Dynamic Delegation of Experimentation

By Yingni Guo[*]

I study a dynamic relationship in which a principal delegates experimentation to an agent. Experimentation is modeled as a one-armed bandit that yields successes following a Poisson process. Its unknown intensity is high or low. The agent has private information, his type being his prior belief that the intensity is high. The agent values successes more than the principal, so prefers more experimentation. I reduce the analysis to a finite-dimensional problem. The optimal mechanism is a cutoff rule in the belief space. The cutoff gives pessimistic types total freedom but curtails optimistic types' behavior. Surprisingly, this delegation rule is time-consistent.
JEL: D82, D83, D86.
Keywords: principal-agent, delegation, experimentation, one-armed bandit, incomplete information, no monetary transfers.

Innovation carries great uncertainty. Firms frequently start R&D projects with little knowledge of eventual success. As experimentation goes on but no success occurs, firms grow pessimistic and taper resource input or even discontinue the project altogether.

This paper studies the optimal mechanism by which a principal (she) delegates experimentation to an agent (he), as is the case of a firm delegating an R&D project to its employee. The literature on experimentation in a principal-agent setting focuses on transferable utilities.[1] Instead, I focus on delegation (Holmström (1977, 1984)) for three reasons. First, from a practical point of view, it is obvious that an overwhelming number of economic activities (conventional and innovative) are organized by delegation: managers delegate tasks to subordinates by authority, rather than transfer-based trading contracts. Second, it is often cheaper to restrict the agent's actions than to devise a possibly complex compensation scheme. This is consistent with the transaction-cost economics which discusses the relative efficiency of authority-based organization ("hierarchies") and contract-based organization ("market") (Coase (1937); Williamson (1975)). Third, there are cases in which transfers are prohibited outright to prevent corruption, such as constituents delegating reforms to politicians. The current literature on delegation focuses on

[1]See for instance Bergemann and Hege (1998, 2005), Halac, Kartik and Liu (2013), Hörner and Samuelson (2013).

static problems which preclude learning. In this paper, I consider the problem of dynamic delegation. As new information arrives over time, the flexibility granted to the agent can be adjusted accordingly.

The R&D project consumes the principal's resources and the agent's time. Both wish to discontinue it if they become pessimistic enough. However, the agent's relative return from the project's successes frequently exceeds the principal's (high cost of principal's resources; principal's moderate benefit from one project out of her many responsibilities; agent's career advancement as an extra benefit); hence the agent prefers to keep the project alive for a longer time.

Promising projects warrant longer experimentation. Building on his expertise, the agent often has private knowledge on the prospect of the project at the outset. If the principal wishes to take advantage of his information, she has to give the agent some flexibility over resource allocation. But misaligned preferences curtail the flexibility that the principal is willing to grant. Therefore the principal faces a trade-off between using the agent's information and containing his bias.

The purpose of this paper is to solve for the optimal delegation rule. It addresses the following questions: In the absence of transfers, what instruments does the principal use to extract the agent's private information? Is there delay in information acquisition? How much of the resource allocation decision should be delegated to the agent? Will some projects be over-experimented and others under-experimented? Is the optimal delegation rule time-consistent?

To answer these questions, I model experimentation as a continuous-time bandit problem. See for instance Presman (1990), Keller, Rady and Cripps (2005), and Keller and Rady (2010). There is one unit of a perfectly divisible resource per unit of time and the agent continually splits the resource between a safe task and a risky one. In any given time interval, the safe task generates a known flow payoff proportional to the resource allocated to it.[2] The risky task's payoff depends on an unknown binary state. In the benchmark setting, if the state is good the risky task yields successes at random times. The arrival rate is proportional to the resource allocated to it. If the state is bad, the risky task yields no successes. I assume that the agent values the safe task's flow payoffs and the risky task's successes differently than the principal. Both prefer to allocate the resource to the risky task in the good state and to the safe task in the bad one. However, the agent values successes relatively more than the principal, so he prefers to experiment longer if faced with prolonged absence of success.[3] At the outset, the agent has private information: his type is his prior belief that the state is good. After experimentation begins, the agent's actions and the arrivals of successes are publicly observed.

The principal delegates the decision on how the agent should allocate the resource over time. This decision is made at the outset. Since the agent has private information before experimentation, the principal offers a set of policies from which the agent chooses his preferred one. A policy specifies how the agent should allocate

---

[2]The flow payoff generated by the safe task can be regarded as the opportunity cost saved or the payoff from conducting conventional tasks.

[3]This assumption is later relaxed and the case in which the bias goes in the other direction is also studied.

the resource in all future contingencies.

Note that the space of all policies is very large. Possible policies include: allocating all resource to the risky task until a fixed time and then switching to the safe task only if no success has realized; gradually reducing the resource input to the risky task if no success occurs and allocating all resource to it after the first success; allocating all resource to the risky task until the first success and then allocating a fixed fraction to the risky task; always allocating a fixed fraction of the unit resource to the risky task; etc.

A key observation is that any policy, in terms of payoffs, can be summarized by a pair of numbers, corresponding to the *total expected discounted resource* allocated to the risky task conditional on the state being good and the *total expected discounted resource* allocated to the risky task conditional on the state being bad. As far as payoffs are concerned, this constitutes a simple, finite-dimensional summary statistic for any given policy. The range of these summary statistics as we vary policies is what I call the feasible set—a subset of the plane. Determining the feasible set is a nontrivial problem in general, but it involves no incentive constraints and therefore reduces to a standard optimization problem which I solve. This reduces the delegation problem to a static one. Given that the resulting problem is static, I use Lagrangian optimization methods (similar to those used by Amador, Werning and Angeletos (2006)) to solve it and determine the optimal delegation rule.

Under a mild regularity condition, the optimal delegation rule takes a very simple form. It is a *cutoff rule* with a properly calibrated prior belief that the state is good. This belief is then updated *as if* the agent had no private information. In other words, this belief drifts down when no success is observed and jumps to one upon the first success. It is updated in the way the principal would if she were carrying out the experiment herself (starting at the calibrated prior belief). The agent freely decides whether to experiment or not as long as the updated belief remains above the cutoff. However, if this belief ever reaches the cutoff, the agent is asked to stop experimenting. This rule turns out not to bind for types with low enough priors, who voluntarily stop experimenting conditional on no success, but does constrain those with high enough priors, who are required to stop when the cutoff is reached.

Given this updating rule, the belief jumps to one upon the first success. Hence, in the benchmark setting the cutoff rule can be implemented by imposing a deadline for experimentation, under which the agent allocates all resource to the risky task after the first success, but is not allowed to experiment past the deadline. Those types with low enough priors stop experimenting before the deadline conditional on no success whereas a positive measure of types with high enough priors stop at the deadline. In equilibrium, there is no delay in information acquisition as the risky task is operated exclusively until either the first success reveals that the state is good or the agent stops.

Among the set of high enough types who are forced to stop when the cutoff (or the deadline) is reached, the highest subset under-experiment even from the principal's point of view. Every other type over-experiments. A prediction of my model is the most promising projects are always terminated too early while less promising ones

are stopped too late due to agency.

An important property of the cutoff rule is time consistency: After any history the principal would not revise the cutoff rule even if she were given a chance to do so. In particular, after the agent experiments for some time, yet no success has realized, the principal still finds it optimal to keep the cutoff (or the deadline) at the level set at the beginning. Hence, surprisingly, implementing the cutoff rule requires no commitment on the principal's side.

I then show that both the optimality of the cutoff rule and its time-consistency generalize to situations in which the risky task generates successes in the bad state as well. When successes are inconclusive, the belief is updated differently than in the benchmark setting. It jumps up upon successes and then drifts down. Consequently, the cutoff rule cannot be implemented by imposing a deadline. Instead, it can be implemented as a sliding deadline. The principal initially extends some time to the agent to operate the risky task. Then, whenever a success realizes, more time is granted. The agent is free to switch to the safe task before he uses up the time granted by the principal. After a long stretch of time without success, the agent is required to switch to the safe task.[4]

I further extend the analysis to the case in which the agent gains less from the experimentation than the principal and therefore tends to under-experiment. This happens when an innovative task yields positive externalities, or when it is important to the firm but does not widen the agent's influence. When the agent's bias is small enough, the optimum can be implemented by imposing a lockup period which is extended upon successes. Instead of placing a cap on the length of experimentation as in the previous case, the principal enacts a floor. The agent has no flexibility but to experiment before the lockup period ends, yet has full flexibility afterwards. Time-consistency is no longer preserved, as whenever the agent stops experimenting voluntarily, he reveals that the principal's optimal experimentation length has yet to be reached. The principal is tempted to order the agent to experiment further. Therefore to implement the sliding lockup period, commitment from the principal is required.

My results have two important implications for the practical design of delegation rules (I assume a larger agent's return in this illustration). First, a (sliding) deadline should be in place as a safeguard against abuse of the principal's resources. The continuation of the project is permitted only upon demonstrated successes. Second, the agent should have the flexibility over resource allocation before the (sliding) deadline is reached. In particular, the agent should be free to terminate the project whenever he finds appropriate. Besides in-house innovation, these results apply to various resource allocation problems with experimentation, such as companies budgeting marketing resources for product introduction and funding agencies awarding grants to scholarly research.[5]

---

[4]In the Web Appendix, I show that the optimality of the cutoff rule and its time-consistent property generalizes to Lévy bandits (Cohen and Solan (2013)).

[5]Suppose that experimentation has a flow cost. The agent is cash-constrained and his action is contractible. Delegating experimentation means funding his research project.

My paper contributes to the literature on delegation. This literature addresses the incentive problems in organizations which arise due to hidden information and misaligned preferences. Holmström (1977, 1984) provides conditions for the existence of an optimal solution to the delegation problem. He also characterizes optimal delegation sets in a series of examples, under the restriction, for the most part, that only interval delegation sets are allowed. Alonso and Matouschek (2008) and Amador and Bagwell (2013) characterize the optimal delegation set in general environments under some conditions and provide conditions under which simple interval delegation is optimal. None of these papers consider dynamic delegation. What distinguishes my model from static delegation problems is that additional information arises over time. The principal ought to use it both to reduce the agent's informational rents and to adjust his behavior. My paper complements the current literature and facilitates the understanding of how to optimally delegate experimentation.

