Delegated Learning and Non-Credible Communication

Peter B. McCrory and Farzad Pourbabaee

August 2020

Abstract

We consider a setting in which an impatient agent acquires payoff-relevant information about the true state of the world. The agent endogenously chooses when to stop learning, at which point an uninformed principal takes an action to maximize her own expected payoff. The agent's preferences are biased relative to the principal's, generating misalignment of expected payoffs. When communication is non-credible, the principal can only rely upon the agent's endogenous stopping rule when strategically specifying her course of action. In the no-communication equilibrium, the agent adopts a one-sided stopping rule as a function of her posterior belief that is consistent with the principal's pre-specified course of action at the time of stopping. When the principal has commitment power, relative to the full-communication equilibrium, the agent is always worse off; for intermediate values of prior beliefs, the principal is better off. The one-sided equilibrium stopping rule (and associated action) can switch discretely as a function of prior beliefs, generating dramatic regime changes for arbitrarily small changes in beliefs. When learning is initiated in the no-communication equilibrium there is a non-zero probability of indefinite delay, in which the agent never ceases learning and the principal never takes an action.

JEL classification: C73; D82; D83

Keywords: Delegated learning; Strategic stopping; Asymmetric information

University of California, Berkeley (Contact: pbmccrory@berkeley.edu and farzad@berkeley.edu). The authors thank Robert Anderson and Philipp Strack for helpful discussions and suggestions.

1 Introduction

In many economic interactions with private information, the agent with superior knowledge often acquires such information over time and with increasing confidence. In such settings, the timing of when the agent approaches a counter-party responsible for taking a payoff-relevant action may be dependent as much upon what the agent has learned as when he has learned it.

Consider the canonical example of the used car market. By virtue of owning, driving, and maintaining his car, the seller likely has greater information about its quality than any potential buyer.[1] Yet, this superior knowledge is not acquired instantaneously upon initial purchase of the car. Rather, it takes time for any new owner of a car to learn about its long-term quality. This learning process may never fully cease so long as the seller owns the car, even as his confidence about the quality of his car improves over time.

Of course, the timing of when an owner of a car decides to sell his car is not random—it likely depends upon both the market price for the used car as well as the confidence the seller has in its underlying quality. Moreover, while the potential buyer of a car may not observe the posterior belief of the seller at the time of sale, the buyer nevertheless is aware that the seller has chosen to sell the car after having spent some time learning about the car's quality. This latter feature of the economic exchange raises a novel question as it relates to economic interactions with private information: How does the time-dependence of the learning process and the endogeneity of when the economic interaction takes place determine the equilibrium strategies employed by both principal and agent?

The economic forces we analyze in this paper manifest themselves in other settings as well, such as in the following example. In advanced economies, large firms with already established sources of revenue nevertheless devote considerable financial resources to R&D activities. In such large and complex organizations, those making the decision for whether an innovation is ultimately developed further oftentimes delegate the R&D process to subordinate, research employees.

The R&D employee likely has a preference for early resolution of the research process, as opposed to the firm, which is much less dependent upon any particular idea for survival. Moreover, the research employee's preferences may not be perfectly aligned with that of his managers if, for example, private benefits accrue to him in the event that the product line

[1] This is, of course, the motivation articulated in Akerlof (1970) for studying the implications of asymmetric information in economic interactions. Akerlof writes: "After owning a specific car, however, for a length of time, the car owner can form a good idea of the quality of this machine"

enhancement is ultimately adopted. But otherwise, he likely has a vested interest in the viability and profitability of the product line enhancement.

Together, these forces determine when the research employee optimally concludes his research. When he does, he submits a report to his superiors, who then make a decision about whether the idea is developed further. In light of this example, our paper has two surprising and interesting results: (i) For a certain range of prior beliefs about the quality of R&D projects, the firm's managers prefer to ignore the informational content of the report by their subordinates—relying instead on the endogenous arrival of reports (which, in turn, disciplines idea quality of submitted reports in equilibrium); (ii) In some cases, minor changes in the prior belief over idea quality can, in the non-credible communication case, produce a sharp regime change in which the types of reports submitted in equilibrium switch from being those believed to be low quality to those believed to be high quality. As such, the firm's managers correspondingly adjust their equilibrium action from low to high.

Contributions. The examples outlined in the introduction provide clear motivation for the general problem we investigate in this paper. Here, we briefly discuss the set-up of the problem and the contributions our paper makes.

We study a principal agent problem in which the principal (she) delegates the task of learning about the payoff relevant state of the world, θ, to the agent (he). Both the principal and the agent share a common prior belief, π, about θ; however, the agent has access to a technology which allows him to acquire information over time about the likely payoff relevant state of the world. Naturally, the longer the agent spends learning, the more precise is his estimate of θ.

When the agent stops learning about θ, he delivers a report to the principal, who then updates her belief about the likely state of the world. Given how we set up the problem, the principal's optimal action is exactly her posterior belief about θ. The principal does not discount the future and is thus willing to delay her action until the agent (optimally) stops learning and delivers his report.

The strategic dimension of the problem arises from two salient differences in preferences between the agent and the principal. First, the agent receives a private benefit simply from delivering the project to the principal. The magnitude of this private benefit helps determine how willing the agent is to devote time to learning about θ as opposed to immediately sending a report to the principal. Second, the agent's preferences are biased *upwards* relative to the principal's—for any belief about the true state of the world, the agent would prefer the principal to take a higher action than what is Bayes-rational for the principal. This tension of preferences can be embedded in the communication framework of Crawford and Sobel (1982).

For this problem, we study two extreme communication environments: full communication and no communication. By doing so, we sidestep issues related to strategic communication. In the former, the full-communication game, the agent truthfully communicates his posterior belief to the principal after devoting some time to learning about θ. Based on the agent's report, the principal takes the action which maximizes her own expected utility.

How long the agent spends learning depends upon the action taken by the principal. Anticipating the principal's optimal action, the agent knows precisely his exit value at the time of stopping. He thus solves an optimal stopping time problem. The unique Markov perfect equilibrium entails the agent continuing the learning process for *intermediate* values of his continuously updated posterior belief (the continuation region is symmetric around 1/2); the agent stops learning when his posterior belief hits either the upper or lower boundary of the continuation region, both of which are in the interior of the unit interval.[2] Because the agent truthfully communicates his posterior belief when learning ceases, a high level of π is matched with a *high* action $(1 - \alpha_c)$ by the principal; a low posterior belief is similarly mapped to a low action α_c. This equilibrium requires that the agent's solution to the optimal stopping time problem be consistent with the induced action of the principal.

A more interesting situation arises when credible communication between the parties is infeasible or prohibitively costly. The lack of formal communication induces, for both parties, a strategic reliance upon the optimal stopping time problem faced by the agent to discipline equilibrium actions. While communication, per se, does not occur, the equilibrium features *ex post* revelation of the agent's posterior belief at the time that the agent stops the learning process.

The solution concept we use is the Markov perfect equilibrium. The equilibrium profile features a one-sided continuation set $\mathcal{C}$ chosen by the agent; when the posterior belief remains in this set he finds it optimal to continue learning. In turn, the principal specifies a single plan of action ξ that she takes when the agent stops learning. Necessarily, $\mathcal{C}$ is the agent's best-response to the principal's action ξ, and ξ is the principal's best-response to $\mathcal{C}$; the principal knows that at the time of stopping the agent's posterior belief has hit the boundary $\partial\mathcal{C}$. This equilibrium feedback substantially disciplines the types of continuation sets that survive the fixed-point. Despite the lack of communication, the agent's ability to hide his posterior belief

[2]Precisely, there exists an open interval $(\alpha_c, 1 - \alpha_c)$ such that the agent continues the learning so long as his posterior process $\boldsymbol{\pi} = \{\pi_t\}$ remains in this set.

from the principal is remarkably limited. In equilibrium, the principal perfectly elicits the agent's posterior belief without communication. Consequently, as we show formally, for all *intermediate* beliefs, the principal's payoff is higher under no communication and the agent is worse-off relative to the game with full communication.

The no-communication environment features the emergence of *discontinuous* regime changes for projects of *ex ante* similar quality. Specifically, we show that a small change in initial belief about θ can, in equilibrium, lead the agent to reverse his one-sided stopping rule. In turn, rather than approaching the principal when his posterior is sufficiently low, the agent only stops learning when his posterior belief about the quality of the project is sufficiently high. As an implication, the principal switches from taking a low action to taking a high action at the time of stopping.

The determining factor for these type of discontinuities depends upon whether the principal has commitment power in the no-communication game. When the principal lacks commitment, the agent has first-mover advantage to specify which of the two continuation regions prevail after initiating the learning process. Thus, the discontinuous regime change occurs in the no-communication game without commitment at exactly that point where the agent's preference over either of the two continuation regions switches.

In the no-communication game with commitment, the principal commits to a course of action at the time of stopping, which in turn disciplines the continuation region chosen by the agent, should he decide to initiate learning. By implication, the discontinuities in this setting arise because the agent has the option to immediately stop/never start learning, thereby preventing the principal to free ride on the agent's research. When this threat is credible and the agent would not initiate learning under the principal's preferred course of action, the principal instead adopts her second-best course of action (e.g. switching from a low to high action at the time of stopping).

Literature review. Our paper builds on the seminal work of Crawford and Sobel (1982) followed by Melumad and Shibano (1991) by allowing the agent to have incomplete knowledge about the hidden state of the world and thus engaging in a noisy-learning problem. However, we restrict the space of communication between the two parties to two extreme cases, namely full communication and no communication. We thereby purposefully downplay the role of strategic communication and emphasize, instead, the strategic interaction of the agent's learning decision on the principal's action choice.

The separation of learning from the action choice in the context of research and approval is studied in Carpenter and Ting (2007), McClellan (2017) and Henry and Ottaviani (2019). In contrast to these papers, we study an environment characterized by both *no communication* between the parties and the absence of a public signal. Our model thus assumes away symmetric information by only letting the agent to have access to the acquired information. Hence, the only available information to the principal at the time of stopping is the shared initial belief and the observation that agent has submitted the project for the principal to take an action.

Our work relates also to Grenadier et al. (2016) and Orlov et al. (2019) in which a principal decides whether and when to undertake an irreversible investment based on a public signal (observable to both parties) and the information provided by the agent about a payoff relevant hidden variable. Both papers highlight the fact that the direction of the bias between the principal and the agent guides both the agent's preference toward early or late submission and his ability to credibly communicate his information to persuade the principal. Aside from the differences in the payoff structure – ours is quadratic (following the hypothesis testing literature) and theirs is quasi-linear utility (following the organizational investment literature) – we deliberately abstract away from communication issues by restricting ourselves to the extreme cases of mandatory full communication and infeasible communication. Particularly, we are interested in understanding how the mandates set by the principal guide the learning decisions of the agent and eventually determine the types of projects submitted by the agent to the principal.

Finally, on a more distant thread, our paper is related to the strategic experimentation literature (c.f., Bolton and Harris (1999) and Keller et al. (2005))—particularly its delegated versions such as Guo (2016) and Klein (2016). In contrast to such delegated strategic experimentation models, our model does not allow the principal to observe the output of the experimentation process carried out by the agent.

Organization of the paper. In section 2 we study the stage game between the parties and their preference profiles. In section 3 we present the benchmark of full communication. Subsequently, in section 4 we present the equilibrium results in the absence of communication between principal and agent. Results on the comparison between two communication extremes are provided in section 5. Finally, we conclude in section 6. The proofs that are not offered in the main body are relegated to the appendix A.

2 Model

An agent is responsible for learning about the viability of a project. When the agent stops learning, the principal is tasked with executing the project. The project's type is a binary random variable $\theta \in \{0, 1\}$, unobserved to both parties. Both parties share a common initial belief $\pi = \mathsf{P}(\theta = 1)$ at time 0. The agent's role is to collect information about θ and the principal gets to take an action $a \in \mathbb{R}$.

The information flow process x is *only* observable to the agent and follows a drift-diffusion $\mathrm{d}x_t = \theta \mathrm{d}t + \sigma \mathrm{d}Z_t$. This implies that the posterior process $\{\pi_t\}$ held by the agent follows[3]

$$\mathrm{d}\pi_t = \frac{\pi_t(1-\pi_t)}{\sigma}\mathrm{d}\bar{Z}_t, \tag{2.1}$$

in that $\sigma \mathrm{d}\bar{Z}_t = \mathrm{d}x_t - \pi_t \mathrm{d}t$ is the innovation process. The agent stops at a random time τ, at which time the principal takes an action. After she takes an action, the value of θ is realized. The utility of agent and principal after the realization of θ are, respectively, equal to

$$U_A(a,\theta) = \kappa - (a - b - \theta)^2$$
$$U_P(a,\theta) = -(a-\theta)^2.$$

The first term in the agent's utility, κ, represents his *private benefits* from the execution of the project. The parameter b, assumed to be positive without loss of generality, represents the extent of *conflict of interest* between the two parties. Further, we assume the agent is impatient with time preference rate of $\rho > 0$, while the principal is infinitely patient.

