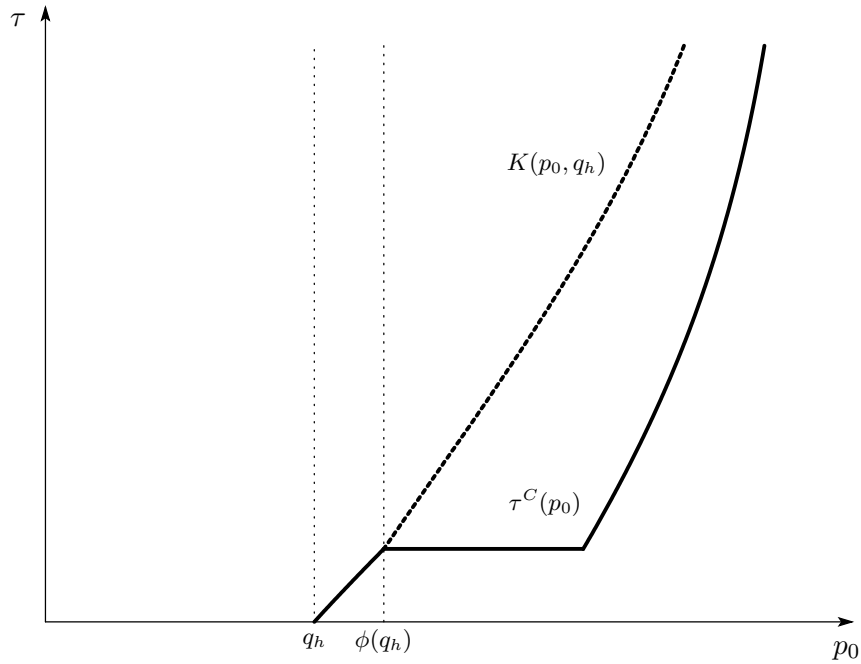


Figure 6a: The commitment solution when λ_l is relatively high ($\lambda_l = 0.07$)

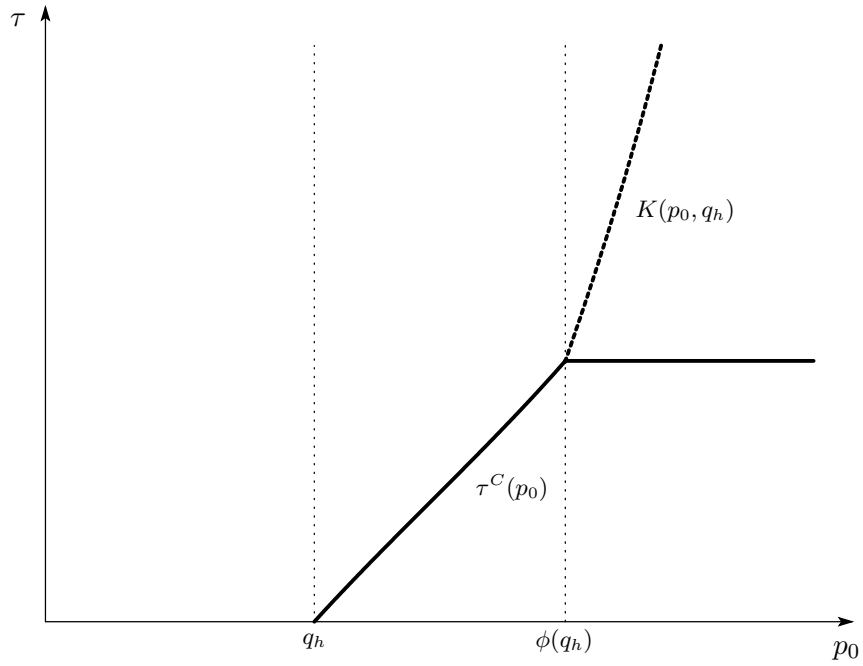


Figure 6b: The commitment solution when λ_l is relatively low ($\lambda_l = 0.04$)

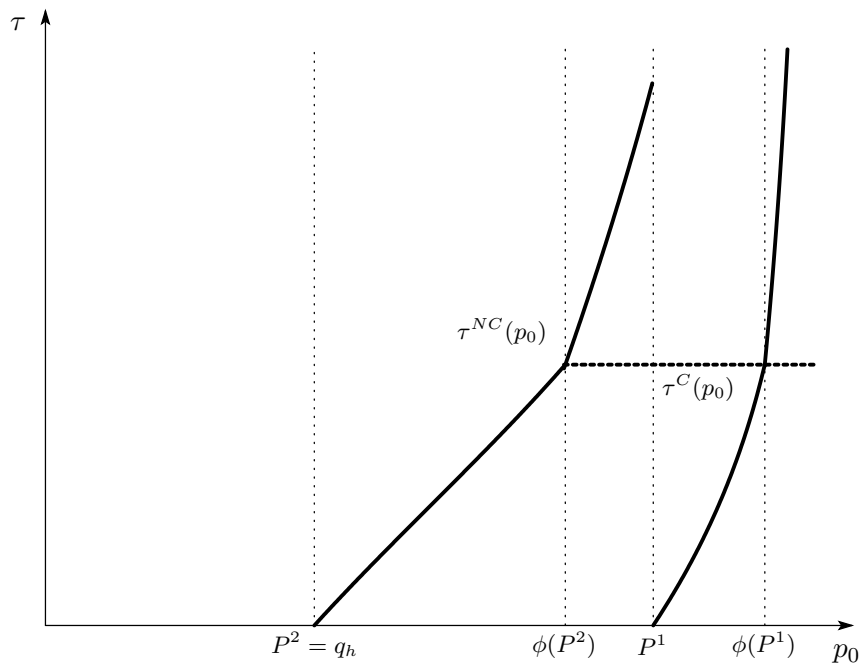


Figure 7: The no-commitment solution