Second, my paper is related to the literature on experimentation in a principal-agent setting.[6] Since most papers address different issues than I do, here I only mention the most related ones. Gomes, Gottlieb and Maestri (2013) study a multiple-period model in which the agent has private information about both the project quality and his cost of effort. The agent's actions are observable. Unlike in my setting, the agent has no benefit from the project and outcome-contingent transfers are allowed. They identify necessary and sufficient conditions under which the principal only pays rents for the agent's information about his cost, but not for the agent's information about the project quality. Garfagnini (2011) studies a dynamic delegation model without hidden information at the beginning. The principal cannot commit to future actions and transfers are infeasible. Agency conflicts arise because the agent prefers to work on the project regardless of the state. He delays information acquisition to prevent the principal from growing pessimistic. In my model, there is pre-contractual hidden information; transfers are infeasible and the principal is able to commit to long-term contract terms; the agent has direct benefit from experimentation and shares the same preferences as the principal conditional on the state. Agency conflicts arise as the agent is inclined to exaggerate the prospects for success and prolong the experimentation.

The paper is organized as follows. The model is presented in Section I. Section II considers a single player's decision problem. In Section III, I illustrate how to reduce the delegation problem to a static one. The main results are presented in Section IV. I extend the analysis to more general stochastic processes in Section V and discuss other extensions of the model. Section VI concludes.

---

[6]Bergemann and Hege (1998, 2005) study the financing of a new venture in which the principal funds the experiment and the agent makes contract offers. Dynamic agency problem arises as the agent can invest or divert the funds. Hörner and Samuelson (2013) consider a similar model in which the agent's effort requires funding and is unobservable. The principal makes short-term contract offers specifying profit-sharing arrangement. Halac, Kartik and Liu (2013) study long-term contract for experimentation with adverse selection about the agent's ability and moral hazard about his effort choice.

I. The model

SMALL CAPS

PLAYERS, TASKS AND STATES

Time $t \in [0, \infty)$ is continuous. There are two risk-neutral players $i \in \{\alpha, \rho\}$, an agent (he) and a principal (she), and two tasks, a safe task S and a risky one R. The principal is endowed with one unit of perfectly divisible resource per unit of time. She delegates resource allocation to the agent, who continually splits the resource between the two tasks. The safe task yields a known deterministic flow payoff that is proportional to the fraction of the resource allocated to it. The risky task's payoff depends on an unknown binary state, $\omega \in \{0, 1\}$.

In particular, if the fraction $\pi_t \in [0, 1]$ of the resource is allocated to R over an interval $[t, t + dt)$, and consequently $1 - \pi_t$ to S, player i receives $(1 - \pi_t)s_i dt$ from S, where $s_i > 0$ for both players. The risky task generates a success at some point in the interval with probability $\pi_t \lambda^1 dt$ if $\omega = 1$ and $\pi_t \lambda^0 dt$ if $\omega = 0$. Each success is worth h_i to player i. Therefore, the overall expected payoff increment to player i conditional on ω is $[(1 - \pi_t)s_i + \pi_t \lambda^\omega h_i]dt$. All this data is common knowledge.[7]

In the benchmark setting, I assume that $\lambda^1 > \lambda^0 = 0$. Hence, R yields no success in state 0. In Subsection V.A, I extend the analysis to the setting in which $\lambda^1 > \lambda^0 > 0$.

CONFLICTS OF INTERESTS

I allow different payoffs to players, *i.e.*, I do not require that $s_\alpha = s_\rho$ or $h_\alpha = h_\rho$. The restriction imposed on payoff parameters is the following:

ASSUMPTION 1:   *Parameters are such that $\lambda^1 h_i > s_i > \lambda^0 h_i$ for $i \in \{\alpha, \rho\}$, and*

$$\frac{\lambda^1 h_\alpha - s_\alpha}{s_\alpha - \lambda^0 h_\alpha} > \frac{\lambda^1 h_\rho - s_\rho}{s_\rho - \lambda^0 h_\rho}.$$

Assumption 1 has two implications. First, there is agreement on how to allocate the resource if the state is known. Both players prefer to allocate the resource to R in state 1 and the resource to S in state 0. Second, the agent values successes over flow payoffs relatively more than the principal does. Let

$$\eta_i = \frac{\lambda^1 h_i - s_i}{s_i - \lambda^0 h_i}$$

denote player i's net gain from R's successes over S's flow payoffs. The ratio η_α / η_ρ, being strictly greater than one, measures how misaligned players' interests are and is referred to as the agent's bias. (The case in which the bias goes in the other direction is discussed in Subsection V.B.)

---

[7]It is not necessary that S generates deterministic flow payoffs. What matters to players is that the expected payoff rates of S are known and equal s_i, and that S's flow payoffs are uncorrelated with the state.

Players do not observe the state. At time 0, the agent has private information about the probability that the state is 1. For ease of exposition, I express the agent's prior belief that the state is 1 in terms of the implied odds ratio of state 1 to state 0, denoted θ and referred to as the agent's type. The agent's type is drawn from a compact interval $\Theta \equiv [\underline{\theta}, \bar{\theta}] \subset \mathbf{R}_+$ according to some continuous density function f. Let F denote the cumulative distribution function.

By the definition of the odds ratio, the agent of type θ assigns probability $p(\theta) = \theta/(1+\theta)$ to the event that the state is 1 at time 0. The principal knows only the type distribution. Hence, her prior belief that the state is 1 is given by

$$\mathbf{E}[p(\theta)] = \int_\Theta \frac{\theta}{1+\theta} \mathrm{d}F(\theta).$$

Actions and successes are publicly observable. The only information asymmetry comes from the agent's private information about the state at time 0. Hence, a resource allocation policy, which I introduce next, conditions on both the agent's past actions and arrivals of successes.

POLICIES AND POSTERIOR BELIEFS

A (pure) resource allocation policy is a non-anticipative stochastic process $\pi = \{\pi_t\}_{t \geq 0}$. Here, $\pi_t \in [0, 1]$ is interpreted as the fraction of the unit resource allocated to R at time t, which may depend only on the history of events up to t. A policy π can be described as follows. At time 0, a choice is made of a deterministic function $\pi(t \mid 0)$, measurable with respect to t, $0 \leq t < \infty$, which takes values in $[0, 1]$ and corresponds to the fraction of the resource allocated to R up to the moment of the first success. If at the random time τ_1 a success occurs, then depending on the value of τ_1, a new function $\pi(t \mid \tau_1, 1)$ is chosen, etc. The space of all policies, including randomized ones, is denoted Π. (See Footnote 9.)

Let N_t denote the number of successes observed up to time t. Both players discount payoffs at rate $r > 0$. Player i's payoff given an arbitrary policy $\pi \in \Pi$ and an arbitrary prior belief $p \in [0, 1]$ consists of the expected discounted payoffs from R's successes and the expected discounted flow payoffs from S

$$U_i(\pi, p) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt} \left[h_i \mathrm{d}N_t + (1 - \pi_t)\, s_i \mathrm{d}t\right] \,\Big|\, \pi, p\right].$$

Here, the expectation is taken over the state ω and the stochastic processes π and N_t. By the Law of Iterated Expectations, I can rewrite player i's payoff as the discounted sum of the expected payoff increments

$$U_i(\pi, p) = \mathbf{E}\left[\int_0^\infty re^{-rt} \left[(1 - \pi_t)s_i + \pi_t \lambda^\omega h_i\right] \mathrm{d}t \,\Big|\, \pi, p\right].$$

7

Given prior p, policy π and trajectory N_s on the time interval $0 \leq s \leq t$, I consider the posterior probability p_t that the state is 1. The function p_t may be assumed to be right-continuous with left-hand limits. Because R yields no success in state 0, before the first success of the process N_t, the process p_t satisfies a differential equation

$$\dot{p}_t = -\pi_t \lambda^1 p_t (1 - p_t). \tag{1}$$

At the first success, p_t jumps to one.[8]

DELEGATION

I consider the situation in which transfers are not allowed and the principal is able to commit to dynamic policies. At time 0, the principal chooses a set of policies from which the agent chooses his preferred one. Since there is hidden information at time 0, by the revelation principle, the principal's problem is reduced to solving for a map $\boldsymbol{\pi} : \Theta \to \Pi$ to maximize her expected payoff subject to the agent' incentive compatibility constraint (IC constraint, hereafter). Formally, I solve

$$\sup \quad \int_\Theta U_\rho(\boldsymbol{\pi}(\theta), p(\theta)) \mathrm{d}F(\theta),$$
$$\text{subject to} \quad U_\alpha(\boldsymbol{\pi}(\theta), p(\theta)) \geq U_\alpha(\boldsymbol{\pi}(\theta'), p(\theta)) \quad \forall \theta, \theta' \in \Theta,$$

over measurable $\boldsymbol{\pi} : \Theta \to \Pi$.[9]

II. The single-player benchmark

In this section, I present player i's preferred policy as a single player. This is a standard problem. The policy preferred by player i is Markov with respect to the posterior belief p_t. It is characterized by a cutoff belief p_i^* such that $\pi_t = 1$ if $p_t \geq p_i^*$ and $\pi_t = 0$ otherwise. By standard results (see Keller, Rady and Cripps (2005) Proposition 3.1, for instance), the cutoff belief is

$$p_i^* = \frac{s_i}{\lambda^1 h_i + \frac{(\lambda^1 h_i - s_i)\lambda^1}{r}} = \frac{r}{r + (\lambda^1 + r)\eta_i}. \tag{2}$$

---

[8]In general, subscripts indicate either time or player. Superscripts refer to state. Parentheses contain type or policy.

[9]Here, I define randomized policies and stochastic mechanisms following Aumann (1964). Let $\mathcal{B}_{[0,1]}$ (resp. $\mathcal{B}^k$) denote the σ-algebra of Borel sets of $[0, 1]$ (resp. $\mathbf{R}_+^k$) and λ the Lebesgue measure on $[0, 1]$, where k is a positive integer. I denote the set of measurable functions from $(\mathbf{R}_+^k, \mathcal{B}^k)$ to $([0, 1], \mathcal{B}_{[0,1]})$ by F^k and endow this set with the σ-algebra generated by sets of the form $\{f : f(s) \in A\}$ with $s \in \mathbf{R}_+^k$ and $A \in \mathcal{B}_{[0,1]}$. The σ-algebra is denoted χ^k. Let Π^* denote the space of pure policies. I impose on Π^* the product σ-algebra generated by $(F^k, \chi^k), \forall k \in N_+$. Following Aumann (1964), I define randomized policies as measurable functions $\hat{\pi} : [0, 1] \to \Pi^*$. According to $\hat{\pi}$, a value $\epsilon \in [0, 1]$ is drawn uniformly from $[0, 1]$ and then the pure policy $\hat{\pi}(\epsilon)$ is implemented. Analogously, I define stochastic mechanisms as measurable functions $\hat{\boldsymbol{\pi}} : [0, 1] \times \Theta \to \Pi^*$. A value $\epsilon \in [0, 1]$ is drawn uniformly from $[0, 1]$, along with the agent's report θ, determines which element of Π is chosen. For ease of exposition, my descriptions assume pure policies and deterministic mechanisms. My results do not.