3 Full Communication at Stopping

Suppose that when the agent stops learning he must perfectly communicate his posterior belief π_τ to the principal. That is, the agent cannot hide his collected information from the principal. As such, the principal naturally takes action π_τ for any stopping time τ, thereby maximizing her expected payoff. This, in turn, determines the exit value function for the agent:

$$g_c(\pi_\tau) = \kappa - \pi_\tau(\pi_\tau - b - 1)^2 - (1-\pi_\tau)(\pi_\tau - b)^2 = \kappa - b^2 + \pi_\tau^2 - \pi_\tau. \tag{3.1}$$

[3] This follows directly from Theorem 9.1 in Lipster and Albert (2001).

The agent then follows the path of his posterior process until he takes an exit decision. Effectively, he solves the stopping time problem[4]

$$V_{A,c}(\pi) = \sup_{\tau} \mathsf{E}\left[e^{-\rho\tau} g_c(\pi_\tau)\right]. \tag{3.2}$$

The standard HJB equation from the perspective of the agent who observes the posterior path is

$$\rho V_{A,c}(\pi) = \frac{1}{2\sigma^2}\pi^2(1-\pi)^2 V''_{A,c}(\pi). \tag{3.3}$$

In the sequel, we repeatedly use the general solution to the HJB equation (3.3), which has the following form[5]

$$c_1\pi^{1-\lambda}(1-\pi)^{\lambda} + c_2\pi^{\lambda}(1-\pi)^{1-\lambda}, \tag{3.4}$$

in which c_1 and c_2 are real constants and $\lambda = \frac{1+\sqrt{1+8\sigma^2\rho}}{2}$, that is strictly larger than one. On the space of twice continuously differentiable functions on the unit interval $C^2[0,1]$, one can define the infinitesimal characteristic operator $\mathbb{K}$ for the posterior process $\{\pi_t\}$ as

$$\mathbb{K}h = -\rho h + \frac{\pi^2(1-\pi)^2}{2\sigma^2}h'', \quad \forall h \in C^2[0,1]. \tag{3.5}$$

Throughout the paper we assume $b^2 < \kappa < 1/4$. This assumption rules out the uninteresting cases of immediate stopping and indefinite delay *irrespective* of the initial belief.

Proposition 1 (Full Communication)**.** *There exists a unique Markov perfect equilibrium in which the agent continues the learning process so long as* $\pi_t \in (\alpha_c, 1-\alpha_c)$*, and stops otherwise. Further, the point* α_c *exists uniquely in* $(0, 1/2)$.

As a result of this proposition, the principal takes the *high action* $(1-\alpha_c)$ when the belief about θ is high and takes the *low action* α_c when the belief is low. When the initial belief is not in the interval $(\alpha_c, 1-\alpha_c)$, the agent stops learning immediately and the principal takes action π. This description amounts to Figure 1 in which we plot the value functions for both parties, with the left y-axis used for V_A (solid blue line) and the right y-axis for V_P (solid black line).[6] At the boundary points $\{\alpha_c, 1-\alpha_c\}$ the agent's value function smoothly meets the stopping value (dashed blue line), whereas kinks appear on the principal's payoff

[4] In the paper, we use the subscript c to refer to the full communication and n to the no-communication situation.

[5] Of course, if the agent pays (receives) a flow cost (benefit) while learning, this would manifest as an additional constant term in the HJB equation and in the general solution without materially affecting the results.

[6] For all figures, we use the following combination of primitives: $\kappa = 0.24, b = 0.17, \lambda = 2.2$.

function. This is owed to the fact that the decision to stop or continue is made by the agent and not the principal. Also, in the continuation region $(\alpha_c, 1-\alpha_c)$ the principal has constant valuation: when the initial belief is π, with probability $\frac{\pi-\alpha_c}{1-2\alpha_c}$ the posterior belief path will hit the upper barrier $(1-\alpha_c)$ and with complementary probability it will hit the lower barrier α_c. In the former case, the principal's action is $1-\alpha_c$ while in the latter case it is α_c. In both cases, her conditional expected payoff is $-\alpha_c(1-\alpha_c)$, resulting in the flat valuation on the continuation region.

To summarize: with perfect communication of the agent's ex post belief, the valuation of both parties is symmetric around $\pi = 1/2$ with high (low) actions associated to high (low) levels of the beliefs. Both of these features of the equilibrium disappear in the absence of communication, as presented in the next section.

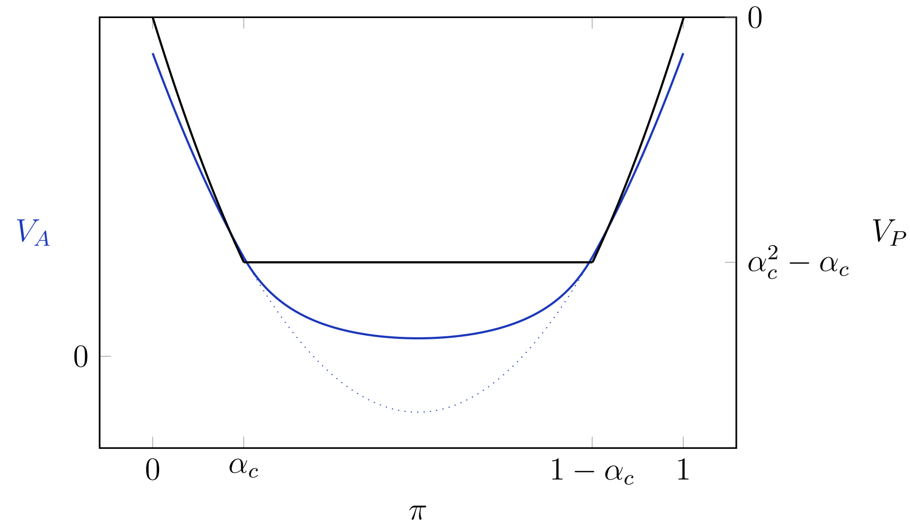

Figure 1: Value Functions in Full Communication

This figure plots the equilibrium value functions of the agent (V_A; solid black) and of the principal (V_P; solid blue) in the full-communication game, each as a function of the initial belief π. At the time of stopping, the agent truthfully communicates his posterior belief to the principal who then takes her utility maximizing action. The dotted-blue line represents the agent's exit value $g_c(\pi_\tau)$ as a function of her continuously updated posterior belief. When the agent's prior/posterior belief is within the continuation region $(\alpha_c, 1-\alpha_c)$, he continues learning. Otherwise, he stops learning and reports his belief to the principal.

4 Strategic Stopping with No Communication

In this section, we consider an environment in which credible communication is either infeasible or prohibitively costly. This modeling choice is, of course, stylized. We make this choice to sidestep issues of strategic communication à la Crawford and Sobel (1982). Such a scenario could arise if, for example, the knowledge acquired by the agent is extremely specialized and there is a fixed cost associated with conveying such knowledge in terms that the principal could understand.[7]

In situations in which it is infeasible for the agent to communicate what he has learned, the principal's ability to commit to a course of action when learning ceases plays an important role in determining the equilibrium outcome of the game. Therefore, we distinguish between two cases: (i) the principal *cannot* commit ex ante to an action profile at the time of stopping and (ii) the principal can commit ex ante and hence her choice *shapes* the agent's learning decision. We present results for both cases in the following subsections.

4.1 Principal Lacks Commitment Power

In this environment the agent starts the game by choosing whether to initiate the learning process or not. This corresponds to a binary decision $d \in \{\texttt{stop}, \texttt{continue}\}$ at $t = 0$. Conditioned on initiating the learning process, i.e $d = \texttt{continue}$, the agent chooses a continuation region $\mathcal{C} \subset [0, 1]$ that encodes all relevant information of the stopping time problem he faces. Particularly, he continues learning so long as $\pi_t \in \mathcal{C}$ and stops otherwise. Therefore, the agent's strategy is summarized by the mapping $(\boldsymbol{d}, \boldsymbol{\mathcal{C}}) : [0, 1] \to \{\texttt{stop}, \texttt{continue}\} \times 2^{[0,1]}$, where the domain of this mapping corresponds to the initial belief $\pi \in [0, 1]$.[8]

In the subsequent stage of the game, the principal specifies a course of action, as depicted in Figure 2. Obviously, if the agent stops right away at $t = 0$, the principal's optimal action is π. For all $t > 0$, we only consider strategies for the principal in which she takes a fixed action ξ, independent of the stopping time.

In this game in which the principal has no commitment, the agent has the first mover advantage. The agent's choice of whether or not initiate the learning disciplines the course of action specified by the principal, both on and off the equilibrium path. Anticipating results presented in the next subsection, relative to the case in which the principal has commitment power, this yields for the agent a weak improvement in his welfare for all initial beliefs, with a strict improvement for a subset of initial beliefs with positive measure.

[7] Alternatively, as we will see below, when she is able to commit to a course of action, the principal is better off in the no-communication equilibrium relative to the full-communication equilibrium for intermediate ranges of initial beliefs about the project's quality. If the principal can *ex ante* commit to ignoring the agent's report (even off-equilibrium) and to a pre-specified course of action, then the principal would choose the no-communication equilibrium.

[8] We use bold-face symbols to refer to the mappings and use regular font to denote a generic output of that map.

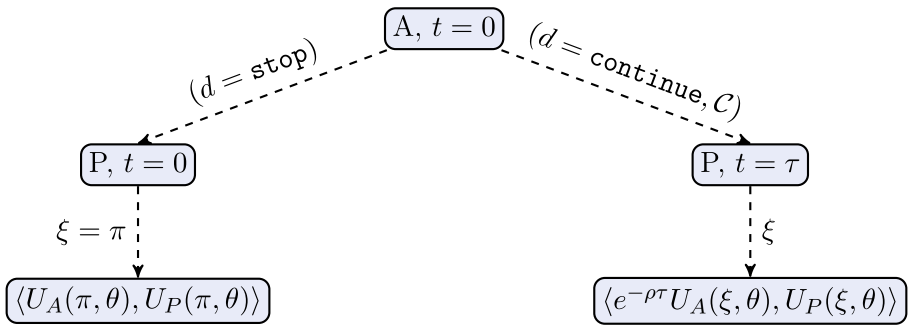

Figure 2: Game tree – no commitment from the principal

Notation 2. For any Borel-measurable subset $B \subset [0, 1]$, denote τ_B as the first time the process $\{\pi_t\}$ hits the set B. That is, $\tau_B = \inf \{t \geq 0 : \pi_t \in B\}$. We denote the boundary of any such set, B, by ∂B.

Before presenting the definition of the equilibrium, we need to introduce one more object. Given that the principal takes action ξ at the stopping time, the *exit-value* of the agent at the stopping time τ is denoted by $g_n(\pi_\tau; \xi)$. For a fixed ξ, this is only a function of his posterior belief at τ. Except where necessary, we suppress the dependence of the agent's exit-value upon the principal's pre-specified course of action ξ. This exit-value is given by

$$g_n(\pi_\tau; \xi) = \mathsf{E}_{\pi_\tau} \left[\kappa - (\xi - b - \theta)^2 \right] = \kappa + \pi_\tau \left[2(\xi - b) - 1 \right] - (\xi - b)^2 \tag{4.1}$$

Definition 3 (Equilibrium for the No-Commitment, No-Communication Game)**.** Given the initial belief π, the tuple of strategy mappings $\langle \boldsymbol{\xi}, (\boldsymbol{d}, \boldsymbol{\mathcal{C}}) \rangle$ constitutes a Markov perfect equilibrium in the no-commitment, no-communication game if

(i) The principal's action $\boldsymbol{\xi}(\pi, d, \mathcal{C})$ is a best-response to every $(d, \mathcal{C})$. Equivalently, given the continuation set $\mathcal{C}$, the principal's action reflects her rational beliefs, that is on $\{\tau < \infty\}$

$$\begin{aligned} &\xi = \mathsf{E}_\pi \left[\theta \mid \pi_\tau \in \partial \mathcal{C}, d = \texttt{continue} \right] \\ &\xi = \pi \quad \text{if } d = \texttt{stop}. \end{aligned} \tag{4.2}$$

(ii) The agent's continuation set $\boldsymbol{\mathcal{C}}(\pi)$ is a best-response to $\boldsymbol{\xi}$.

(iii) The agent's decision at time 0 is $d = \texttt{continue}$ iff

$$\begin{aligned} V_{A,n}(\pi) &:= \sup_{\tau} \mathsf{E}_\pi \left[e^{-\rho\tau} U_A(\xi,\theta) \right] = \sup_{\tau} \mathsf{E}_\pi \left[e^{-\rho\tau} g_n(\pi_\tau) \right] \\ &\geq Q_A(\pi) := \kappa - b^2 + \pi^2 - \pi, \end{aligned} \tag{4.3}$$

where $\xi \equiv \boldsymbol{\xi}(\pi, \texttt{continue}, \boldsymbol{\mathcal{C}}(\pi))$.

The three conditions in our definition of the equilibrium are simply the standard best-response requirements along each sub-game of the full-game in Figure 2, after solving for the sub-game equilibria by backwards induction.

The first condition places the most discipline on the equilibrium outcome. Equation (4.2) states that, in equilibrium, the principal's action rationally accounts for the fact that π_τ is on the boundary of the continuation region chosen by the agent. The second condition (ii) requires that, conditioned on deciding to continue, the agent picks the optimal continuation region in response to his (correct) belief that the principal will follow her equilibrium strategy $\boldsymbol{\xi}$. And finally, (iii) requires that the agent initiate learning if his expected continuation payoff at time 0 dominates his outside option of immediately stopping: $Q_A(\pi)$. This outside option of never initiating the learning is determined by principal's action that is equal to her *symmetric* belief of π. In what follows, we solve for the equilibrium step-by-step, by backwards induction.

Principal's turn. We first present the best-response reaction of the principal as a result of the optimality conditions required by condition (i). We claim that

$$\boldsymbol{\xi}(\pi, d, \mathcal{C}) = \begin{cases} \pi & d = \texttt{stop} \text{ or } |\partial\mathcal{C} \cap (0,1)| \geq 2 \\ \partial\mathcal{C} \setminus \{0,1\} & \text{o.w} \end{cases} \tag{4.4}$$

It is straightforward to verify (4.4). If $d = \texttt{stop}$ then her optimal decision is to match the action with her initial belief π; if the boundary of $\mathcal{C}$ contains more than two points in $(0,1)$ (e.g., $\pi \in (\alpha, \beta) \subset \mathcal{C}$; $\alpha, \beta \in \partial\mathcal{C}$; and $0 < \alpha < \beta < 1$), then the principal is uncertain with regard to whether $\pi_\tau = \alpha$ or $\pi_\tau = \beta$. Consequently from the Martingale property of (2.1) and as a result of the Optional Stopping theorem, the principal infers that

$$\mathsf{E}_\pi \left[\theta \mid \pi_\tau \in \{\alpha, \beta\} \right] = \alpha \frac{\beta - \pi}{\beta - \alpha} + \beta \frac{\pi - \alpha}{\beta - \alpha} = \pi, \tag{4.5}$$

and takes the action π, again matching her action with her initial belief. Alternatively, if

$|\partial\mathcal{C} \cap (0,1)| = 1$, then the principal knows exactly the agent's posterior at the stopping time and thus takes the corresponding action. Having characterized the principal's equilibrium strategy under no-commitment, we are now ready to discuss the sub-game in which the agent chooses his continuation set conditioned on having decided to initiate the learning process.

Agent's choice of $\mathcal{C}$ in the sub-game $d = \texttt{continue}$. In this sub-game, the agent knows that principal's optimal reaction in the subsequent sub-game follows (4.4). Therefore, if he chooses a *fully-inscribed* continuation set – that is $\mathcal{C} \subset (0,1)$ – his choice is followed by the principal's response of π.[9] And if he chooses a *one-sided* continuation region – namely $(\alpha, 1]$ or $[0, \beta)$ – the principal perfectly elicits his posterior belief at his stopping time, i.e α or β depending on which subset is chosen. We next show a fully inscribed continuation region never satisfies equilibrium requirement (ii).

Toward the contradiction, assume the agent chooses $\mathcal{C} \subset (0,1)$, $\pi \in (\alpha, \beta) \subset \mathcal{C}$, and $\alpha, \beta \in \partial\mathcal{C}$. As outlined above, given this continuation region, the principal will best respond back by taking action π at the time of stopping. Inserting this action in the exit value function of the agent in (4.1), $g_n(\pi_\tau)$ becomes an affine function in π_τ—increasing or decreasing depending on whether the initial belief π is greater or smaller than $b + 1/2$. In either case, such a *monotone* exit value function does not support a fully inscribed open interval as the continuation region. For instance, when g_n is increasing in π_τ, the agent can profitably deviate by slightly lowering α, the lower threshold determining the stopping time, and continuing the learning process over a larger interval. A formal proof ruling out fully inscribed continuation regions is presented in the next lemma.

Lemma 4. *Suppose the principal takes a constant action a, that may or may not be equal to the initial belief π, then the equilibrium continuation set can never have a fully inscribed open component (α, β).*

Proof. Recalling the principles of optimality for stopping time problems, it must be the case that on the continuation region $V_{A,n} > g_n$ and on the stopping region $[0,1] \setminus \mathcal{C}$, $V_{A,n} = g_n$. Further, $\mathbb{K}V_{A,n} \leq 0$ on the entire unit interval.[10] Assume without loss of generality that $a > b + 1/2$ so g_n is increasing in π_τ. Also, toward contradiction suppose (α, β) is an open component of $\mathcal{C}$.[11]

[9] A fully-inscribed set $B \in [0,1]$ is one in which $0, 1 \notin \partial B$. That is, neither ends of the unit interval are on the boundary of the set B.

[10] The reader can consult section 2, specifically equations (2.2.81) and (2.2.92) of Peskir and Shiryaev (2006) for more detailed exposition.

[11] That is for some $x \in (\alpha, \beta) \subset \mathcal{C}$, this interval is the maximal connected subset of $\mathcal{C}$ that contains x.

First, note that $\mathbb{K}g_n(\alpha) = -\rho g_n(\alpha) \leq 0$ because of the aforementioned principles of optimality. So, $V_{A,n}(\alpha) = g_n(\alpha) \geq 0$. Since g_n is increasing and $V_{A,n} > g_n$ on (α, β), then it must be the case that $V_{A,n} > 0$ on this interval. Hence, in light of the HJB equation on (α, β) ($\rho V_{A,n} = \frac{\pi^2(1-\pi)^2}{2\sigma^2} V''_{A,n}$), one deduces the strict convexity of $V_{A,n}$; equivalently, $V''_{A,n} > 0$, on this interval.

Define $\mathfrak{r} := V_{A,n} - g_n$. Then from the principle of smooth fit on the boundaries, we would have $\mathfrak{r}'(\alpha) = \mathfrak{r}'(\beta) = 0$.[12] Therefore, either $\mathfrak{r}' = 0$ on the entire $[\alpha, \beta]$, or there exists a point $x \in (\alpha, \beta)$ such that $\mathfrak{r}''(x) = V''_{A,n}(x) = 0$. The former case cannot happen because then $\mathfrak{r}$ will be constant on $[\alpha, \beta]$ and equals to zero, that is in contrast with $V_{A,n} > g_n \geq 0$ on the open interval. The latter case also is in contrast with the fact that $V''_{A,n}$ is always positive on (α, β). Therefore, $\mathcal{C}$ can never have an open component with both end points differing from $\{0, 1\}$. □

Emboldened by the previous lemma, we can now safely restrict our search for continuation sets that survive conditions (i) and (ii) to one-sided intervals. In this regard, we call $(\alpha, 1]$ the *α-right-sided-interval* and denote it by $\mathcal{R}_\alpha$ and $[0, \beta)$ is called *β-left-sided-interval* and denoted by $\mathcal{L}_\beta$. Therefore, the set of equilibrium continuation sets must belong to $\{\mathcal{R}_\alpha : \alpha \in [0, 1]\} \cup \{\mathcal{L}_\beta : \beta \in [0, 1]\}$.

Proposition 5. *There exists a unique right-sided-interval denoted by $\mathcal{R}_{\alpha_n}$ and a unique left-sided-interval $\mathcal{L}_{\beta_n}$ that satisfy conditions (i) and (ii) of definition 3. In addition, $\alpha_n < 1 - \beta_n$.*

In particular, condition (i) is satisfied if and only if the principal takes action α_n (resp. β_n) when agent's continuation region is $\mathcal{R}_{\alpha_n}$ (resp. $\mathcal{L}_{\beta_n}$). In Figure 3, we plot the agent's payoff functions under the continuation sets $\mathcal{R}_{\alpha_n}$ and $\mathcal{L}_{\beta_n}$: $V_{A,\mathcal{R}}$ and $V_{A,\mathcal{L}}$, respectively. The tangent lines at α_n and β_n depict the agent's exit value functions – when the principal takes action α_n and β_n respectively at the exit time – that in turn induce the continuation value functions $V_{A,\mathcal{R}}$ on $\mathcal{R}_{\alpha_n} = (\alpha_n, 1]$ and $V_{A,\mathcal{L}}$ on $\mathcal{L}_{\beta_n} = [0, \beta_n)$. We refer to $V_{A,\mathcal{R}}$ as the right-sided value function and $V_{A,\mathcal{L}}$ as the left-sided value function.

Recalling equation (3.4), we can thus write

$$V_{A,\mathcal{R}}(\pi) = \left(\kappa - b^2 + \alpha_n^2 - \alpha_n\right) \left(\frac{\pi}{\alpha_n}\right)^{1-\lambda} \left(\frac{1-\pi}{1-\alpha_n}\right)^{\lambda}, \tag{4.6a}$$

$$V_{A,\mathcal{L}}(\pi) = \left(\kappa - b^2 + \beta_n^2 - \beta_n\right) \left(\frac{\pi}{\beta_n}\right)^{\lambda} \left(\frac{1-\pi}{1-\beta_n}\right)^{1-\lambda}. \tag{4.6b}$$

[12] This is a standard result in the stopping time contexts where the underlying process is a diffusion; see section 9.1 of Peskir and Shiryaev (2006).

The right-sided value function is decreasing and the left-sided value function is increasing; thus, the two functions intersect at a unique point, which we denote by η_n. As it can be confirmed from Figure 3, *conditioned* on $d = \texttt{continue}$, the agent's choice of continuation region is $\mathcal{L}_{\beta_n}$ on $[0, \alpha_n)$, $\mathcal{R}_{\alpha_n}$ on $[\alpha_n, \eta_n]$, $\mathcal{L}_{\beta_n}$ on $(\eta_n, \beta_n]$ and $\mathcal{R}_{\alpha_n}$ on $(\beta_n, 1]$.[13]

Having established the strategies adopted by the agent in this sub-game, we now can turn to the final requirement of the sub-game perfect equilibrium: the optimality condition at the root of the game tree, condition (iii).

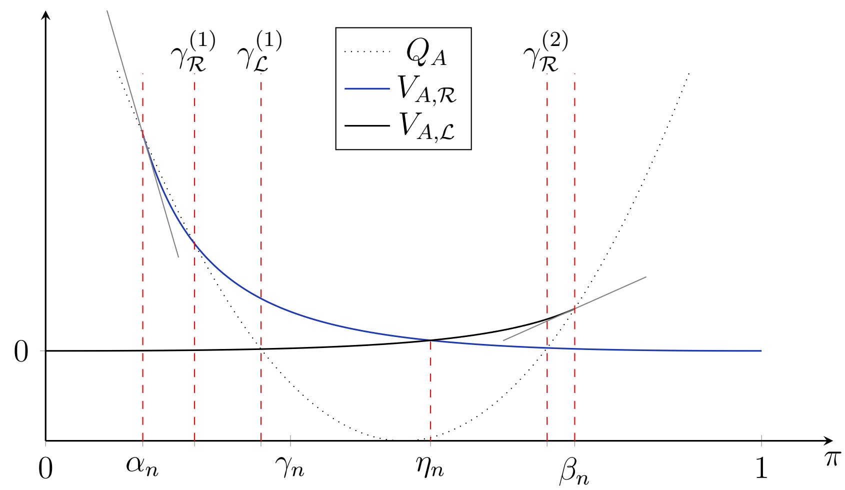

Figure 3: Agent's Value Functions

This figure plots the agent's value functions under two choices of continuation regions: $\mathcal{R}_{\alpha_n}$ (solid blue) and $\mathcal{L}_{\beta_n}$ (solid black). The two tangent lines, positioned *below* the value functions, represent the agent's exit valuation when principal takes action α_n and β_n respectively. The dotted black line represents (Q_A)—the agent's payoff under immediate stopping (and principal taking action π). The intersections of $V_{A,\mathcal{R}}$ ($V_{A,\mathcal{L}}$) with Q_A are listed by the sequence $\gamma_{\mathcal{R}}^{(i)}$ ($\gamma_{\mathcal{L}}^{(i)}$).

Agent's choice of d. At the root of the game-tree, the agent decides whether or not to initiate the learning process under the correct belief that, when continuing, he subsequently selects the one-sided continuation region that, in turn, determines the principal's best-response ξ in (4.4).

When not initiating the learning process, the agent obtains the expected payoff of $Q_A(\pi)$ (defined in condition (iii)) upon immediate delivery of the project to the principal. Therefore, condition (iii) implies that the agent initiates learning whenever $\max\{V_{A,\mathcal{R}}(\pi), V_{A,\mathcal{L}}(\pi)\} >$

[13] At first glance, it may seem surprising that the left-sided continuation region prevails on $[0, \alpha_n)$. But this is simply because on this region, the only continuation region *consistent* with $d = \texttt{continue}$ is the left-sided region $\mathcal{L}_{\beta_n}$. The same logic holds for $\mathcal{R}_{\alpha_n}$ on the interval $(\beta_n, 1]$.

$Q_A(\pi)$. Otherwise, the agent immediately delivers the project to the principal.

This condition adds an element of nonstationarity to the agent's decision. Specifically, at $t = 0$ (and only at $t = 0$) continuation of learning is optimal if the continuation value dominates the *immediate stopping* payoff Q_A as well as the exit value function g_n. Viewing the x axis in Figure 3 as the *initial* belief held by both principal and agent, the agent decides to initiate the learning process on the subset of $\mathcal{R}_{\alpha_n}$ on which $V_{A,\mathcal{R}} > Q_A$. Therefore, the decision to initiate the learning process depends on whether π belongs to that subset or not, whereas the decision to terminate the learning depends on whether π_τ belongs to $\mathcal{R}_{\alpha_n}$. The same logic holds for the subset of initial beliefs in which the left-sided region $\mathcal{L}_{\beta_n}$ prevails.