Note that the cutoff belief p_i^* decreases in η_i. Therefore, the agent's cutoff belief p_α^* is lower than the principal's p_ρ^*, as he values R's successes over S's flow payoffs more than the principal does.

Given the law of motion of belief (1) and the cutoff belief (2), player i's preferred policy given prior $p(\theta)$ can be identified with a fixed stopping time $\tau_i(\theta)$: if the first success occurs before the stopping time, use R forever after the first success; otherwise, use R until the stopping time and then switch to S. Player i's preferred stopping time for a given θ is stated as follows:

CLAIM 1: *Player i's stopping time given odds ratio $\theta \in \Theta$ is*

$$\tau_i(\theta) = \begin{cases} \frac{1}{\lambda^1} \log \frac{(r+\lambda^1)\theta\eta_i}{r} & if \ \frac{(r+\lambda^1)\theta\eta_i}{r} \geq 1, \\ 0 & if \ \frac{(r+\lambda^1)\theta\eta_i}{r} < 1. \end{cases}$$

Figure 1 illustrates the two players' cutoff beliefs and their preferred stopping times associated with two possible odds ratios θ', θ'' (with $\theta' < \theta''$). The prior beliefs are thus $p(\theta'), p(\theta'')$ (with $p(\theta') < p(\theta'')$). The x-axis variable is time and the y-axis variable is the posterior belief. On the y-axis are the two players' cutoff beliefs p_ρ^* and p_α^*. The solid and dashed lines depict how posterior beliefs evolve when R is used exclusively and no success realizes.

The figure on the left-hand side shows that for a given odds ratio the agent prefers to experiment longer than the principal does because his cutoff is lower than the principal's. The figure on the right-hand side shows that for a given player i, the stopping time increases in the odds ratio, *i.e.*, $\tau_i(\theta') < \tau_i(\theta'')$. Therefore, both players prefer to experiment longer given a higher odds ratio. Figure 1 makes clear what agency problem the principal faces. The principal's stopping time $\tau_\rho(\theta)$ is an increasing function of θ. The agent prefers to stop later than the principal for a given θ and thus has incentives to misreport his type. More specifically, lower types (those types with a lower θ) have incentives to mimic high types to prolong the experimentation.
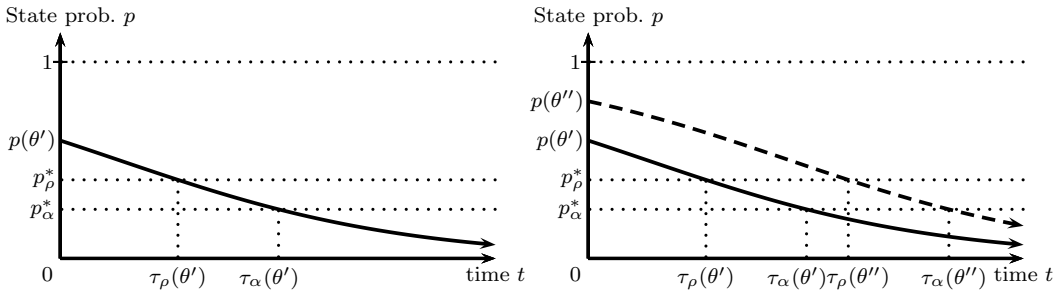

FIGURE 1. THRESHOLDS AND STOPPING TIMES

Note: Parameters are $\eta_\alpha = 3/2, \eta_\rho = 3/4, r/\lambda^1 = 1, \theta' = 3/2, \theta'' = 4$.

Given that a single player's preferred policy is always characterized by a stopping time, one might expect that the solution to the delegation problem is a set of stopping times. This is the case if there is no private information or no bias. For example, if the distribution F is degenerate, information is symmetric. The optimal delegation set is the principal's preferred stopping time given her prior. If η_α/η_ρ equals one, the two players' preferences are perfectly aligned. The principal, knowing that for any prior the agent's preferred stopping time coincides with hers, offers the set of her preferred stopping times $\{\tau_\rho(\theta) : \theta \in \Theta\}$ for the agent to choose from.

However, if the agent has private information and is also biased, it is unclear how the principal should restrict his actions. Particularly, it is unclear whether the principal would still offer a set of stopping times.[10] For this reason, I am led to consider the space of all policies.

III. A finite-dimensional characterization of the policy space

The space of all policies is large. In the first half of this section, I associate to each policy—a (possibly complicated) stochastic process—a pair of numbers, called *total expected discounted resource* pair, and show that this pair is a sufficient statistic for this policy in terms of both players' payoffs. Then, I solve for the set of feasible *total expected discounted resource* pairs, which is a subset of $\mathbf{R}^2$ and can be treated as the space of all policies.

This transformation allows me to reduce the dynamic delegation problem to a static one. In the second half of this section, I characterize players' preferences over the feasible pairs and reformulate the delegation problem.

A. A policy as a pair of numbers

For a fixed policy π, I define $\mathbf{w}^1(\pi)$ and $\mathbf{w}^0(\pi)$ as follows:

$$(3) \quad \mathbf{w}^1(\pi) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt}\pi_t \mathrm{d}t \;\middle|\; \pi, 1\right] \text{ and } \mathbf{w}^0(\pi) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt}\pi_t \mathrm{d}t \;\middle|\; \pi, 0\right].$$

The term $\mathbf{w}^1(\pi)$ is the expected discounted sum of the resource allocated to R under π in state 1. I refer to $\mathbf{w}^1(\pi)$ as the *total expected discounted resource* (*expected resource*, hereafter) allocated to R under π in state 1.[11] Similarly, the term $\mathbf{w}^0(\pi)$ is the *expected resource* allocated to R under π in state 0. Both $\mathbf{w}^1(\pi)$ and $\mathbf{w}^0(\pi)$ are in $[0,1]$ because π takes values in $[0,1]$. Therefore, $(\mathbf{w}^1, \mathbf{w}^0)$ defines a mapping from the policy space Π to $[0,1]^2$.

A policy's resource pair is important because all payoffs from implementing this policy can be written in terms of its resource pair. Conditional on the state, the

---

[10]There exist examples in which restricting attention to stopping-time policies is sub-optimal. See the Web Appendix.

[11]For a fixed policy π, the *expected resource* spent on R in state 1 is proportional to the expected discounted number of successes, *i.e.*, $\mathbf{w}^1(\pi) = \mathbf{E}\left[\int_0^\infty re^{-rt}\mathrm{d}N_t \mid \pi, 1\right]/\lambda^1$.

payoff rate of R is known, so the payoff from implementing this policy only depends on how the resource is allocated between R and S. For instance, conditional on state 1, $(\lambda^1 h_i, s_i)$ is the payoff rate of R and S; $(\mathbf{w}^1(\pi), 1 - \mathbf{w}^1(\pi))$ is the resource allocated to R and S in expectation. Their product is the payoff conditional on state 1.[12] Multiplying the payoffs conditional on the state by the initial state distribution gives the payoff of this policy. I summarize the analysis above in the following lemma.

LEMMA 1 (A policy as a pair of numbers):
For a given policy $\pi \in \Pi$ and a given prior $p \in [0,1]$, player i's payoff can be written as

(4) $\qquad U_i(\pi, p) = p \left(\lambda^1 h_i - s_i\right) \mathbf{w}^1(\pi) + (1 - p) \left(\lambda^0 h_i - s_i\right) \mathbf{w}^0(\pi) + s_i.$

PROOF:
Player i's payoff given policy $\pi \in \Pi$ and prior $p \in [0,1]$ is

$$
\begin{aligned}
U_i(\pi, p) =& \mathbf{E}\left[\int_0^\infty r e^{-rt}\left[(1 - \pi_t)s_i + \pi_t \lambda^\omega h_i\right] \mathrm{d}t \;\middle|\; \pi, p\right] \\
=& p\mathbf{E}\left[\int_0^\infty r e^{-rt}\left[s_i + \pi_t\left(\lambda^1 h_i - s_i\right)\right] \mathrm{d}t \;\middle|\; \pi, 1\right] \\
& + (1 - p)\mathbf{E}\left[\int_0^\infty r e^{-rt}\left[s_i + \pi_t\left(\lambda^0 h_i - s_i\right)\right] \mathrm{d}t \;\middle|\; \pi, 0\right] \\
=& p\left(\lambda^1 h_i - s_i\right)\mathbf{E}\left[\int_0^\infty r e^{-rt}\pi_t \mathrm{d}t \;\middle|\; \pi, 1\right] \\
& + (1 - p)\left(\lambda^0 h_i - s_i\right)\mathbf{E}\left[\int_0^\infty r e^{-rt}\pi_t \mathrm{d}t \;\middle|\; \pi, 0\right] + s_i \\
=& p\left(\lambda^1 h_i - s_i\right)\mathbf{w}^1(\pi) + (1 - p)\left(\lambda^0 h_i - s_i\right)\mathbf{w}^0(\pi) + s_i.
\end{aligned}
$$

∎

Lemma 1 shows that $(\mathbf{w}^1(\pi), \mathbf{w}^0(\pi))$ is a sufficient statistic for policy π for the payoffs. Instead of working with a generic policy π, it is without loss of generality to focus on $(\mathbf{w}^1(\pi), \mathbf{w}^0(\pi))$.

B. Feasible set

Let Γ denote the image of the mapping $(\mathbf{w}^1, \mathbf{w}^0) : \Pi \to [0,1]^2$. I call Γ the feasible set since it contains all possible (w^1, w^0) pairs that can be achieved by some policy π. The following lemma shows that the feasible set is the convex hull of the image of Markov policies under $(\mathbf{w}^1, \mathbf{w}^0)$.

LEMMA 2 (Feasible set):

---

[12]Recall that, conditional on the state and over any interval, S generates a flow payoff proportional to the resource allocated to it and R yields a success with probability proportional to the resource allocated to it.

The feasible set is the convex hull of $\left\{ (\mathbf{w}^1(\pi), \mathbf{w}^0(\pi)) : \pi \in \Pi^M \right\}$, where Π^M is the set of Markov policies with respect to the posterior belief of state 1.

PROOF:

The feasible set is convex given that the policy space Π is convexified. (Recall that Π includes all randomized policies. See Footnote 9.) Therefore, I only need to characterize its extreme points. A bundle $\hat{w}$ is an extreme point of Γ if and only if there exists $(p_1, p_2) \in \mathbf{R}^2, \|(p_1, p_2)\| = 1$ such that $\hat{w} \in \arg\max_{w \in \Gamma}(p_1, p_2) \cdot w$. Therefore, the feasible set can be found by taking the convex hull of the following set

$$(5) \qquad \left\{ \hat{w} \in [0,1]^2 \; \middle| \; \exists (p_1, p_2) \in \mathbf{R}^2, \|(p_1, p_2)\| = 1, \hat{w} \in \arg\max_{w \in \Gamma}(p_1, p_2) \cdot w \right\}.$$

Comparing the objective $(p_1, p_2) \cdot w$ with (4), I can rewrite $(p_1, p_2) \cdot w$ as the expected payoff of a single player i whose prior belief of state 1 is $|p_1|/(|p_1| + |p_2|)$ and whose payoff parameters are $s_i = 0, \lambda^1 h_i = \mathrm{sgn}(p_1), \lambda^0 h_i = \mathrm{sgn}(p_2)$. (Here, $\mathrm{sgn}(\cdot)$ is the sign function.) It follows that the argument of the maximum, $\arg\max_{w \in \Gamma}(p_1, p_2) \cdot w$, coincides with player i's optimal policy. This transforms the problem to a standard optimization problem. Markov policies are sufficient. ∎

Here, I calculate the image of $(\mathbf{w}^1, \mathbf{w}^0)$ for two classes of policies, which turn out to be important for characterizing Γ. The first class consists of stopping-time policies: allocate all resource to R until a fixed time; if at least one success occurs by then, allocate all resource to R forever; otherwise, switch to S forever. As mentioned before, a single player's optimal policy if he wants to maximize his payoff is always a stopping-time policy. The image of all stopping-time policies under $(\mathbf{w}^1, \mathbf{w}^0)$ is denoted Γ^{st}. It is easy to verify that

$$\Gamma^{\mathrm{st}} = \left\{ (w^1, w^0) \; \middle| \; w^0 = 1 - (1 - w^1)^{\frac{r}{r + \lambda^1}}, w^1 \in [0,1] \right\}.$$

The second class consists of slack-after-success policies: allocate all resource to R until the first success occurs; then allocate a fixed fraction to R. If the state is 0, all resource is directed to R because no success ever occurs. Therefore, $\mathbf{w}^0$ equals one for all slack-after-success policies. Meanwhile, $\mathbf{w}^1$ varies as I vary the fraction of the resource directed to R after the first success. If no resource is directed to R after the first success, $\mathbf{w}^1$ is the lowest. Note that this policy is the optimal one when a single player wants to minimize his payoff. The image of all slack-after-success policies under $(\mathbf{w}^1, \mathbf{w}^0)$ is denoted Γ^{sl}. It is easy to verify that

$$\Gamma^{\mathrm{sl}} = \left\{ (w^1, w^0) \; \middle| \; w^0 = 1, w^1 \in \left[\frac{r}{r + \lambda^1}, 1 \right] \right\}.$$

The following lemma shows that the convex hull of these two classes is the feasible set, because they correspond to a single player's optimal policy if he wants to maximize or minimize his payoff.

12

LEMMA 3 (Conclusive news—feasible set):
The feasible set is $\mathrm{co}\{\Gamma^{st} \cup \Gamma^{sl}\}$, *the convex hull of the image of stopping-time and slack-after-success policies.*

PROOF:

Based on the proof of Lemma 2, I want to show that the maximum in (5) is achieved by either a stopping-time or slack-after-success policy. If $p_1 \geq 0, p_2 \geq 0$ (resp. $p_1 \leq 0, p_2 \leq 0$), $\max_{w \in \Gamma}(p_1, p_2) \cdot w$ is achieved by the policy which directs all resources to R (resp. S). If $p_1 > 0, p_2 < 0$, $\max_{w \in \Gamma}(p_1, p_2) \cdot w$ is achieved by a lower-cutoff Markov policy under which R is used exclusively if the posterior belief is above the cutoff and S is used below. This Markov policy is effectively a stopping time policy. If $p_1 < 0, p_2 > 0$, according to Keller and Rady (forthcoming), $\max_{w \in \Gamma}(p_1, p_2) \cdot w$ is achieved by a upper-cutoff Markov policy under which R is used exclusively if the posterior belief is below the cutoff and S is used if above. This is either a policy allocating all resource to R until the first success and then switching to S, or a policy allocating all resource to S. ∎

Figure 2 depicts the image of all stopping-time policies and slack-after-success policies when r/λ^1 equals 1. The shaded area is $\mathrm{co}\{\Gamma^{st} \cup \Gamma^{sl}\}$. The (w^1, w^0) pairs on the southeast boundary correspond to stopping-time policies. Those pairs on the north boundary correspond to slack-after-success policies.
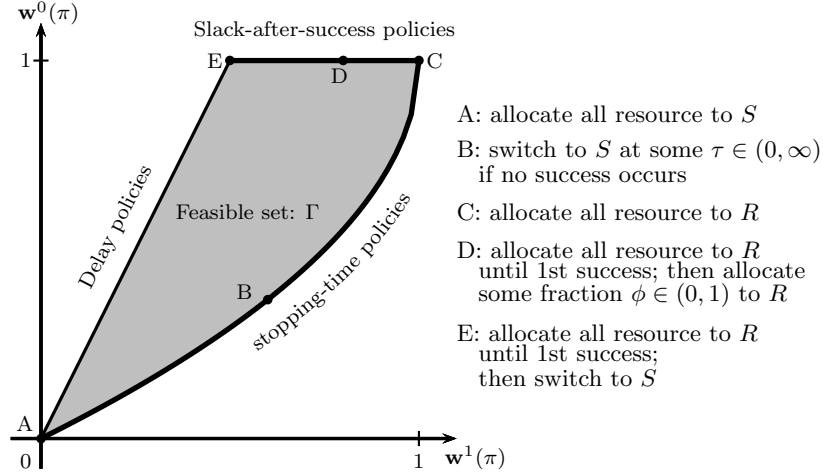

FIGURE 2. FEASIBLE SET AND EXAMPLE POLICIES ($r/\lambda^1 = 1$)

According to Lemma 3, the image of any policy π under $(\mathbf{w}^1, \mathbf{w}^0)$ is in $\mathrm{co}\{\Gamma^{st} \cup \Gamma^{sl}\}$. Also, for any $(w^1, w^0) \in \mathrm{co}\{\Gamma^{st} \cup \Gamma^{sl}\}$, I can identify a policy π (not necessarily a unique one) such that $(w^1, w^0) = (\mathbf{w}^1(\pi), \mathbf{w}^0(\pi))$. From now on, when I refer to a pair $(w^1, w^0) \in \Gamma$, I have in mind a policy such that w^1 is the *expected resource* allocated to R under this policy in state 1 and w^0 is that in state 0. A $(w^1, w^0) \in \Gamma$ pair is also called a bundle.

The feasible set Γ is bounded from above by the union of Γ^{sl} and $\{(w^1, w^0) \mid \epsilon(r/(r+\lambda^1), 1) + (1-\epsilon)(0,0), \epsilon \in [0,1]\}$. The latter set can be achieved by delaying the policy corresponding to point E (see Figure 2) for some fixed amount of time. I call this class of policies delay policies. From now on, I also refer to the union of Γ^{sl} and $\{(w^1, w^0) \mid \epsilon(r/(r+\lambda^1), 1) + (1-\epsilon)(0,0), \epsilon \in [0,1]\}$ as the northwest boundary of Γ. The fact that the northwest boundary is piecewise linear is peculiar to the benchmark setting due to its degenerate feature that R yields no success in state 0. In Subsection V.A and the Web Appendix, I characterize the feasible sets of more general stochastic processes.

The shape of the feasible set only depends on the ratio r/λ^1. Figure 3 shows that the feasible set expands as r/λ^1 decreases. Intuitively, if future payoffs are discounted to a lesser extent, a player has more time to learn about the state. As a result, he is more capable of directing resources to R in one state while avoiding wasting resources on R in the other state.
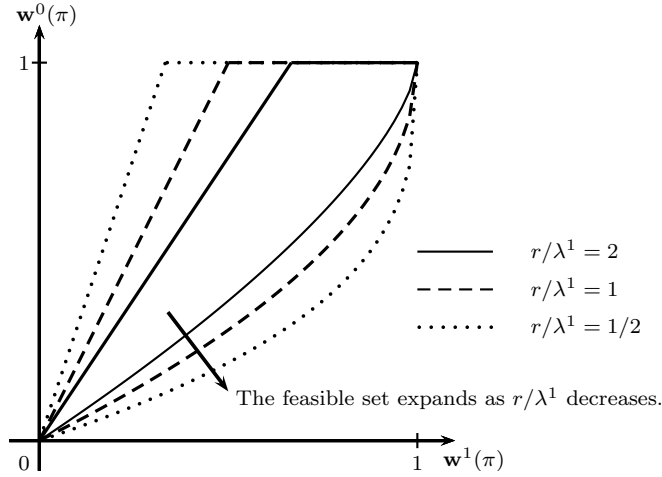

Figure 3. Feasible sets as r/λ^1 varies

For future reference, I also write the feasible set as $\Gamma = \{(w^1, w^0) \mid \beta^{\mathrm{se}}(w^1) \le w^0 \le \beta^{\mathrm{nw}}(w^1), w^1 \in [0,1]\}$ where $\beta^{\mathrm{se}}, \beta^{\mathrm{nw}}$ are functions from $[0,1]$ to $[0,1]$, characterizing the southeast and northwest boundaries of the feasible set,

$$\beta^{\mathrm{se}}\left(w^1\right) \equiv 1 - \left(1 - w^1\right)^{\frac{r}{r+\lambda^1}},$$

$$\beta^{\mathrm{nw}}\left(w^1\right) \equiv \begin{cases} \left(r + \lambda^1\right) w^1 & \text{if } w^1 \in \left[0, \frac{1}{r+\lambda^1}\right], \\ 1 & \text{if } w^1 \in \left(\frac{1}{r+\lambda^1}, \frac{1}{r}\right]. \end{cases}$$

14

C. Preferences over feasible pairs

If a player knew the state, he would allocate all resources to R in state 1 and all resources to S in state 0. However, this allocation cannot be achieved if the state is unknown. If a player wants to direct more resources to R in state 1, he has to allocate more resources to R before the arrival of the first success. Inevitably, more resources will be wasted on R if the state is actually 0.