Next, we characterize the agent's optimal choice of d at the root of the tree. For this we denote the i-th intersection of $V_{A,\mathcal{R}}$ with Q_A by $\gamma_{\mathcal{R}}^{(i-1)}$, letting $\gamma_{\mathcal{R}}^{(0)} = \alpha_n$. The same convention is used to refer to the intersections of $V_{A,\mathcal{L}}$ with Q_A. The agent thus chooses to stop at $t = 0$ for $\pi \in [0, \gamma_{\mathcal{R}}^{(1)})$, continue and select $\mathcal{R}_{\alpha_n}$ on $(\gamma_{\mathcal{R}}^{(1)}, \eta_n)$, continue and select $\mathcal{L}_{\beta_n}$ on (η_n, β_n), and finally stop on $(\beta_n, 1]$. The agent is indifferent on the boundary of these regions, so we break the tie in favor of the principal's payoff.

Despite the fact that principal does not have commitment power, it is nevertheless enlightening to express her expected payoff under both the left- and right-sided regions. In the next subsection, these payoffs will become strategically important since we allow the principal to have commitment power, thereby disciplining the agent's choices of whether to initiate learning and under what conditions to stop the learning process.

In the one-sided continuation regions, there is always a positive probability that $\{\pi_t\}$ hits the end-point that *belongs* to the continuation region (e.g $1 \in \mathcal{R}_{\alpha_n}$ or $0 \in \mathcal{L}_{\beta_n}$) before the other end-point. In such situations, the agent *never* delivers the project to the principal. As such, we need to take a stand on how the principal values *indefinite delay*, and precisely what her payoff is when the project is never delivered to her?

Recall that we assume the principal is infinitely patient. Intuitively, we view her time preference as the limit of finitely patient profiles with diminishing exponential discount rate $\rho_P \to 0$. Thus, it is natural to treat her payoff from an indefinite delay by the agent as zero, since as the stopping time tends towards infinity (with declining, though positive probability), for any fixed discount factor the principal's payoff tends towards zero.

Having specified the principal's payoff from indefinite delay, we can express her expected payoff under each one-sided continuation region. For instance suppose $\mathcal{R}_{\alpha_n}$ is chosen by the agent. Then, he stops learning when either $\pi_\tau = \alpha_n$ or $\pi_\tau = 1$. In the latter case the principal's payoff is zero, because 1 belongs to the continuation region. Even though the agent is entirely confident about the underlying quality of the project, he nevertheless never delivers the project to the principal. In the former case:[14]

$$V_{P,\mathcal{R}}(\pi) = (\alpha_n^2 - \alpha_n)\mathsf{P}\left(\tau_{\alpha_n} < \tau_1\right)$$
$$= (\alpha_n^2 - \alpha_n)\frac{1-\pi}{1-\alpha_n} = -\alpha_n(1-\pi) \tag{4.7}$$

Similarly, when agent selects $\mathcal{L}_{\beta_n}$ as his continuation set, the principal's expected payoff is

$$V_{P,\mathcal{L}}(\pi) = (\beta_n^2 - \beta_n)\mathsf{P}\left(\tau_{\beta_n} < \tau_0\right)$$
$$= (\beta_n^2 - \beta_n)\frac{\pi}{\beta_n} = -(1-\beta_n)\pi. \tag{4.8}$$

Together, the two expressions imply that there exists a cut-off $\gamma_n := \alpha_n/(\alpha_n + 1 - \beta_n)$ such that principal prefers the action α_n to β_n iff $\pi > \gamma_n$. These two value function, together with the principal's payoff under agent's immediate stopping, i.e $Q_P = \pi^2 - \pi$, are depicted in Figure 4.

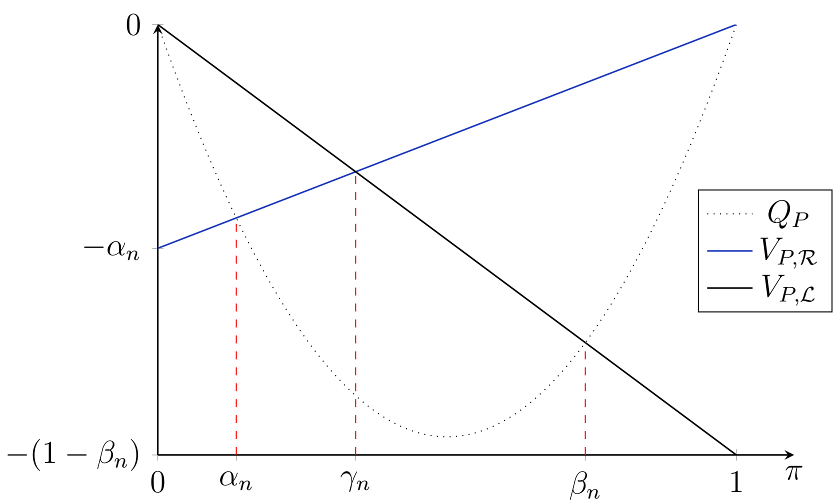

Figure 4: Principal's Value Functions

This figure shows the principal's value function: $V_{P,\mathcal{R}}$ under continuation region $\mathcal{R}_{\alpha_n}$ (solid blue) and $V_{P,\mathcal{L}}$ under continuation region $\mathcal{L}_{\beta_n}$ (solid black), each as a function of the initial belief π. The two graphs are linear, and intersect at $\gamma_n < 0.5$. For low (high) levels of initial belief, the principal prefers the high (low) action. Also, her payoff when the agent stops immediately (Q_P) is shown by the dotted black line.

At this point, we are fully equipped to present the equilibrium outcome of the extensive

[14] Appendix A.4 provides the derivation of the probability of hitting α_n prior to the absorbing state of $\pi_\tau = 1$ conditional on $\theta = 1$. It is straightforward to extend this to derive the unconditional probability of hitting α_n before τ_1.

game in which the principal does not have the ability to make an ex ante commitment to a certain course of action, as represented in Figure 2.

Theorem 6 (Equilibrium for the No-Commitment, No-Communication Game)**.** *In the absence of ex ante principal commitment, there exists a Markov perfect equilibrium*[15]*, in which the agent stops immediately when the initial belief is either sufficiently low or high; otherwise, he specifies a one-sided stopping rule with a low threshold α_n (resp. high threshold β_n) for moderately low (resp. high) initial beliefs.*

In Figure 5 we plot both players' equilibrium value functions in the absence of commitment by the principal. Notably, the principal's payoff features discontinuities, because it is the agent who has the first mover advantage. Accordingly, the agent smooths out his payoff by strategically determining the equilibrium continuation region.

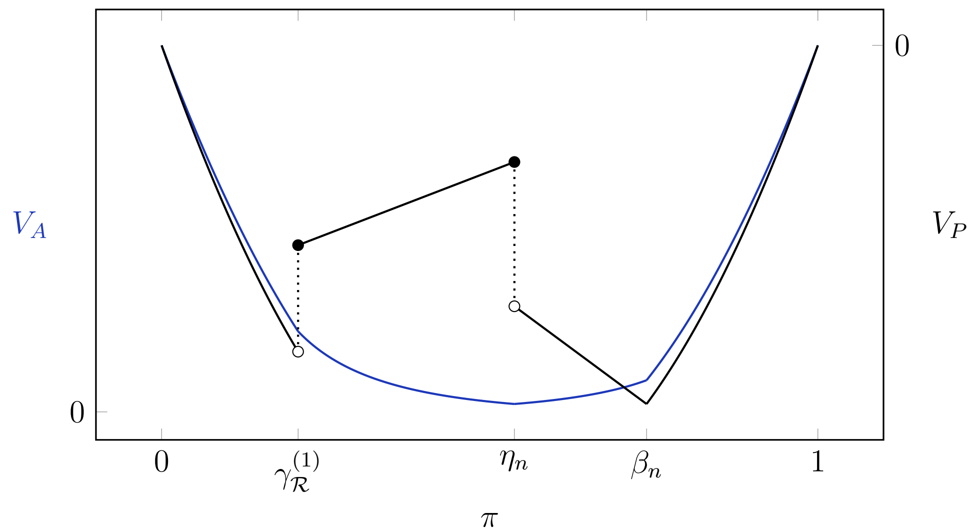

Figure 5: Equilibrium Value Functions when Principal Lacks Commitment Power

This figure represents the *equilibrium* value functions of the agent (solid blue) and the principal (solid black) as a function of the initial belief π in the no-communication game in which the principal lacks commitment power. The agent's payoff function is continuous because he has the first mover advantage, while the principal's payoff function exhibits discontinuities at $\gamma_{\mathcal{R}}^{(1)}$ (where agent switches from immediate stopping to leaning under $\mathcal{R}_{\alpha_n}$) and at η_n (where agent switches from the right sided continuation region $\mathcal{R}_{\alpha_n}$ to the left sided continuation region $\mathcal{L}_{\beta_n}$).

[15]The uniqueness is up to the choice of the continuation region on the boundary points, where we choose to pick the Pareto optimal outcome, thereby breaking the ties in favor of the principal seeing as the agent has already smoothed out his payoff.

4.2 Principal Has Commitment Power

In this part, we ask: What happens when the principal is the first mover and *commits* to a certain action at the stopping time? Specifically, the principal's strategy is the mapping $\boldsymbol{\xi} : [0,1] \times \{\texttt{stop}, \texttt{continue}\} \to [0,1]$, by which, at time 0, she sets a target action that only depends on the initial belief and the agent's choice of whether to initiate learning (d).

In the no-communication game with commitment, the principal signals her commitment power by restricting herself to strategies that do not depend upon the agent's choice of $\mathcal{C}$. We view this as the appropriate strategy space for a principal with commitment power since, in equilibrium, this will lead to a weak improvement in the principal's value function relative to the game in which her strategy is further dependent upon the agent's continuation region.

Following the principal's specification of choice of action at the time of stopping, the game continues with the agent's action i.e $(d, \mathcal{C})$. The agent's strategy mappings are $\boldsymbol{d} : [0,1]^2 \to \{\texttt{stop}, \texttt{continue}\}$ and $\boldsymbol{\mathcal{C}} : [0,1]^2 \to 2^{[0,1]}$, where both maps take in $(\pi, \xi) \in [0,1]^2$ as the input; the first one returns the agent's learning decision at $t = 0$, and the second one returns his choice of continuation region. The game-tree is plotted in Figure 6. In the

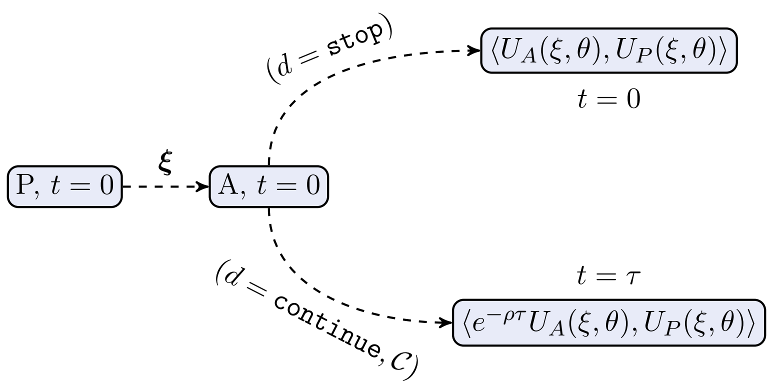

Figure 6: Game tree – with commitment from the principal

following definition we present the requirements of the Markov perfect equilibrium of this game.

Definition 7 (Equilibrium for the Principal-Commitment, No-Communication Game)**.** Given the initial belief π, the tuple of strategy mappings $\langle \boldsymbol{\xi}, (\boldsymbol{d}, \boldsymbol{\mathcal{C}})\rangle$ constitutes a Markov perfect equilibrium when principal holds commitment power if

(i) Conditioned on $d = \texttt{continue}$, $\boldsymbol{\mathcal{C}}(\pi, \xi)$ is the optimal continuation region for every $(\pi, \xi) \in [0,1]^2$.

(ii) At $t = 0$, agent chooses $\boldsymbol{d}(\pi, \xi) = \texttt{continue}$ if

$$\sup_{\tau} \mathsf{E}_{\pi} \left[e^{-\rho\tau} U_A(\xi, \theta) \right] > Q_A(\pi) = \kappa - b^2 + \pi^2 - \pi, \tag{4.9}$$

for every (π, ξ).

(iii) For every π, $\boldsymbol{\xi}(\pi, \texttt{stop}) = \pi$, and

$$\boldsymbol{\xi}(\pi, \texttt{continue}) \in \arg\max_{a} \mathsf{E} \left[-(a-\theta)^2 1_{\{\tau<\infty\}} \mid \pi_{\tau} \in \partial \boldsymbol{\mathcal{C}}(\pi, a) \right]. \tag{4.10}$$

The first two conditions above are natural requirements for the sub-game perfect equilibrium. The third condition, however, is the notable departure from the no-communication game in which the principal does not have commitment power. This third condition requires the principal to specify an optimal action in the root of the game-tree in Figure 6, anticipating, that, at the time of stopping, the agent's posterior belief is on the boundary of the continuation region.

The main pillars behind the characterization of the equilibrium in the no-communication game with principal-commitment have already been developed in lemma 4 and proposition 5. In particular, condition (i) of the definition above together with lemma 4 imply that the agent's best-response mapping, $\boldsymbol{\mathcal{C}}$, must yield one-sided continuation regions.

Intuitively, the last item in definition 7 implies that, through a process of introspection, the principal pre-specifies an exit action, a, that matches the boundary point of the continuation region, $\boldsymbol{\mathcal{C}}(\pi, a)$, subsequently chosen by the agent. In this sense, the principal's promised course of action is self-fulfilling because the agent finds it optimal to stop his learning when his posterior hits the pre-specified benchmark of the principal. The principal, in turn, knows that the posterior belief of the agent at the time of stopping is equal to the promised course of action (which is, in turn, her optimal action).

However, the crucial difference with the no-commitment case is that now, as a result of her first-mover advantage, the principal can effectively choose which of the two continuation sets, $\mathcal{R}_{\alpha_n}$ or $\mathcal{L}_{\beta_n}$, prevails when the agent initiates the learning process. Of course this mandate by the principal can be refused by the agent, in which case the agent never initiates learning; otherwise, he finds it optimal to follow the continuation region whose boundary matches the promised exit action of the principal.

As a result of the first-mover advantage now conferred to the principal, it is the principal's payoff function, depicted in Figure 4, that determines which continuation region prevails in equilibrium. A striking feature implied by this figure is that the principal's preferred continuation region is inversely related to the initial belief π. Specifically, for low levels of the initial belief, the principal prefers the continuation region consistent with a high exit action (for $\pi < \gamma_n$, $V_{P,\mathcal{L}} \geq V_{P,\mathcal{R}}$). Conversely, for high levels of initial belief, the principal prefers the continuation region consistent with the low exit action (for $\pi > \gamma_n$, $V_{P,\mathcal{L}} \leq V_{P,\mathcal{R}}$).

What may seem surprising at first glance is that the principal *always* prefers one of the two continuation regions over immediate stopping (i.e. $\max\{V_{P,\mathcal{R}}, V_{P,\mathcal{L}}\} > Q_P$ for $\pi \in (0,1)$). To understand the intuition behind this result, consider the initial belief $\pi = 0.5 > \gamma_n$, a value for which the right-sided continuation region is preferred by the principal to the left-sided region. By invoking the right-sided continuation region, the principal improves her payoff relative to immediate stopping in two ways: First, by initiating the low-action regime, the principal's ex post payoff conditional on stopping (i.e $\tau_{\alpha_n} < \tau_1$) is improved relative to her ex ante payoff when learning is never initiated. This is obviously the case since Q_P attains its minimum at $\pi = 0.5$. Second, the principal's expected payoff under indefinite delay is higher than what she would attain under immediate stopping (indefinite delay occurs with positive probability when $\theta = 1$; $0 > Q_P$).

To see these two forces formally, decompose the principal's payoff from immediate stopping for $\pi = 0.5 > \gamma_n$:

$$\begin{aligned} Q_P &= (\pi^2 - \pi)\mathsf{P}(\tau_{\alpha_n} < \tau_1) + (\pi^2 - \pi)(1 - \mathsf{P}(\tau_{\alpha_n} < \tau_1)) \\ &< (\alpha_n^2 - \alpha_n)\mathsf{P}(\tau_{\alpha_n} < \tau_1) + (\pi^2 - \pi)(1 - \mathsf{P}(\tau_{\alpha_n} < \tau_1)) \\ &< (\alpha_n^2 - \alpha_n)\mathsf{P}(\tau_{\alpha_n} < \tau_1) + 0(1 - \mathsf{P}(\tau_{\alpha_n} < \tau_1) \\ &= V_{P,\mathcal{R}} \end{aligned} \tag{4.11}$$

where the first line (representing the first force) follows from the fact that $\pi^2 - \pi$ achieves its minimum at $\pi = 0.5$. The second inequality follows from the fact that $\pi^2 - \pi < 0$ (representing the second force).[16]

In the next theorem we use the preference rankings over continuation regions and immediate stopping for both principal and agent to identify the equilibrium outcome of the no-communication game when the principal has commitment power. Our result is stated formally in the following theorem:

[16]This result is qualitatively unchanged when instead modeling the principal's utility function as $U_P(a,\theta) = \phi - (a-\theta)^2$, for some $\phi > 0$. In this general setting, the principal will prefer Q_P to either continuation region for sufficiently low or high initial beliefs, since indefinite delay now incorporates the opportunity cost ϕ of never taking an action. Qualitatively, however, the model is unchanged.

Theorem 8 (Equilibrium for the Principal-Commitment, No-Communication Game)**.** *There exists a generically*[17] *unique Markov perfect equilibrium satisfying the conditions listed in definition 7.*

Sketch of the proof. In Figure 3 we denote the i-th intersection of $V_{A,\mathcal{R}}$ with Q_A by $\gamma_{\mathcal{R}}^{(i-1)}$ in which $\gamma_{\mathcal{R}}^{(0)} = \alpha_n$. A similar convention is used for the intersections of $V_{A,\mathcal{L}}$ and Q_A. Given the agent's and principal's payoff functions over the entire belief region, depicted respectively in Figure 3 and Figure 4, we express the equilibrium behavior over different intervals. The relative ranking of intersection points, namely $\{\gamma_{\mathcal{R}}^{(1)}, \gamma_{\mathcal{L}}^{(1)}, \gamma_n, \gamma_{\mathcal{R}}^{(2)}\}$, determines whic continuation region is implemented in the equilibrium.[18] Below, we deliberate on the ordering $\gamma_{\mathcal{R}}^{(1)} \leq \gamma_{\mathcal{L}}^{(1)} \leq \gamma_n \leq \gamma_{\mathcal{R}}^{(2)}$.

(i) – $[0, \gamma_{\mathcal{R}}^{(1)})$: regardless of principal's choice, the immediate stopping over this region is the agent's dominant choice.

(ii) – $[\gamma_{\mathcal{R}}^{(1)}, \gamma_{\mathcal{L}}^{(1)})$: on this region the principal's preference ranking is $V_{P,\mathcal{L}} \geq V_{P,\mathcal{R}} \geq Q_P$. If she specifies the left-sided continuation region, with action β_n, the agent responds by immediately stopping, since on this interval $V_{A,\mathcal{R}} < Q_A$. Therefore, the principal commits to the right-sided continuation region, with stopping action α_n; the agent in turn initiates the learning process, subject to the continuation region $\mathcal{R}_{\alpha_n}$.

(iii) – $[\gamma_{\mathcal{L}}^{(1)}, \gamma_n)$: on this region the principal can induce her most favorable action, i.e β_n, because the agent no longer has the credible threat of immediate stopping; i.e., $V_{A,\mathcal{L}} \geq Q_A$. Therefore, the principal chooses stopping action β_n and the agent initiates the learning process subject to the continuation set $\mathcal{L}_{\beta_n}$.

(iv) – $[\gamma_n, \gamma_{\mathcal{R}}^{(2)}]$: the principal's dominant action is α_n, and the agent naturally initiates learning under the right-sided continuation region because $V_{A,\mathcal{R}} \geq Q_A$. Therefore, the principal chooses α_n and the agent performs learning subject to the continuation set $\mathcal{R}_{\alpha_n}$.

(v) – $(\gamma_{\mathcal{R}}^{(2)}, \beta_n)$: even though the principal prefers α_n on this region, the agent now has the credible threat of immediately stopping should the principal specify action α_n. Therefore, the principal capitulates to the agent's threat by taking action β_n. In turn, the agent performs the learning process subject to $\mathcal{L}_{\beta_n}$.

(vi) – $[\beta_n, 1]$: The agent prefers immediate stopping regardless of the stopping action specified by the principal. Thus, in equilibrium the agent never initiates learning and the principal takes action π.

[17] Ties are broken in favor of the Pareto dominating outcome.

[18] Reducing the number of possibilities, recall that it is always the case by definition that $\gamma_{\mathcal{R}}^{(1)} < \gamma_{\mathcal{R}}^{(2)}$.

The equilibrium pattern characterized above is generically unique on the interval $[0,1]$, with multiple equilibria at the boundary points of the intervals. To rule out multiple equilibria, in each case we choose the Pareto dominant equilibrium. □

Table 1 below summarizes the equilibrium strategy of each party under each range of prior beliefs:

	$[0,\gamma_{\mathcal{R}}^{(1)})$	$[\gamma_{\mathcal{R}}^{(1)},\gamma_{\mathcal{L}}^{(1)})$	$[\gamma_{\mathcal{L}}^{(1)},\gamma_n)$	$[\gamma_n,\gamma_{\mathcal{R}}^{(2)}]$	$(\gamma_{\mathcal{R}}^{(2)},\beta_n)$	$[\beta_n,1]$
Agent	$d=\texttt{stop}$	$d=\texttt{cont}$ $\mathcal{C}=\mathcal{R}_{\alpha_n}$	$d=\texttt{cont}$ $\mathcal{C}=\mathcal{L}_{\beta_n}$	$d=\texttt{cont}$ $\mathcal{C}=\mathcal{R}_{\alpha_n}$	$d=\texttt{cont}$ $\mathcal{C}=\mathcal{L}_{\beta_n}$	$d=\texttt{stop}$
Principal's action	π if $d=\texttt{stop}$ β_n if $d=\texttt{cont}$	π if $d=\texttt{stop}$ α_n if $d=\texttt{cont}$	π if $d=\texttt{stop}$ β_n if $d=\texttt{cont}$	π if $d=\texttt{stop}$ α_n if $d=\texttt{cont}$	π if $d=\texttt{stop}$ β_n if $d=\texttt{cont}$	π if $d=\texttt{stop}$ α_n if $d=\texttt{cont}$

Table 1: Equilibrium strategy pair

The equilibrium value functions in the absence of communication but with the principal having commitment power are plotted in Figure 7. In contrast to Figure 1, the value functions for both agent and principal now feature discontinuities.

These discontinuities for both parties arise because of the interaction between the principal's commitment power and the agent's threat of never initiating the learning process (which is at times credible and non-credible). The first type of discontinuity for the principal arises because for some prior beliefs the agent prefers immediate stopping to either continuation region. For such initial beliefs ($\pi \notin [\gamma_{\mathcal{R}}^{(1)},\beta_n]$), the agent credibly threatens to stop learning immediately regardless of which continuation region the principal recommends.

For all other beliefs, the principal prefers either of the two continuation regions over immediate stopping (for intuition see equation (4.11)). Whether or not the principal is able to induce the agent to adopt her preferred continuation region depends upon the credibility of the agent's threat to never initiate learning. Discontinuities in the value functions arise for both principal and agent at those values for which the agent's threat of immediate stopping under the principal's preferred continuation region switches from credible to non-credible (and vice versa).

Remark 9. There are certain equilibrium features that are robust under other orderings of intersection points. For example, if $\gamma_{\mathcal{R}}^{(2)} > \beta_n$, the discontinuities in value functions at $\gamma_{\mathcal{R}}^{(2)}$ occur at a value greater than β_n. However, the prediction that moderately high levels of belief, namely $\pi \in (\gamma_n,\gamma_{\mathcal{R}}^{(2)})$, induce the low action in equilibrium (α_n) is robust. As another case, the prediction that there exists a range of moderately low beliefs on which the high action β_n unfolds remains robust as long as $\gamma_{\mathcal{L}}^{(1)} < \gamma_n$. A more comprehensive analysis connecting the primitives (κ, b, λ) to the orderings of the equilibrium intersections is highly

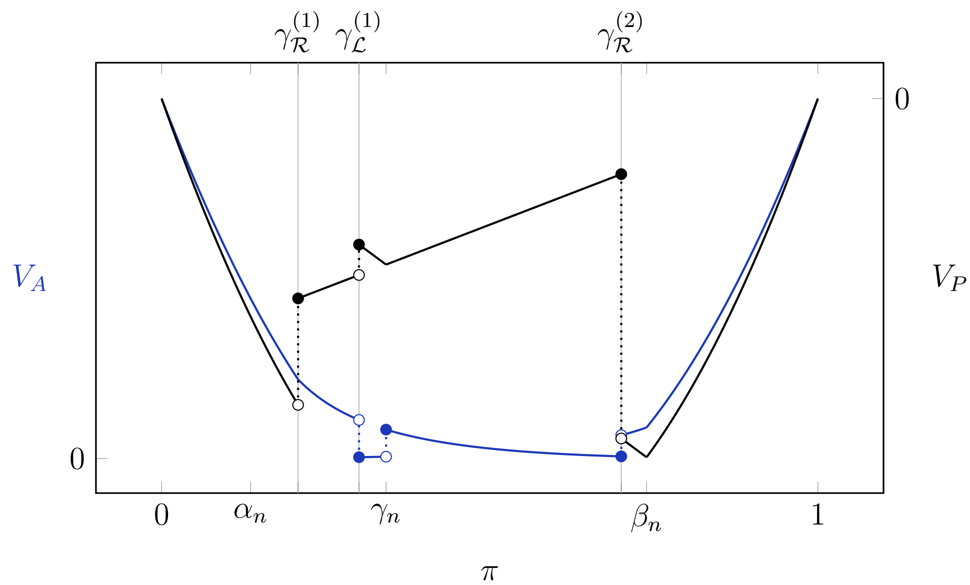

Figure 7: Equilibrium Value Functions with No Communication

This figure plots the players' *equilibrium* value functions in the no-communication game when principal has commitment power, each as a function of the initial belief π. The agent's value function (V_A; solid blue) exhibits discontinuities because, with commitment power, the principal has the first mover advantage. This commitment power affords the principal the privilege of selecting which continuation region prevails in equilibrium, which in turn prevents the agent from value matching. In addition, the principal's payoff (V_P; solid black) has discontinuous points at $\{\gamma_{\mathcal{R}}^{(1)}, \gamma_{\mathcal{L}}^{(1)}, \gamma_{\mathcal{R}}^{(2)}\}$ because at times the agent's threat of immediate stopping is credible, consequently forcing the principal to propose her *second most preferred* choice

intractable, and not quite fruitful in terms of equilibrium robust predictions. In the next section we offer individual welfare predictions that are robust across equilibrium orderings.