A player's attitude toward the trade-off between spending more resources on R in state 1 and wasting fewer resources on R in state 0 depends on how likely the state is 1 and how much he gains from R's successes over S's flow payoffs. According to Lemma 1, player i's payoff given policy π and odds ratio θ is

$$
U_i(\pi, p(\theta)) = \left(\frac{\theta}{1+\theta} \eta_i \mathbf{w}^1(\pi) - \frac{1}{1+\theta} \mathbf{w}^0(\pi) \right) \left(s_i - \lambda^0 h_i \right) + s_i.
$$

Recall that $p(\theta) = \theta/(1 + \theta)$ is the prior that the state is 1 given θ. Player i's preferences over (w^1, w^0) are characterized by upward-sloping indifference curves, with slope $\theta\eta_i$. For a given player, the indifference curves are steeper, the higher the odds ratios. For a given odds ratio, the agent's indifference curves are steeper than the Principal's (see Figure 4). Player i's preferred bundle given θ, denoted $(w_i^1(\theta), w_i^0(\theta))$, is the point at which his indifference curve is tangent to the southeast boundary of Γ. It is easy to verify that $(w_i^1(\theta), w_i^0(\theta))$ corresponds to a stopping-time policy with $\tau_i(\theta)$ being the stopping time.
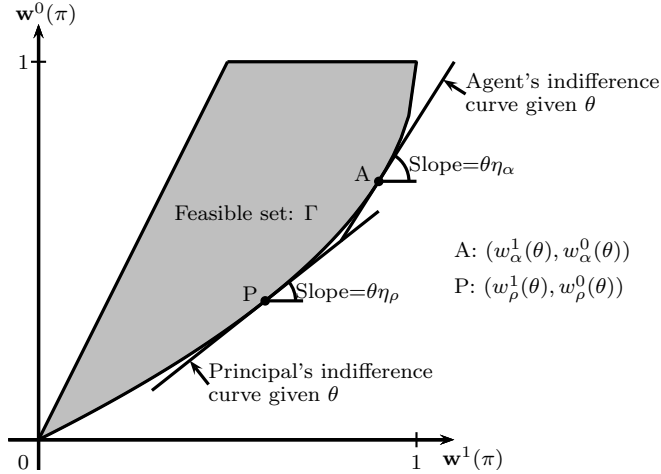

FIGURE 4. INDIFFERENCE CURVES AND PREFERRED BUNDLES

Note: Parameters are $\eta_\alpha = 3/2, \eta_\rho = 3/4, r/\lambda^1 = 1, \theta = \sqrt{10}/3$.

15

D. Delegation problem reformulated

Based on Lemma 1 and 3, I reformulate the delegation problem. Given that each policy can be represented by a bundle in Γ, the principal simply offers a direct mechanism $(w^1, w^0) : \Theta \to \Gamma$, called a contract, such that

$$(6) \qquad \max_{w^1, w^0} \int_\Theta \left(\frac{\theta}{1+\theta} \eta_\rho w^1(\theta) - \frac{1}{1+\theta} w^0(\theta) \right) dF(\theta),$$

$$(7) \qquad \text{subject to} \quad \theta\eta_\alpha w^1(\theta) - w^0(\theta) \geq \theta\eta_\alpha w^1(\theta') - w^0(\theta'), \quad \forall \theta, \theta' \in \Theta.$$

The IC constraint (7) ensures that the agent reports his type truthfully. The data relevant to this problem include: (i) two payoff parameters η_α, η_ρ; (ii) the feasible set parametrized by r and λ^1; and (iii) the type distribution F. The solution to this problem, called the optimal contract, is denoted $(w^{1*}(\theta), w^{0*}(\theta))$.[13]

IV. Main results

To solve for the optimal contract, I start with a further simplification. Given a direct mechanism $(w^1(\theta), w^0(\theta))$, let $U_\alpha(\theta)$ denote the payoff that the agent of type θ gets by maximizing over his report, *i.e.*, $U_\alpha(\theta) = \max_{\theta' \in \Theta} [\theta\eta_\alpha w^1(\theta') - w^0(\theta')]$. As the optimal mechanism is truthful, $U_\alpha(\theta)$ equals $\theta\eta_\alpha w^1(\theta) - w^0(\theta)$ and the envelope condition implies that $U'_\alpha(\theta) = \eta_\alpha w^1(\theta)$. The principal's payoff for a fixed θ is

$$\frac{\theta}{1+\theta} \eta_\rho w^1(\theta) - \frac{1}{1+\theta} w^0(\theta) = \frac{U_\alpha(\theta)}{1+\theta} + \frac{(\eta_\rho - \eta_\alpha)\theta w^1(\theta)}{1+\theta}.$$

The first term on the right-hand side corresponds to the "shared preference" between the two players because they both prefer higher w^1 value for a higher θ. The second term captures the "preference divergence" as the principal is less willing to spend resources on R in state 0 for a given increase in w^1 than the agent.

By integrating the envelope condition, one obtains the standard integral condition

$$(8) \qquad \theta\eta_\alpha w^1(\theta) - w^0(\theta) = \eta_\alpha \int_{\underline{\theta}}^\theta w^1(\tilde{\theta}) d\tilde{\theta} + \underline{\theta}\eta_\alpha w^1(\underline{\theta}) - w^0(\underline{\theta}).$$

Incentive compatibility of (w^1, w^0) also requires w^1 to be a nondecreasing function of θ: higher types (those types with a higher θ) are more willing to spend resources on R in state 0 for a given increase in w^1 than low types. Thus, condition (8) and the monotonicity of w^1 are necessary for incentive compatibility. As is standard, these two conditions are also sufficient.

The principal's problem is thus to maximize the expected payoff (6) subject to the feasibility constraint $w^0(\theta) \in [\beta^{\text{se}}(w^1(\theta)), \beta^{\text{nw}}(w^1(\theta))]$, the IC constraint (8), and monotonicity constraint $w^1(\theta') \geq w^1(\theta)$ for $\theta' > \theta$. Note that this problem is

---

[13]Since both players' payoffs are linear in (w^1, w^0), the optimal mechanism is deterministic.

convex because the expected payoff (6) is linear in $(w^1(\theta), w^0(\theta))$ and the constraint set is convex.

Substituting the IC constraint (8) into (6) and the feasibility constraint, and integrating by parts allows me to eliminate $w^0(\theta)$ from the problem except for its value at $\underline{\theta}$. I denote $w^0(\underline{\theta})$ by $\underline{w}^0$. Consequently, the principal's problem reduces to finding a function $w^1 : \Theta \to [0, 1]$ and a scalar $\underline{w}^0$ that solves

$$\text{(OBJ)} \qquad \max_{w^1, \underline{w}^0 \in \Phi} \left(\eta_\alpha \int_{\underline{\theta}}^{\overline{\theta}} w^1(\theta) G(\theta) \mathrm{d}\theta + \underline{\theta} \eta_\alpha w^1(\underline{\theta}) - \underline{w}^0 \right),$$

subject to

$$\text{(9)} \qquad \beta^{\mathrm{se}}(w^1(\theta)) \leq \theta \eta_\alpha w^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^{\theta} w^1(\tilde{\theta}) \mathrm{d}\tilde{\theta} - \underline{\theta} \eta_\alpha w^1(\underline{\theta}) + \underline{w}^0, \quad \forall \theta \in \Theta,$$

$$\text{(10)} \qquad \beta^{\mathrm{nw}}(w^1(\theta)) \geq \theta \eta_\alpha w^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^{\theta} w^1(\tilde{\theta}) \mathrm{d}\tilde{\theta} - \underline{\theta} \eta_\alpha w^1(\underline{\theta}) + \underline{w}^0, \quad \forall \theta \in \Theta,$$

where

$$\Phi \equiv \left\{ w^1, \underline{w}^0 \mid w^1 : \Theta \to [0, 1], w^1 \text{ nondecreasing}; \underline{w}^0 \in [0, 1] \right\},$$

$$G(\theta) = \left(1 - \frac{H(\theta)}{H(\overline{\theta})} \right) + \left(\frac{\eta_\rho}{\eta_\alpha} - 1 \right) \frac{\theta h(\theta)}{H(\overline{\theta})}, \text{ where } h(\theta) = \frac{f(\theta)}{1 + \theta}, H(\theta) = \int_{\underline{\theta}}^{\theta} h(\tilde{\theta}) \mathrm{d}\tilde{\theta}.$$

Here, $G(\theta)$ consists of two terms. The first term is positive as it corresponds to the "shared preference" between the two players toward higher w^1 value for a higher θ. The second term is negative as it captures the impact of the incentive problem on the principal's expected payoff due to the agent's bias toward longer experimentation.

I denote this problem by $\mathcal{P}$. The set Φ is convex and includes the monotonicity constraint. Any contract $(w^1, \underline{w}^0) \in \Phi$ uniquely determines an incentive compatible direct mechanism based on (8). A contract is admissible if $(w^1, \underline{w}^0) \in \Phi$ and the feasibility constraint, (9) and (10), is satisfied.

A. A robust result: pooling at the top

With a continuum of types, I first show that types above some threshold are offered the same (w^1, w^0) bundle. Intuitively, types at the very top prefer to experiment more than what the principal would like for any prior. Therefore, the cost of separating those types exceeds the benefit. This can be seen from the fact that the first term—the "shared preference" term—of $G(\theta)$ reduces to 0 as θ approaches $\overline{\theta}$. As a result, the principal finds it optimal to pool those types at the very top.

Let θ_p be the lowest value in Θ such that

$$\text{(11)} \qquad \int_{\hat{\theta}}^{\overline{\theta}} G(\theta) \mathrm{d}\theta \leq 0, \text{ for any } \hat{\theta} \geq \theta_p.$$

17

My next result shows that types with $\theta \geq \theta_p$ are pooled.

PROPOSITION 1 (Pooling at the top):
An optimal contract $(w^{1*}, \underline{w}^{0*})$ satisfies $w^{1*}(\theta) = w^{1*}(\theta_p)$ for $\theta \geq \theta_p$. It is optimal for (9) or (10) to hold with equality at θ_p.

PROOF:
The contribution to (OBJ) from types with $\theta > \theta_p$ is $\eta_\alpha \int_{\theta_p}^{\bar{\theta}} w^1(\theta) G(\theta) d\theta$. Substituting $w^1(\theta) = \int_{\theta_p}^{\theta} dw^1 + w^1(\theta_p)$ and integrating by parts, I obtain

$$(12) \qquad \eta_\alpha w^1(\theta_p) \int_{\theta_p}^{\bar{\theta}} G(\theta) d\theta + \eta_\alpha \int_{\theta_p}^{\bar{\theta}} \int_{\theta}^{\bar{\theta}} G(\tilde{\theta}) d\tilde{\theta} dw^1(\theta).$$

The first term only depends on $w^1(\theta_p)$. The second term depends on $dw^1(\theta)$ for all $\theta \in [\theta_p, \bar{\theta}]$. According to the definition of θ_p, the integrand of the second term, $\int_{\theta}^{\bar{\theta}} G(\tilde{\theta}) d\tilde{\theta}$, is weakly negative for all $\theta \in [\theta_p, \bar{\theta}]$. Therefore, it is optimal to set $dw^1(\theta) = 0$ for all $\theta \in [\theta_p, \bar{\theta}]$. If $\theta_p = \underline{\theta}$, all types are pooled. The principal offers her preferred uninformed bundle, which is on the southeast boundary of Γ. If $\theta_p > \underline{\theta}$, the first term of (12) is zero as well because $\int_{\theta_p}^{\bar{\theta}} G(\theta) d\theta = 0$. Adjusting $w^1(\theta_p)$ does not affect the objective function, so $w^1(\theta_p)$ can be increased until either (9) or (10) binds. ∎

The slope of the principal's indifference curves is bounded from above by $\bar{\theta}\eta_\rho$, which is the slope if she believes that the agent's type is $\bar{\theta}$. The slope of the agent's indifference curves is steeper than $\underline{\theta}\eta_\rho$ if his type is above $\bar{\theta}\eta_\rho/\eta_\alpha$. The following corollary states that types above $\bar{\theta}\eta_\rho/\eta_\alpha$ are offered the same bundle.

COROLLARY 1: *The threshold of the top pooling segment, θ_p, is below $\bar{\theta}\eta_\rho/\eta_\alpha$.*

PROOF:
Substituting $G(\theta) = \left(\int_{\theta}^{\bar{\theta}} h(\tilde{\theta}) d\tilde{\theta} + \left(\frac{\eta_\rho}{\eta_\alpha} - 1 \right) \theta h(\theta) \right) / H(\bar{\theta})$ and integrating by parts yields

$$\int_{\hat{\theta}}^{\bar{\theta}} G(\theta) d\theta = \frac{1}{H(\bar{\theta})} \int_{\hat{\theta}}^{\bar{\theta}} \left(\frac{\eta_\rho}{\eta_\alpha} \theta - \hat{\theta} \right) h(\theta) d\theta.$$

For any $\hat{\theta} \in [\bar{\theta}\eta_\rho/\eta_\alpha, \bar{\theta}]$, the integrand $(\theta\eta_\rho/\eta_\alpha - \hat{\theta})h(\theta)$ is weakly negative for any $\theta \in \Theta$. Therefore, $\theta_p \leq \bar{\theta}\eta_\rho/\eta_\alpha$. If $\theta_p > \underline{\theta}$, θ_p is strictly below $\bar{\theta}\eta_\rho/\eta_\alpha$. Otherwise, $h(\theta)$ must equal 0 for $\theta \in [\bar{\theta}\eta_\rho/\eta_\alpha, \bar{\theta}]$. A contradiction.[14] ∎

Note that for a given type distribution, the value of θ_p depends only on the ratio η_ρ/η_α but not on the magnitudes of η_ρ, η_α. If $\eta_\rho/\eta_\alpha = 1$, both parties' preferences are perfectly aligned. The function G is positive for any θ, and thus the principal optimally sets θ_p to be $\bar{\theta}$. As η_ρ/η_α decreases, the agent's bias grows. The principal

---

[14] If $\theta_p > \underline{\theta}$, I obtain that $\int_{\theta_p}^{\bar{\theta}} (\eta_\rho \theta - \eta_\alpha \theta_p) h(\theta) d\theta = 0$.

enlarges the top pooling segment by lowering θ_p. When η_ρ/η_α is sufficiently close to zero, the principal optimally sets θ_p to be $\underline{\theta}$ in which case all types are pooled.

COROLLARY 2: *For a fixed type distribution, θ_p increases in η_ρ/η_α. Moreover, $\theta_p = \bar{\theta}$ if $\eta_\rho/\eta_\alpha = 1$, and there exists $z^* \in [0,1)$ such that $\theta_p = \underline{\theta}$ if $\eta_\rho/\eta_\alpha \leq z^*$.*

PROOF:

Let $G(\theta, z)$ be the value of $G(\theta)$ if η_ρ/η_α equals $z \in [0,1]$. Let $\theta_p(z)$ be the lowest value in Θ such that $\int_{\hat{\theta}}^{\bar{\theta}} G(\theta, z)\mathrm{d}\theta \leq 0$ for any $\hat{\theta} \geq \theta_p(z)$. Because $G(\theta, z)$ increases in z for a fixed θ, $\theta_p(z)$ also increases in z. If $z = 1$, $G(\theta, 1) = (H(\bar{\theta}) - H(\theta))/H(\bar{\theta})$. Suppose $\theta_p(1) < \bar{\theta}$. I have $\int_{\theta_p(1)}^{\bar{\theta}} G(\theta, 1)\mathrm{d}\theta = \int_{\theta_p(1)}^{\bar{\theta}}(\theta - \theta_p(1))h(\theta)\mathrm{d}/H(\bar{\theta})\theta \leq 0$. This implies that $h(\theta) = 0$ for all $\theta \in [\theta_p(1), \bar{\theta}]$. A contradiction. Therefore, $\theta_p(1) = \bar{\theta}$. For any $\hat{\theta} \in \Theta$ and $z \in [0,1]$, I have $\int_{\hat{\theta}}^{\bar{\theta}} G(\theta, z)\mathrm{d}\theta = \left(z \int_{\hat{\theta}}^{\bar{\theta}} \theta h(\theta)\mathrm{d}\theta - \hat{\theta} \int_{\hat{\theta}}^{\bar{\theta}} h(\theta)\mathrm{d}\theta\right)/H(\bar{\theta})$. Let z^* be $\min_{\hat{\theta} \in \Theta}\left(\hat{\theta} \int_{\hat{\theta}}^{\bar{\theta}} h(\theta)\mathrm{d}\theta\right)/\left(\int_{\hat{\theta}}^{\bar{\theta}} \theta h(\theta)\mathrm{d}\theta\right)$. If $z \leq z^*$, $\int_{\hat{\theta}}^{\bar{\theta}} G(\theta, z)\mathrm{d}\theta \leq 0$ for any $\hat{\theta} \in \Theta$ and thus $\theta_p(z) = \underline{\theta}$. ∎

If $\theta_p = \underline{\theta}$, all types are pooled. The optimal contract consists of the principal's preferred uninformed bundle. From now on, I focus on the more interesting case in which $\theta_p > \underline{\theta}$.

<center>B.   Imposing a cutoff</center>

To make progress, I assume for the rest of this section and Subsection V.A that the type distribution satisfies the following condition. In the Web Appendix, I examine how results change when this condition fails.

ASSUMPTION 2: *For all $\theta \leq \theta_p$, $1 - G(\theta)$ is nondecreasing.*

When the density function f is differentiable, Assumption 2 is equivalent to the following condition:

$$\frac{\eta_\alpha}{\eta_\alpha - \eta_\rho} \geq -\left(\theta\frac{f'(\theta)}{f(\theta)} + \frac{1}{1+\theta}\right), \quad \forall \theta \leq \theta_p.$$

This condition is satisfied for all density functions that are nondecreasing and holds for the exponential distribution, the log-normal, the Pareto and the Gamma distribution for a subset of their parameters. Also, it is satisfied for any density f with $\theta f'/f$ bounded from below when η_α/η_ρ is sufficiently close to 1.

My next result (Proposition 2) shows that under Assumption 2 the optimal contract takes a very simple form. To describe it formally, I introduce the following:

DEFINITION 1: *The cutoff rule is the contract (w^1, w^0) such that*

$$(w^1(\theta), w^0(\theta)) = \begin{cases} (w_\alpha^1(\theta), w_\alpha^0(\theta)) & \text{if } \theta \leq \theta_p, \\ (w_\alpha^1(\theta_p), w_\alpha^0(\theta_p)) & \text{if } \theta > \theta_p. \end{cases}$$

<center>19</center>

Under the cutoff rule, types below θ_p are offered their preferred bundles $(w_\alpha^1(\theta), w_\alpha^0(\theta))$ whereas types above θ_p are pooled at $(w_\alpha^1(\theta_p), w_\alpha^0(\theta_p))$. I denote the cutoff rule by $(w_{\theta_p}^1, \underline{w}_{\theta_p}^0)$. Figure 5 shows the delegation set corresponding to the cutoff rule. With a slight abuse of notation, I identify a bundle on the southeast boundary of Γ with the slope of the tangent line at that point. As θ varies, the principal's preferred bundle ranges from $\underline{\theta}\eta_\rho$ to $\overline{\theta}\eta_\rho$ and the agent's ranges from $\underline{\theta}\eta_\alpha$ to $\overline{\theta}\eta_\alpha$. The delegation set is the interval between $\underline{\theta}\eta_\alpha$ and $\theta_p\eta_\alpha$. According to Corollary 1, $\theta_p\eta_\alpha$ is smaller than $\overline{\theta}\eta_\rho$. Therefore, the upper bound of the delegation set is lower than $\overline{\theta}\eta_\rho$, the principal's preferred bundle given the highest type $\overline{\theta}$.
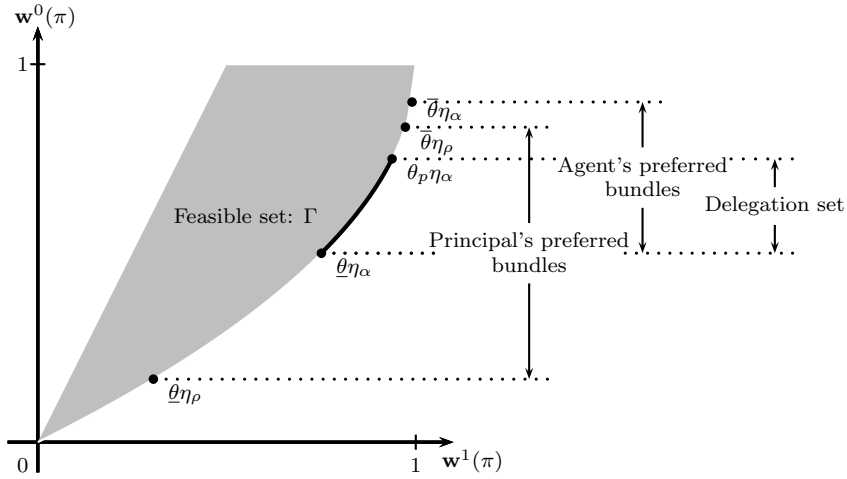

FIGURE 5. DELEGATION SET UNDER CUTOFF RULE

Note: Parameters are $\eta_\alpha = 1, \eta_\rho = 3/5, r/\lambda^1 = 1, \underline{\theta} = 1, \overline{\theta} = 5$ with θ being uniformly distributed. The pooling threshold is $\theta_p \approx 1.99$.