5 The value of Non-Communication

5.1 Comparison of Payoffs Under Each Communication Regime

In this subsection we compare the payoffs of both principal and agent in the full-communication game to their payoffs in the no-communication game with principal-commitment. Figure 8 plots the full communication payoffs (indexed by c; dashed lines) against the non-communication payoffs (indexed by n; solid lines).

When communication is infeasible and the principal has commitment power, the agent is weakly worse off (for intermediate prior beliefs strictly so), whereas in most regions the principal is better off. Thus one can see why in many instances the principal would prefer not to facilitate formal communication with the agent and instead pre-specify an action at the time of stopping to discipline the ex post quality of projects delivered by the agent in equilibrium. In the next two propositions we formalize the two welfare comparisons suggested by this figure, first for the agent and then for the principal.

Proposition 10. *The agent is always worse off in the absence of communication.*

Proof. Note that for every $\pi \in [0,1]$, $V_{A,n}(\pi) \in \{Q_A(\pi), V_{A,\mathcal{R}}(\pi), V_{A,\mathcal{L}}(\pi)\}$ where $\mathcal{R} = \mathcal{R}_{\alpha_n}$ and $\mathcal{L} = \mathcal{L}_{\beta_n}$. The agent's value function in the full communication follows (3.2):

$$V_{A,c}(\pi) = \sup_{\tau} \mathsf{E}_\pi \left[e^{-\rho\tau} \left(\kappa - b^2 + \pi_\tau^2 - \pi_\tau \right) \right] \tag{5.1}$$

Setting $\tau = 0, \tau = \tau_{\alpha_n}$ and $\tau = \tau_{\beta_n}$ in the *rhs* of the above equation implies respectively that $V_{A,c}(\pi) \geq Q_A(\pi), V_{A,c}(\pi) \geq V_{A,\mathcal{R}}(\pi)$ and $V_{A,c}(\pi) \geq V_{A,\mathcal{L}}(\pi)$. □

Proposition 11. *There exists a function $\underline{b}(\cdot) \geq 0$ such that for all $b \in (\underline{b}(\kappa, \lambda), \sqrt{\kappa})$*[19]

(i) $\alpha_n \leq \alpha_c$*,*

(ii) the principal is better off in the absence of communication on the intermediate belief region $(\alpha_c \vee \gamma_{\mathcal{R}}^{(1)}, \gamma_{\mathcal{R}}^{(2)} \wedge \beta_n)$*.*

The essential content behind this proposition is that as the bias between the two parties widens, in equilibrium the principal commits to a more conservative action under the right sided continuation region (i.e., lowering α_n) relative to the lower threshold chosen by the agent under full communication (α_c). This occurs because the action that maximizes the agent's utility is (a) always greater than the action which maximizes the principal's utility (b) increasing in b.

5.2 Project Realization and Indefinite Delay

As a relaxed notion of efficiency, we examine the conditional likelihood under the right-sided continuation region that a project will never be undertaken despite its high quality; that is, $\mathsf{P}\left(\tau_1 < \tau_\alpha \mid \theta = 1\right)$, where again τ_1 is the first time that the posterior belief hits one. Recall the law of motion for the posterior belief:

$$\mathrm{d}\pi_t = \sigma^{-1}\pi_t(1-\pi_t)\mathrm{d}\bar{B}_t = \sigma^{-2}\pi_t(1-\pi_t)\left(\mathrm{d}x_t - \pi_t \mathrm{d}t\right) \tag{5.2}$$

[19]Recall that b is a parameter which governs the misalignment of the agent's and principal's preferences.

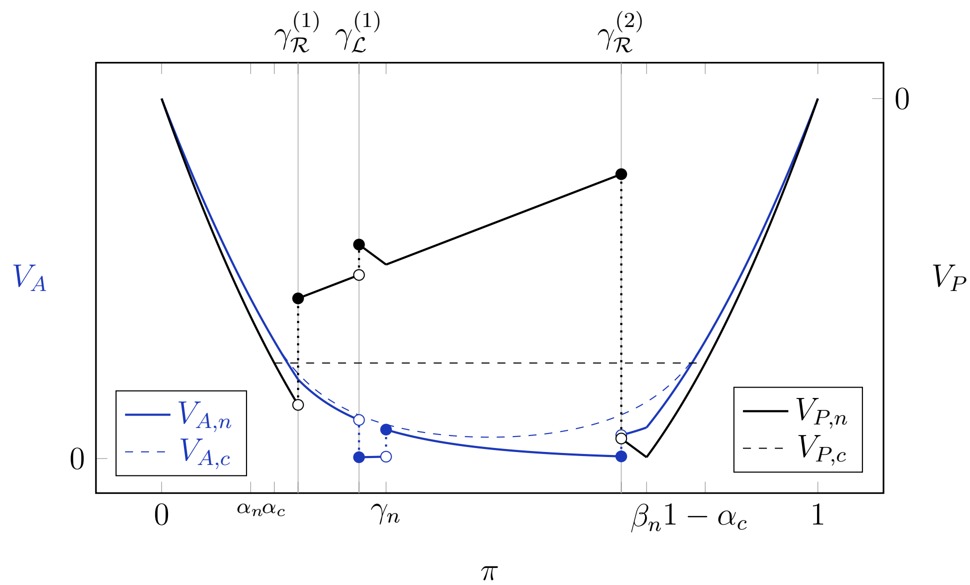

Figure 8: Individual Payoff Comparison

This figure overlays the players' value functions under full communication (dotted lines) on their value functions under no communication and principal commitment (solid lines). The agent's value functions are colored in blue, and the principal's are colored in black. In the intermediate belief region, the principal is better off without communication and the agent is worse off. For sufficiently low or high initial beliefs, the agent never initiates learning in either the full- or no-communication game; thus, both parties are indifferent between either of the two communication environments.

Therefore, conditioned on $\theta = 1$,

$$\mathrm{d}\pi_t = \sigma^{-1}\pi_t(1-\pi_t)\left(\frac{1-\pi_t}{\sigma}\mathrm{d}t + \mathrm{d}B_t\right). \tag{5.3}$$

Define $q(\pi) = \mathsf{P}_\pi\left(\tau_1 < \tau_\alpha \mid \theta = 1\right)$. With $\{\pi_t\}$ following the above law of motion, an infinitesimal analysis of $q(\pi_t)$ amounts to the ODE $q'(\pi) + \pi q''(\pi)/2 = 0$ (see appendix A.4), with the general solution $q(\pi) = q_0/\pi + q_1$. The coefficients $\{q_0, q_1\}$ are easily solved for using the boundary conditions $q(\alpha_n) = 0$ and $q(1) = 1$.

When $\pi \in (\gamma_{\mathcal{R}}^{(1)}, \gamma_{\mathcal{L}}^{(1)}) \cup (\gamma_n, \gamma_{\mathcal{R}}^{(2)})$, the equilibrium continuation region is $\mathcal{R}_{\alpha_n}$. Thus, as implied by equation (5.3), the probability of the low action α_n being undertaken conditioned on $\theta = 1$ is

$$\mathsf{P}_\pi\left(\tau_{\alpha_n} \leq \tau_1 \mid \theta = 1\right) = \frac{\alpha_n(1-\pi)}{\pi(1-\alpha_n)}, \tag{5.4}$$

and with complementary probability

$$\mathsf{P}_\pi\left(\tau_{\alpha_n} > \tau_1 \mid \theta = 1\right) = \frac{\pi - \alpha_n}{\pi(1-\alpha_n)} \tag{5.5}$$

the project is never delivered by the agent to the principal.

The first possibility, that a low action is taken when the project type is high, could also arise in the perfect communication case, in which a stream of bad news lowers the agent's posterior belief to his lower threshold, at which point he stops learning and the principal takes the lower action. In contrast, the second event, in which the good project is never carried out, only occurs in the absence of communication.

For $\theta = 1$ and $\pi \in (\gamma_{\mathcal{L}}^{(1)}, \gamma_n)$, the left sided continuation region is adopted by the agent in equilibrium. Thus, the path of $\{\pi_t\}$ almost surely ends up at β_n, even though the initial belief that the project is a high type is less than half. Similarly, when $\theta = 1$ and $\pi \in (\gamma_{\mathcal{R}}^{(2)}, \beta_n)$ the prevailing continuation region is again $\mathcal{L}_{\beta_n}$, so that a high type project eventually hits the upper threshold β_n, at which point the principal takes the high action.

A similar analysis can be done for a low type project, i.e $\theta = 0$. When the initial belief is $\pi \in (\gamma_{\mathcal{R}}^{(1)}, \gamma_{\mathcal{L}}^{(1)}) \cup (\gamma_n, \gamma_{\mathcal{R}}^{(2)})$, the continuation region $\mathcal{R}_{\alpha_n}$ is adopted by the agent in equilibrium. Thus, when the project is the low type, almost surely $\{\pi_t\}$ hits the lower boundary α_n. However, for $\pi \in (\gamma_{\mathcal{L}}^{(1)}, \gamma_n) \cup (\gamma_{\mathcal{R}}^{(2)}, \beta_n)$ the equilibrium continuation region is $\mathcal{L}_{\beta_n}$, so that, conditioned on $\theta = 0$, the law of motion for $\{\pi_t\}$ is

$$\mathrm{d}\pi_t = \frac{\pi_t(1-\pi_t)}{\sigma}\left(\frac{-\pi_t}{\sigma}\mathrm{d}t + \mathrm{d}B_t\right). \tag{5.6}$$

As a result,

$$\begin{aligned} \mathsf{P}_\pi\left(\tau_{\beta_n} \le \tau_0 \,|\, \theta = 0\right) &= \frac{\pi(1-\beta_n)}{\beta_n(1-\pi)}, \\ \mathsf{P}_\pi\left(\tau_{\beta_n} > \tau_0 \,|\, \theta = 0\right) &= \frac{\beta_n - \pi}{\beta_n(1-\pi)}. \end{aligned} \tag{5.7}$$

Again the first event, hitting the upper boundary point consistent with a high action even though the project is a low type, could also occur in the full communication case: a stream of positive signals increases the posterior belief of a low type project sufficiently so that the agent stops learning at the upper threshold. However, the second event – in which the low type project is never undertaken (with a low action) – only happens in the absence of communication.

Given the expressions discussed above, the probability of indefinite delay – in which the agent never ceases learning and the principal never takes an action – can be drawn as a function of the initial belief, which is what is shown in Figure 9. At the two ends of the belief spectrum, the chance of indefinite delay is zero, because the agent immediately delivers the project to the principal. However, for most intermediate values of the prior belief π, where the equilibrium continuation region is $\mathcal{R}_{\alpha_n}$, the likelihood of indefinite delay increases as a function of the initial belief. While this may seem counterintuitive – since one might expect that projects with better initial assessments are carried out earlier – due to the non-credible nature of communication, an increasing share of such projects are doomed to inaction.

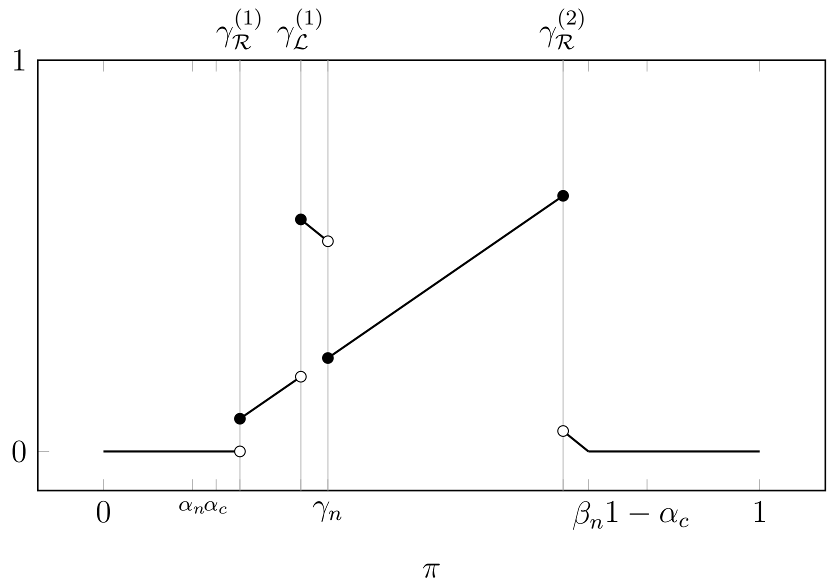

Figure 9: Probability of Indefinite Delay

This figure represents the probability of indefinite delay as a function of the initial belief π when communication is not feasible and the principal has commitment power. At the two ends of the belief interval, the probability of indefinite delay is zero because the agent never initiates learning and the principal immediately takes an action. In the intermediate region, this probability is increasing in π whenever the right sided continuation region $\mathcal{R}_{\alpha_n}$ prevails in equilibrium. Otherwise, on those intervals in which $\mathcal{L}_{\beta_n}$ prevails in equilibrium, the probability of indefinite delay is decreasing in π. In the particular parameterization plotted here, $q_\pi(\gamma_{\mathcal{R}}^{(2)}) \approx 0.65$, indicating that, at times, indefinite delay can occur for the majority of projects in equilibrium.

6 Conclusion

In this paper we study the implications of delegated learning under credible and non-credible communication. By studying these two extreme cases, we are able to abstract away from the well-studied issue of strategic communication between principal and agent.

Surprisingly, the non-credible communication equilibrium is preferred by the principal for intermediate ranges of prior beliefs about the true state of the world. In this equilibrium, the principal pre-specifies a course of action that she will take once approached by the agent.

This, in turn, induces an equilibrium in which, by virtue of approaching the principal, the agent fully reveals his privately observed posterior belief.

The non-credible communication equilibrium, moreover, produces two interesting implications relative to the full communication equilibrium. First, there is scope for indefinite delay on the part of the agent and, in turn, a lack of action by the principal. This never occurs in the full-communication environment. Second, under non-credible communication, the equilibrium strategies of both principal and agent are discontinuous functions of the prior beliefs of both parties about the true state of the world.

A Proofs

A.1 Proof of proposition 1

Existence. Recall the general form of solution to the HJB equation in (3.4). In the case of perfect communication of the posterior belief, the value of stopping is $g_c(\pi_\tau) = \kappa - b^2 + \pi_\tau^2 - \pi_\tau$. Given the convex shape of exit value function g, one would guess that the continuation region for the agent's stopping time problem follows an inscribed open interval in $[0,1]$, denoted by (α, β). On the boundaries of such interval the standard principles of continuous and smooth fit are held. So, there are four equations available resulting from continuous and smooth fit at α and β, together with four unknowns $\{\alpha, \beta, c_1, c_2\}$. From continuous fit conditions:

$$c_1\alpha^{1-\lambda}(1-\alpha)^{\lambda} + c_2\alpha^{\lambda}(1-\alpha)^{1-\lambda} = \kappa - b^2 + \alpha^2 - \alpha \tag{A.1a}$$

$$c_1\beta^{1-\lambda}(1-\beta)^{\lambda} + c_2\beta^{\lambda}(1-\beta)^{1-\lambda} = \kappa - b^2 + \beta^2 - \beta \tag{A.1b}$$

And from the smooth-pasting relations at α and β:

$$c_1(1-\lambda)\alpha^{-\lambda}(1-\alpha)^{\lambda} - c_1\lambda\alpha^{1-\lambda}(1-\alpha)^{\lambda-1} + c_2\lambda\alpha^{\lambda-1}(1-\alpha)^{1-\lambda} - c_2(1-\lambda)\alpha^{\lambda}(1-\alpha)^{-\lambda} = 2\alpha - 1 \tag{A.2a}$$

$$c_1(1-\lambda)\beta^{-\lambda}(1-\beta)^{\lambda} - c_1\lambda\beta^{1-\lambda}(1-\beta)^{\lambda-1} + c_2\lambda\beta^{\lambda-1}(1-\beta)^{1-\lambda} - c_2(1-\lambda)\beta^{\lambda}(1-\beta)^{-\lambda} = 2\beta - 1 \tag{A.2b}$$

We conjecture that $\alpha = 1 - \beta$ and $c_1 = c_2 = \mathfrak{c}$, therefore the above four equations are summarized to the following two:

$$\mathfrak{c}\alpha^{1-\lambda}(1-\alpha)^{\lambda} + \mathfrak{c}\alpha^{\lambda}(1-\alpha)^{1-\lambda} = \kappa - b^2 + \alpha^2 - \alpha \tag{A.3a}$$

$$\mathfrak{c}(1-\lambda)\alpha^{-\lambda}(1-\alpha)^{\lambda} - \mathfrak{c}\lambda\alpha^{1-\lambda}(1-\alpha)^{\lambda-1} + \mathfrak{c}\lambda\alpha^{\lambda-1}(1-\alpha)^{1-\lambda} - \mathfrak{c}(1-\lambda)\alpha^{\lambda}(1-\alpha)^{-\lambda} = 2\alpha - 1 \tag{A.3b}$$

Let us call the odd ratio at α by $m = \alpha/(1-\alpha)$, then $\alpha = m/(1+m)$ and $1-\alpha = 1/(1+m)$, so

$$\mathfrak{c}m^{1-\lambda} + \mathfrak{c}m^{\lambda} = (1+m)(\kappa - b^2) - \frac{m}{1+m} \tag{A.4a}$$

$$\mathfrak{c}(1-\lambda)m^{-\lambda} - \mathfrak{c}\lambda m^{1-\lambda} + \mathfrak{c}\lambda m^{\lambda-1} - \mathfrak{c}(1-\lambda)m^{\lambda} = \frac{m-1}{m+1}. \tag{A.4b}$$

Substituting m^{λ} from (A.4a) into (A.4b) and multiplying both sides by m lead to

$$c(2\lambda-1)m^{2-\lambda}c(\lambda-1)m^{1-\lambda} - c\lambda m^{\lambda} = (\lambda-1)m(1+m)(\kappa-b^2) - \frac{m(\lambda m-1)}{1+m}. \tag{A.5}$$

Now replace m^{λ} from (A.4a) into the above expression and obtain

$$\mathfrak{c}(2\lambda-1)m^{1-\lambda} = (\kappa-b^2)(\lambda m + \lambda - m) - \frac{\lambda m}{m+1} + \frac{m}{(m+1)^2}. \tag{A.6}$$

Dividing corresponding sides of the above relation by that of (A.4a) implies that the solution m must satisfy

$$1 + m^{2\lambda-1} = (2\lambda-1)\frac{(m+1)^2(\kappa-b^2) - m}{(m+1)(\kappa-b^2)(\lambda m+\lambda-m) - \lambda m + \frac{m}{m+1}}. \tag{A.7}$$

At $m = 0$, the *lhs* of the above relation is strictly smaller than *rhs*. Further, at $m = 1$, the *lhs* is equal to 2, while the *rhs* at $m = 1$ is equal to $2/(2\lambda-1)$. Because $\lambda > 1$, the *lhs* value of 2 is always greater than the *rhs* at $m = 1$. This implies that there exists $m_c \in (0,1)$ satisfying (A.7) and thus the system in (A.4). So there exists $\alpha_c \in (0, 1/2)$ satisfying (A.1) and (A.2).

Uniqueness. We prove that m_c is indeed the unique solution to (A.7). Toward this, first note that the *lhs* of this equation is increasing in m on positive reals, hence it suffices to prove the *rhs* is decreasing in m. Let's denote the expression on the *rhs* of (A.7) excluding

$(2\lambda - 1)$ by M, then

$$\frac{\mathrm{d}M}{\mathrm{d}m} \propto (m+1)^4(\kappa - b^2)^2 - (m^2-1)^2(\kappa - b^2) - m^2 \equiv N, \tag{A.8}$$

where the constant of proportionality is positive. Note that N is convex in $\kappa - b^2$, therefore its maximum is achieved at the extremes where $\kappa - b^2 \in \{0, 1/4\}$. Obviously, when $\kappa - b^2 = 0$, then $N \leq 0$. At $\kappa - b^2 = 1/4$:

$$N = \frac{(m+1)^4}{16} - \frac{(m^2-1)^2}{4} - m^2 = -\frac{(m+1)^2(m-3)(3m-1)}{16} - m^2 \tag{A.9}$$

Then one can easily check the above expression achieves its maximum over $[0,1]$ at $m = 1$ that is equal to 0. Therefore, $\mathrm{d}M/\mathrm{d}m \leq 0$, and hence the solution shown above is unique.

The verification step, showing that the above continuation set $(\alpha_c, 1-\alpha_c)$, is indeed the optimal continuation set is pretty standard, therefore we omit that and refer the reader to the verification methods developed in Peskir and Shiryaev (2006). □

A.2 Proof of proposition 5

Unique existence of $\mathcal{R}_{\alpha_n}$. Let us first prove the unique existence of a right-sided interval satisfying the stated two conditions. Let $\mathcal{R}_\alpha = (\alpha, 1] = \mathcal{C}$ be a candidate continuation set. Upon the stopping the principal knows that $\pi_\tau = \alpha$, because $\partial\mathcal{C}$ is a singleton. Therefore, she takes action α leading to the following exit value function for the agent:

$$g_n(\pi_\tau) = \kappa + [2(\alpha - b) - 1]\,\pi_\tau - (\alpha - b)^2 \tag{A.10}$$

Because $\mathcal{C}$ is right-sided then g_n must be decreasing in π_τ, namely stopping at the low levels of belief and continuing in the larger levels. This amounts to $\alpha < b + 1/2$.[20] Further, since $1 \in \mathcal{C}$, the coefficient c_2 in (3.4) must be zero and we let $c_1 = c$. Because of condition (ii), $V_{A,n}$ has to satisfy continuous and smooth fit at the boundary α:

$$\begin{aligned} c\alpha^{1-\lambda}(1-\alpha)^\lambda &= \kappa + (2(\alpha - b) - 1)\,\alpha - (\alpha - b)^2 \\ c(1-\lambda)\alpha^{-\lambda}(1-\alpha)^\lambda - c\lambda\alpha^{1-\lambda}(1-\alpha)^{\lambda-1} &= (2(\alpha - b) - 1) \end{aligned} \tag{A.11}$$

[20] In the case of $\alpha = b + 1/2$, the function g_n is constant and particularly negative. Therefore, the agent would rather to continue forever than stopping in finite time and incur a negative payoff. So this case never supports a proper continuation set, and hence is eliminated from the analysis.

After some regroupings one gets

$$\frac{\lambda}{1-\alpha} = \frac{\kappa - (\alpha - b)^2}{\kappa - (\alpha - b)^2 + (2(\alpha - b) - 1)\,\alpha}. \tag{A.12}$$

So the candidate α must be a root to the following cubic polynomial that should exist in $[0, b + 1/2)$:

$$\mathfrak{R}(\alpha) := \alpha^3 - (1 + 2b + \lambda)\,\alpha^2 + \big(b^2 + 2b + \lambda - \kappa\big)\,\alpha + (b^2 - \kappa)(\lambda - 1) \tag{A.13}$$

Note that $\mathfrak{R}(0) = -(\lambda - 1)(\kappa - b^2) < 0$, $\mathfrak{R}(b + 1/2) = -(\kappa - 1/4)(\lambda + b - 1/2) > 0$ and $\mathfrak{R}(1) = -\lambda(\kappa - b^2) < 0$. Using the intermediate value theorem together with the fact that $\lim_{x\to\infty} \mathfrak{R}(x) = \infty$ implies that $\mathfrak{R}$ has a *unique* root in the interval $(0, b + 1/2)$. Therefore, one can always find a unique α that solves the system (A.11). The *verification* step, showing the above α solving the system (A.11) does indeed give rise to the optimal continuation set $\mathcal{R}_\alpha$, is standard and hence omitted from the proof.

Unique existence of $\mathcal{L}_{\beta_n}$. Now we examine a generic left-sided interval $[0, \beta) = \mathcal{L}_\beta$ as a candidate for the continuation set. Naturally, this would be the case only if g_n is increasing, so we require $\beta > b + 1/2$. Analogous to what was presented above one gets the following system for the optimal β:

$$\begin{aligned} c\beta^\lambda(1-\beta)^{1-\lambda} &= \kappa + (2(\beta - b) - 1)\,\beta - (\beta - b)^2 \\ c\lambda\beta^{\lambda-1}(1-\beta)^{1-\lambda} - c(1-\lambda)\beta^\lambda(1-\beta)^{-\lambda} &= 2(\beta - b) - 1. \end{aligned} \tag{A.14}$$

That after some manipulation amounts to

$$\frac{(2(\beta - b) - 1)\,\beta}{(\beta - b)^2 - \kappa} = \frac{\lambda - \beta}{\lambda - 1} \Leftrightarrow \beta = \frac{\lambda\,(\kappa - (\beta - b)^2) + (\lambda - 1)\beta\,(2(\beta - b) - 1)}{\kappa - (\beta - b)^2}. \tag{A.15}$$

Hence the optimal β must be a solution to the following cubic polynomial:

$$\mathfrak{L}(\beta) := \beta^3 - (2(b+1) - \lambda)\,\beta^2 + \big((b+1)^2 - \kappa - \lambda\big)\,\beta + \lambda(\kappa - b^2) = 0 \tag{A.16}$$

A sign determination exercise implies that $\mathfrak{L}(0) = \lambda(\kappa - b^2) > 0$, $\mathfrak{L}(b + 1/2) = (\kappa - 1/4)(\lambda - b - 1/2)$ and $\mathfrak{L}(1) = (\lambda - 1)(\kappa - b^2) > 0$. Since $b^2 < 1/4$ and $\lambda > 1$ then $\lambda > b + 1/2$. Hence, applying the intermediate value theorem combined with $\lim_{x\to-\infty} \mathfrak{L}(x) = -\infty$ imply that there must a exist a unique root of $\mathfrak{L}$ in the region $(b + 1/2, 1)$.