The next proposition shows that $(w_{\theta_p}^1, \underline{w}_{\theta_p}^0)$ is the optimal contract if Assumption 2 holds. Intuitively, Assumption 2 requires that the type distribution cannot be too skewed toward low types. If some low types were very probable, the principal could tailor her menu to exploit this concentration of low types, deviating from the cutoff rule. Assumption 2 ensures that the type distribution is sufficiently smooth, so that the principal has no particular interest to screen certain low types. When this condition holds, the principal finds it optimal to fill in the "holes" in the delegate set. Therefore, the interval delegation set is optimal.[15]

PROPOSITION 2 (Sufficiency):
The cutoff rule $(w_{\theta_p}^1, \underline{w}_{\theta_p}^0)$ is optimal.

[15]The discussion so far also suggests that Assumption 2 is necessary for a cutoff rule to be optimal. In the Web Appendix, I show that no x_p-cutoff contract is optimal for any $x_p \in \Theta$ if Assumption 2 does not hold. The x_p-cutoff contract is defined as $(w^1(\theta), w^0(\theta)) = (w_\alpha^1(\theta), w_\alpha^0(\theta))$ for $\theta < x_p$ and $(w^1(\theta), w^0(\theta)) = (w_\alpha^1(x_p), w_\alpha^0(x_p))$ for $\theta \geq x_p$.

20

To prove Proposition 2, I utilize Lagrangian optimization methods (similar to those used by Amador, Werning and Angeletos (2006)). It suffices to show that $(w_{\theta_p}^1, \underline{w}_{\theta_p}^0)$ maximizes some Lagrangian functional. Then I establish the sufficient first-order conditions and prove that they are satisfied at the conjectured contract and Lagrange multipliers. The detailed proof is relegated to Appendix. In what follows, I first illustrate how to implement the cutoff rule and then prove that it is time-consistent.

C. Properties of the cutoff rule

Implementation

Under the cutoff rule, the agent need not report his type at time 0. Instead, the optimal outcome for the principal can be implemented indirectly by calibrating a constructed belief that the state is 1. It starts with the prior belief $p(\theta_p)$ and then is updated as if the agent had no private information about the state. More specifically, if no success occurs this belief is downgraded according to the differential equation $\dot{p}_t = -\lambda^1 p_t(1 - p_t)$. Upon the first success this belief jumps to one.

The principal imposes a cutoff at p_α^*. As long as the constructed belief stays above the cutoff, the agent can decide whether to continue experimenting or not. As soon as it drops to the cutoff, the agent is not allowed to operate R any more. This rule does not bind for those types below θ_p, who switch to S voluntarily conditional on no success, but does constrain those types above θ_p, who are forced to stop by the principal.

Figure 6 illustrates how the constructed belief evolves over time. The solid arrow shows that the belief is downgraded when R is used and no success has realized. The dashed arrow shows that the belief jumps to one at the first success. The gray area shows that those types below θ_p stop voluntarily as their posterior beliefs drop to the agent's cutoff p_α^*. As illustrated by the black dot, a mass of higher types are required to stop when the cutoff is reached.

There are many other ways to implement the cutoff rule. For example, the constructed belief may start with the prior belief $p(\theta_p \eta_\alpha / \eta_\rho)$ and the principal imposes a cutoff at p_ρ^*. What matters is that the prior belief and the cutoff are chosen collectively to ensure that exactly those types below θ_p are given the freedom to decide whether to experiment or not.

Time consistency

Given that the agent need not report his type at time 0 and that the principal commits to the cutoff rule before experimentation, a natural question to follow is whether the cutoff rule is time-consistent. Does the principal find it optimal to fulfill the contract that she commits to at time 0 after any history? Put differently, were the principal given a chance to offer a new contract at any time, would she benefit from doing so? As my next result shows, the cutoff rule is indeed time-consistent.
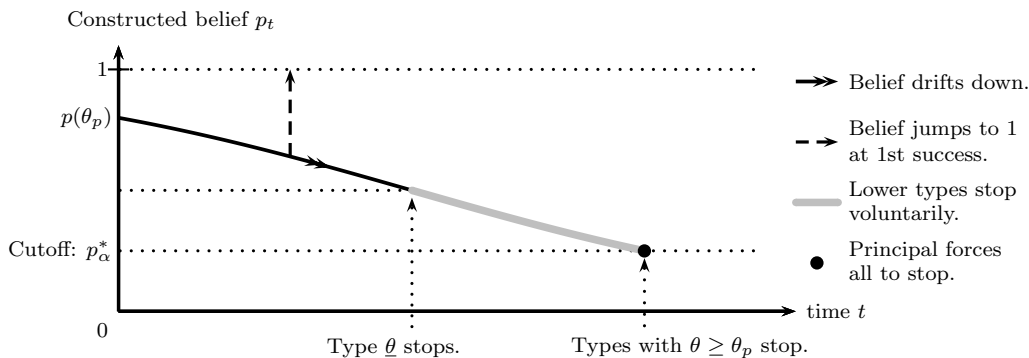
Constructed belief p_t

1

$p(\theta_p)$

Cutoff: p_α^*

0

time t

Type $\underline{\theta}$ stops.　　　Types with $\theta \geq \theta_p$ stop.

→→ Belief drifts down.

- - → Belief jumps to 1 at 1st success.

━━ Lower types stop voluntarily.

● Principal forces all to stop.

FIGURE 6. IMPLEMENTING THE CUTOFF RULE

PROPOSITION 3 (Time consistency):
The cutoff rule is time-consistent.

To show time-consistency, I need to consider three classes of histories on the equilibrium path: (i) the first success occurs before the cutoff is reached; (ii) the agent stops experimenting before the cutoff is reached; (iii) no success has occurred and the agent has not stopped. Clearly, the principal has no incentives to alter the contract after the first two classes of histories. Upon the arrival of the first success, it is optimal to let the agent use R exclusively thereafter. Also, if the agent stops experimenting before the cutoff is reached, his type is revealed. From the principal's point of view, the agent already over-experiments. Hence, she has no incentives to ask the agent to use R any more.

It remains to show that the principal finds the cutoff set at time 0 to be optimal if the agent has not stopped and there is no success. To gain some intuition, it is useful to note that the cutoff rule separates the type space into two segments: low types that are not constrained by the cutoff and high types that are constrained. The cutoff rule does not distort those low types' behaviors at all as they implement their preferred policies. Hence, by setting a cutoff the principal gets no benefit from those low types who are not constrained by the cutoff. Given that the cutoff has no effect on low types' behaviors, the principal chooses the cutoff by looking at only those high types that will be constrained by the cutoff. The cutoff θ_p is chosen such that conditional on the agent's type being above θ_p, the principal finds it optimal to stop experimenting when the cutoff becomes binding. This explains why commitment is not required even if the principal learns more about the agent's type as time passes by.

Here, I sketch how to prove time consistency. First, I calculate the principal's updated belief about the type distribution given no success and that the agent has not stopped. By continuing experimenting, the agent signals that his type is above some level. Hence, the updated type distribution is a truncated one. Since the agent

22

has also updated his belief about the state, I then rewrite the type distribution in terms of the agent's updated odds ratio. Next I show that given the new type distribution the optimal contract is to continue the cutoff rule set at time 0. The detailed proof is relegated to Appendix.

Over- and under-experimentation

The result of Proposition 2 can also be represented by a delegation rule mapping types into stopping times as only stopping-time policies are assigned in equilibrium. Figure 7 depicts such a rule. The x-axis variable is θ, ranging from $\underline{\theta}$ to $\overline{\theta}$. The dotted line represents the agent's preferred stopping time and the dashed line represents the principal's. The delegation rule consists of (i) segment $[\underline{\theta}, \theta_p]$ where the stopping time equals the agent's preferred stopping time and (ii) segment $[\theta_p, \overline{\theta}]$ where the stopping time is independent of the agent's report (*i.e.,* pooling segment). To implement, the principal simply imposes a deadline at $\tau_\alpha(\theta_p)$.

Those types with $\theta \geq \theta_p$ all stop at $\tau_\alpha(\theta_p)$, which is the agent's preferred stopping time given type θ_p. Since the principal's stopping time given type $\theta_p \eta_\alpha/\eta_\rho$ equals the agent's stopping time given θ_p, the delegation rule intersects the principal's stopping time at type $\theta_p \eta_\alpha/\eta_\rho$. From the principal's point of view, those types with $\theta < \theta_p \eta_\alpha/\eta_\rho$ experiment too long while those types with $\theta > \theta_p \eta_\alpha/\eta_\rho$ stop too early.
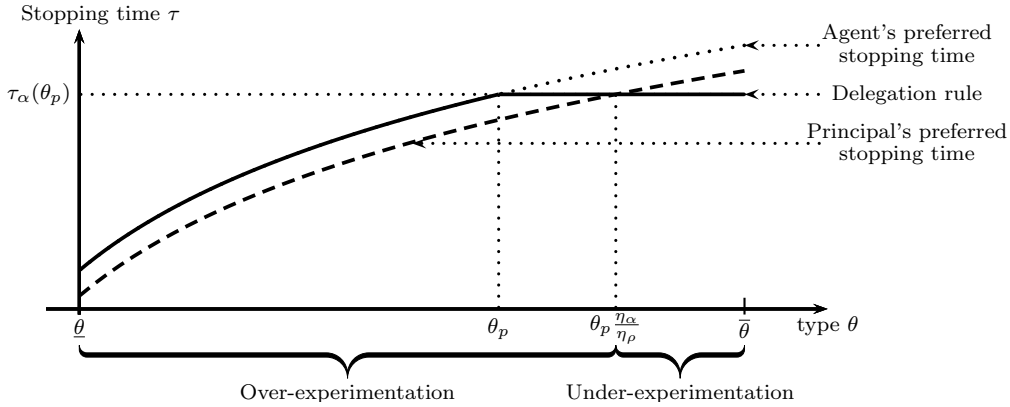
Figure 7. Equilibrium stopping times

Note: Parameters are $\eta_\alpha = 6/5, \eta_\rho = 1, r/\lambda^1 = 1, \underline{\theta} = 1, \overline{\theta} = 5$ with θ being uniformly distributed.