Proof for $\alpha_n < 1-\beta_n$. To show the last part of the proposition, i.e $\alpha_n < 1-\beta_n$, let us plug $1-\beta_n$ instead of α_n in equation (A.13). After some simplifications one obtains

$$\mathfrak{R}(1-\beta_n) = -\beta_n^3 + \beta_n^2\,(2-2b-\lambda) + \beta_n\left(-1+2b-b^2+\kappa+\lambda\right) - \lambda(\kappa-b^2). \tag{A.17}$$

Recall that β_n is a solution to (A.16), thus one can leverage this and replace β_n^3 from that equation into the above one, and get

$$\mathfrak{R}(1-\beta_n) = 4b\beta_n(1-\beta_n) > 0. \tag{A.18}$$

Given that α_n is the unique root of $\mathfrak{R}=0$ in the region $(0, b+1/2)$ and $\mathfrak{R}(0)<0$, we can now conclude that $1-\beta_n$ must be larger than α_n. □

A.3 Proof of proposition 11

We first prove the existence of $\underline{b}(\cdot)$ such that $\alpha_n < \alpha_c$ for all $b > \underline{b}(\kappa,\lambda)$. For minimizing the use of variables define $\Delta := 1-4(\kappa-b^2)$ and let $\nu := \frac{1-\sqrt{\Delta}}{1+\sqrt{\Delta}}$. Then, a sharper analysis on equation (A.7) implies that when $\frac{m}{(1+m)^2} = \kappa - b^2$, its *lhs* is positive while the *rhs* is zero, therefore $m_c \in (0,\nu)$. Next, by rearranging equation (A.7) one obtains the following equivalent characterization pinning down m_c:

$$m^{2\lambda-1} = \mathfrak{T}_{\mathfrak{c}}(m) := \frac{(\kappa-b^2)(\lambda m+\lambda-1)(1+m) - \lambda m + \frac{m^2}{1+m}}{(\kappa-b^2)(\lambda m+\lambda-m)(1+m) - \lambda m + \frac{m}{1+m}} \tag{A.19}$$

In the next lemma we propose a lower bound for m_c, i.e the unique solution to the above equation on $(0,\nu)$.

Lemma 12. *The equation $\nu m = \mathfrak{T}_{\mathfrak{c}}(m)$ has a unique solution on $(0,\nu)$, denoted by m_ν, that is smaller than m_c.*

Proof. We claim the function $\mathfrak{T}_{\mathfrak{c}}$ is decreasing in m. Differentiating in m, one can see $\mathfrak{T}_{\mathfrak{c}}{}'(m)$ is positively proportional to

$$\mathfrak{Q}_{\mathfrak{c}}(m) := (m^4+4m^3+6m^2+4m+1)(\kappa-b^2)^2 - (m^4-2m^2+1)(\kappa-b^2) - m^2, \tag{A.20}$$

that is convex in $(\kappa-b^2)$ hence achieving its maximum at the boundaries of $(\kappa-b^2)$, namely

$\{0, 1/4\}$. At $\kappa - b^2 = 0$, $\mathfrak{Q}_{\mathfrak{c}}$ is obviously negative. Also at $\kappa - b^2 = 1/4$, $\mathfrak{Q}_{\mathfrak{c}}$ equals to

$$\mathfrak{Q}_{\mathfrak{c}}(m) = \frac{-1}{16}(3m^2 + 2m + 3)(1 - m)^2 \leq 0. \tag{A.21}$$

Therefore, $\mathfrak{T}_{\mathfrak{c}}{}'(m) \leq 0$. Denote the solution to $\nu m = \mathfrak{T}_{\mathfrak{c}}(m)$ by m_ν that exists uniquely because $\mathfrak{T}_{\mathfrak{c}}$ is shown to be decreasing and $\mathfrak{T}_{\mathfrak{c}}(0) > 0$. Furthermore, since $m_c < 1$ then on $(0, m_c)$, $\nu m > m^{2\lambda - 1}$, therefore $m_\nu < m_c$.‖

The next step in the proof of the proposition is to show $m_\nu > m_n$, where m_n is the odd ratio associated to α_n.

Lemma 13. *The solution to $\nu m = \mathfrak{T}_{\mathfrak{c}}(m)$ denoted by m_ν is bigger than m_n.*

Proof. Let us represent (A.13) in terms of the odd-ratio m:

$$\mathfrak{Ro}(m) := m^2 - 2bm(1+m) + (1+m)^2(\kappa - b^2)(\lambda - 1) - \lambda m(1+m)\left(1 - (1+m)(\kappa - b^2)\right) = 0. \tag{A.22}$$

From the analysis done in proposition 5 there exists a *unique* solution to the above equation in the interval $\left[0, \frac{1/2+b}{1/2-b}\right]$ denoted by m_n. Since $\mathfrak{Ro}(0) \geq 0$, then $\mathfrak{Ro}(m) < 0$ iff $m > m_n$. By substituting $(1 + m_\nu)(\lambda - 1)(\kappa - b^2)$ from $\nu m_\nu = \mathfrak{T}_{\mathfrak{c}}(m_\nu)$ into $\mathfrak{Ro}$ we see at $m = m_\nu$:

$$\begin{aligned}\mathfrak{Ro}(m_\nu) = -(1 + m_\nu)\times \\ \left[2bm_\nu - \nu m_\nu \left((\kappa - b^2)(\lambda m_\nu + \lambda - m_\nu)(1 + m_\nu) - \lambda m_\nu + \frac{m_\nu}{1 + m_\nu}\right)\right]\end{aligned} \tag{A.23}$$

Let $\alpha_\nu = \frac{m_\nu}{1+\mu}$ and revert the analysis back to the probability domain. Then, $\mathfrak{Ro}(m_\nu) < 0$ iff

$$\begin{aligned} 2b > \nu \left[\alpha_\nu - \frac{\lambda \alpha_\nu}{1 - \alpha_\nu} + \frac{\kappa - b^2}{1 - \alpha_\nu}\left(\frac{(\lambda - 1)\alpha_\nu}{1 - \alpha_\nu} + \lambda\right)\right] \\ \Leftrightarrow 2b(1 - \alpha_\nu) + \nu \alpha_\nu(\alpha_\nu + \lambda - 1) > \frac{(\kappa - b^2)(\lambda - \alpha_\nu)\nu}{1 - \alpha_\nu}. \end{aligned} \tag{A.24}$$

Recall the definition of Δ, and define the following functions that respectively correspond to the *lhs* and *rhs* of the above inequality:

$$\begin{aligned} \mathfrak{A}_{\mathfrak{l}}(\alpha) &:= 2b(1 - \alpha) + \nu\alpha(\alpha + \lambda - 1) \\ \mathfrak{A}_{\mathfrak{r}}(\alpha) &:= \frac{(1 - \Delta)(\lambda - \alpha)\nu}{4(1 - \alpha)} \end{aligned} \tag{A.25}$$

Note that $\mathfrak{A}_{\mathfrak{l}}$ is convex in α thus is always greater than the tangent line at $\alpha = 0$, so

$$\mathfrak{A}_{\mathfrak{l}}(\alpha) \geq 2b + (\nu(\lambda - 1) - 2b)\,\alpha. \tag{A.26}$$

In addition, $\mathfrak{A}_{\mathfrak{r}}$ is convex in α, therefore on $\alpha \leq \frac{1-\sqrt{\Delta}}{2}$ (equivalently $m \leq \nu$) it is upper-bounded by the line connecting $(0, \mathfrak{A}_{\mathfrak{r}}(0))$ to $\left(\frac{1-\sqrt{\Delta}}{2}, \mathfrak{A}_{\mathfrak{r}}(\frac{1-\sqrt{\Delta}}{2})\right)$. We refer to this line by ℓ. As a result, a sufficient condition for (A.24) is that for all $\alpha \leq \frac{1-\sqrt{\Delta}}{2}$, the linear lower approximation for $\mathfrak{A}_{\mathfrak{l}}$ in (A.26) dominates the upper-envelope line ℓ. This is the case iff the domination occurs at the two ends $\alpha = 0$ and $\alpha = \frac{1-\sqrt{\Delta}}{2}$. The former is equivalent to

$$2b \geq \mathfrak{A}_{\mathfrak{r}}(0) = \frac{(1-\Delta)\lambda\nu}{4} = \frac{\lambda(1-\sqrt{\Delta})^2}{4}, \tag{A.27}$$

and the latter is equivalent to

$$\begin{aligned}
2b\left(1 - \frac{1-\sqrt{\Delta}}{2}\right) &\geq \mathfrak{A}_{\mathfrak{r}}\left(\frac{1-\sqrt{\Delta}}{2}\right) - \frac{\nu(\lambda-1)(1-\sqrt{\Delta})}{2} \\
\Leftrightarrow b(1+\sqrt{\Delta}) &\geq \frac{(1-\Delta)(1-\sqrt{\Delta})(2\lambda - 1 + \sqrt{\Delta})}{4(1+\sqrt{\Delta})^2} - \frac{(1-\sqrt{\Delta})^2(\lambda-1)}{2(1+\sqrt{\Delta})} \\
\Leftrightarrow b &\geq \frac{(1-\Delta)^2}{4(1+\sqrt{\Delta})^3}.
\end{aligned} \tag{A.28}$$

Putting together (A.27) and (A.28) imply that for

$$b \geq \max\left\{\frac{\lambda(1-\sqrt{\Delta})^2}{8}, \frac{(1-\Delta)^2}{4(1+\sqrt{\Delta})^3}\right\}, \tag{A.29}$$

$\mathfrak{Ro}(m_\nu) < 0$, and hence $m_\nu > m_n$. Lastly, note that (A.29) leads to a well-defined function $\underline{b}(\kappa, \lambda)$, because as $b \to \sqrt{\kappa}$ the *lhs* increases while the *rhs* approaches to zero (because $\Delta \to 1$).‖

Proof of part (i): Using the previous two lemmas we can now conclude that $m_c > m_\nu > m_n$ thereby $\alpha_c > \alpha_n$.

Proof of part (ii): Note that on $(\alpha_c \vee \gamma_{\mathcal{R}}^{(1)}, \gamma_{\mathcal{R}}^{(2)} \wedge \beta_n)$ the principal's full communication payoff is $-\alpha_c(1-\alpha_c)$ whereas her no-communication payoff is always larger than $-\alpha_n(1-\pi)$. Since $\alpha_c \geq \alpha_n$ and on this interval $\pi \geq \alpha_n$ then

$$-\alpha_n(1-\pi) \geq -\alpha_n(1-\alpha_n) \geq -\alpha_c(1-\alpha_c), \tag{A.30}$$

and the proof follows. □

A.4 Forward equation for the hitting probability

In section 5 we defined the hitting probability as $q(\pi) = \mathsf{P}\left(\tau_1 < \tau_\alpha \middle| \, \theta = 1\right)$ and offered the forward Kolmogorov ODE for that. Here, we present a short and intuitive verification of that result. The law of iterated expectations implies:

$$\begin{aligned} q(\pi) &= \mathsf{E}\left[q(\pi + \mathrm{d}\pi) \middle| \, \theta = 1\right] \\ &= \mathsf{E}\left[q(\pi) + q'(\pi)\mathrm{d}\pi + \frac{1}{2}q''(\pi)\mathrm{d}\langle\pi,\pi\rangle \middle| \, \theta = 1\right] \end{aligned} \tag{A.31}$$

Following the prescription for the diffusion process $\{\pi_t\}$ in (5.3), we apply Ito's lemma to the above expression and obtain the sought ODE: $q'(\pi) + \pi q''(\pi)/2 = 0$.

This ODE has the general solution: $q(\pi) = \frac{q_0}{\pi} + q_1$, for constants $\{q_0, q_1\}$. The boundary conditions $q(\alpha_n) = 0$ and $q(1) = 1$ can be used to identify the unknown coefficients, yielding:

$$q(\pi) = \mathsf{P}_\pi(\tau_1 < \tau_\alpha | \theta = 1) = \frac{\alpha_n(1-\pi)}{\pi(1-\alpha_n)} \tag{A.32}$$

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Bolton, P. and Harris, C. (1999). Strategic experimentation. *Econometrica*, 67(2):349–374.

Carpenter, D. and Ting, M. M. (2007). Regulatory errors with endogenous agendas. *American Journal of Political Science*, 51(4):835–852.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451.

Grenadier, S. R., Malenko, A., and Malenko, N. (2016). Timing decisions in organizations: Communication and authority in a dynamic environment. *American Economic Review*, 106(9):2552–81.

Guo, Y. (2016). Dynamic delegation of experimentation. *American Economic Review*, 106(8):1969–2008.

Henry, E. and Ottaviani, M. (2019). Research and the approval process: the organization of persuasion. *American Economic Review*, 109(3):911–55.

Keller, G., Rady, S., and Cripps, M. (2005). Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68.

Klein, N. (2016). The importance of being honest. *Theoretical Economics*, 11(3):773–811.

Lipster, R. and Albert, S. (2001). *Statistics of Random Processes*. Springer.

McClellan, A. (2017). Experimentation and approval mechanisms. *Unpublished working paper*.

Melumad, N. D. and Shibano, T. (1991). Communication in settings with no transfers. *The RAND Journal of Economics*, pages 173–198.

Orlov, D., Skrzypacz, A., and Zryumov, P. (2019). Persuading the principal to wait. *Journal of Political Economy*.

Peskir, G. and Shiryaev, A. (2006). *Optimal stopping and free-boundary problems*. Springer.