V. Discussions

A. More general stochastic processes

In the benchmark setting, I assume that the risky task never generates any successes in state 0 so the first success reveals that the state is 1. Here, I examine

23

the case in which $0 < \lambda^0 < \lambda^1$. The risky task generates successes in state 0 as well so players never fully learn the state. This happens, for instance, when a bad project generates successes but at a lower intensity. Successes are indicative of a good project but are not conclusive proof. The methodology can be readily extended to incorporate this setting. I first show how to reduce the policy space to a two-dimensional feasible set. Then I prove the optimality of the cutoff rule and its time-consistent property. In the Web Appendix, I extend the analysis to the general class of Lévy bandits (Cohen and Solan (2013)).

For the rest of this subsection, unless otherwise specified I use the same notation as in the benchmark setting. Recall that $(\mathbf{w}^1(\pi), \mathbf{w}^0(\pi))$ denotes the *expected resource* allocated to R under π conditional on state 1 and state 0, and Γ the image of the mapping $(\mathbf{w}^1, \mathbf{w}^0) : \Pi \to [0,1]^2$. The following lemma characterizes the feasible set Γ.

LEMMA 4 (Inconclusive news—feasible set):
There exist two functions $\beta^{se}, \beta^{nw} : [0,1] \to [0,1]$ such that $\Gamma = \{(w^1, w^0) \mid \beta^{se}(w^1) \leq w^0 \leq \beta^{nw}(w^1), w^1 \in [0,1]\}$. The southeast boundary is given by $\beta^{se}(w^1) = 1 - (1 - w^1)^{a^*/(1+a^*)}$, for some constant $a^* > 0$. The northwest boundary β^{nw} is concave, nondecreasing, once continuously differentiable, having end points $(0,0)$ and $(1,1)$.

PROOF:
Based on the proof of Lemma 2, I want to show that the maximum in (5) is achieved by either a lower-cutoff or upper-cutoff policy. If $p_1 \geq 0, p_2 \geq 0$ (resp. $p_1 \leq 0, p_2 \leq 0$), $\max_{w\in\Gamma}(p_1, p_2) \cdot w$ is achieved by the policy which directs all resources to R (resp. S). If $p_1 > 0, p_2 < 0$, according to Keller and Rady (2010), $\max_{w\in\Gamma}(p_1, p_2) \cdot w$ is achieved by a lower-cutoff Markov policy which directs all resource to R if the posterior belief is above the cutoff and to S if below. The cutoff belief, denoted p^*, is given by $p_i^*/(1-p_i^*) = a^*/(1+a^*)$ where a^* is the positive root of equation $r + \lambda^0 - a^*(\lambda^1 - \lambda^0) = \lambda^0(\lambda^0/\lambda^1)^{a^*}$. Let $K(p) \equiv \max_{w\in\Gamma}(p_1, p_2) \cdot w$. If $|p_1|/(|p_1| + |p_2|) \leq p^*$, $K(p)$ equals zero. If $|p_1|/(|p_1| + |p_2|) > p^*$, I obtain $K(p) = -p_2 \left(\frac{-p_2 a^*}{p_1(1+a^*)}\right)^{a^*} /(a^* + 1) + p_1 + p_2$. It is easy to verify that the functional form of the southeast boundary is

$$\beta^{\text{se}}(w^1) = 1 - (1 - w^1)^{\frac{a^*}{a^*+1}}, w^1 \in [0,1].$$

If $p_1 < 0, p_2 > 0$, according to Keller and Rady (forthcoming), $\max_{w\in\Gamma}(p_1, p_2) \cdot w$ is achieved by a upper-cutoff Markov policy under which R is used exclusively if the posterior belief is below the cutoff and S is used if above. Let p^{**} denote this cutoff belief. The function $K(p)$ is continuous, convex, and non-increasing in $|p_1|/(|p_1| + |p_2|)$. Except for a kink at p^{**}, $K(p)$ is once continuously differentiable. Hence, the northwest boundary of Γ is concave, nondecreasing, once continuously differentiable, with the end points being $(0,0)$ and $(1,1)$. ∎

Figure 8 depicts the feasible set when $r = 1/5, \lambda^1 = 2/5, \lambda^0 = (2 - \sqrt{2})/5$. Unlike the benchmark setting, the northwest boundary is characterized by a once continuously differentiable function. My next result shows that the cutoff rule as

defined in Definition 1 is optimal under Assumption 2. This is the case because the proof of Proposition 2, which only relies on the properties of the southeast boundary of the feasible set, applies directly to the current setting.
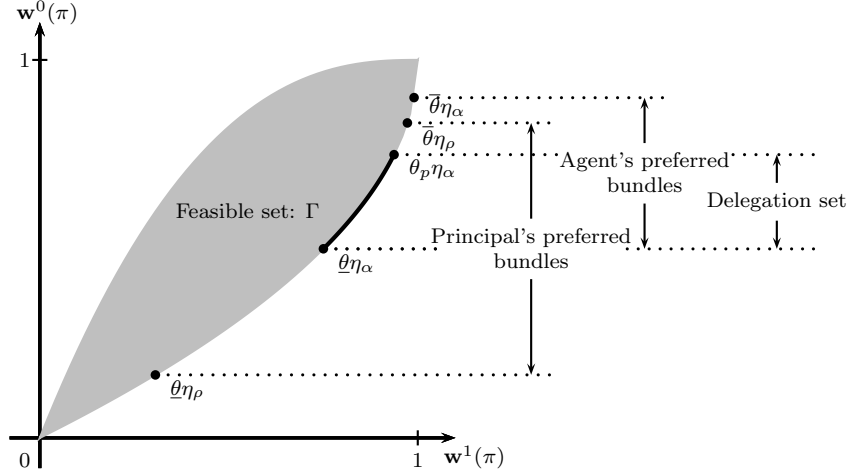

FIGURE 8. DELEGATION SET UNDER CUTOFF RULE: INCONCLUSIVE NEWS CASE

PROPOSITION 4 (Inconclusive news—sufficiency):
The cutoff rule is optimal when $0 < \lambda^0 < \lambda^1$.

PROOF:
According to Lemma 4, the southeast boundary of the feasible set is given by $\beta^{\mathrm{se}}(w^1) = 1 - (1 - w^1)^{a^*/(1+a^*)}$, for some constant $a^* > 0$. The proof of Proposition 2 applies directly here. ∎

The implementation of the cutoff rule, similar to that of the benchmark setting, is achieved by calibrating a constructed belief which starts with the prior belief $p(\theta_p)$. This belief is updated as follows: (i) if no success occurs, the belief drifts down according to the differential equation $\dot{p}_t = -(\lambda^1 - \lambda^0)p_t(1 - p_t)$; and (ii) if a success occurs at time t, the belief jumps from p_{t-} to

$$p_t = \frac{\lambda^1 p_{t-}}{\lambda^1 p_{t-} + \lambda^0(1 - p_{t-})}.$$

The principal imposes a cutoff at p_α^*. The agent can choose to experiment or not if this belief stays above the cutoff and are required to stop when it drops to the cutoff. The gray areas in Figure 9 show that lower types stop voluntarily as their posterior beliefs reach p_α^*. The black dot shows that those types with $\theta > \theta_p$ are required to stop.

Figure 9 also highlights the difference between the benchmark setting, where the belief jumps to one after the first success, and the inconclusive news setting, where

25

the belief jumps up upon successes and then drifts down. Consequently, when successes are inconclusive, the optimum can no longer be implemented by imposing a fixed deadline. Instead, it takes the form of a sliding deadline. The principal initially extends some time to the agent. Whenever a success realizes, more time is granted. The agent is allowed to give up his time voluntarily. That is, the agent can choose to switch to S before he uses up the time granted by the principal. After a long stretch of time without success, the principal requires the agent to switch to S.
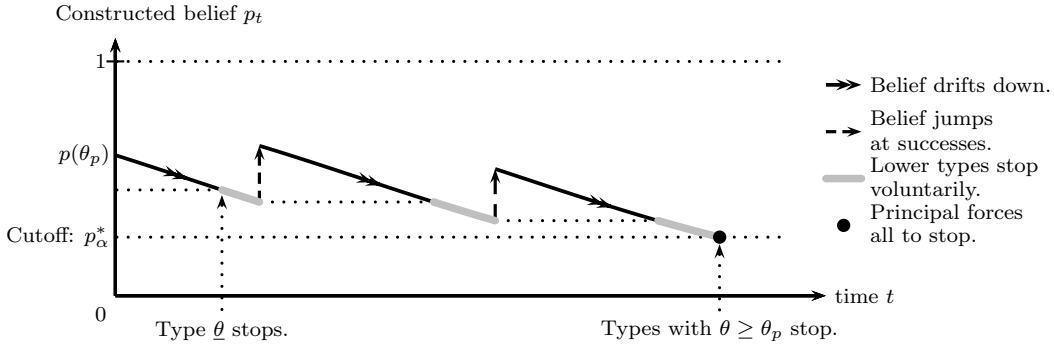

FIGURE 9. IMPLEMENTING THE CUTOFF RULE: INCONCLUSIVE NEWS

The cutoff rule is time consistent when successes are inconclusive. The intuition is the same as in the benchmark setting: since the cutoff creates no distortion on low types' behaviors, the principal chooses the cutoff optimally conditional on the agent being unconstrained by the cutoff. The proof is slightly different than the benchmark setting as successes never fully reveal the state. I need to show that the principal finds it optimal not to adjust the cutoff upon successes. The detailed proof is relegated to Appendix.

PROPOSITION 5 (Inconclusive news—time consistency):
The cutoff rule is time-consistent when $0 < \lambda^0 < \lambda^1$.

B. Biased toward the safe task: a lockup period

In this subsection, I consider the situations in which the agent is biased toward the safe task, *i.e.*, $\eta_\alpha < \eta_\rho$. This happens, for example, when a division (agent) conducts an experiment which yields positive externalities to other divisions. Hence, the agent does not internalize the total benefit that the experiment brings to the organization (principal). Another possibility is that the agent does not perceive the risky task to generate significant career opportunities compared with alternative activities. In both cases, the agent has a preference to stop experimenting earlier.

To illustrate the main intuition, I assume that the lowest type agent prefers a positive length of experimentation, *i.e.*, $\tau_\alpha(\underline{\theta}) > 0$. Using the same methods as in

the main model, I first show that types below some threshold are pooled. Intuitively, types at the bottom prefer to experiment less than what the principal prefers to do for any prior. The cost of separating those types exceeds the benefit. Then, I show that under certain condition the optimal outcome for the principal can be implemented by starting with a properly calibrated prior belief that the state is 1. This belief is then updated as if the agent had no private information. As long as this belief remains above a cutoff belief, the agent is required to operate R. As soon as it drops to the cutoff, the principal keeps her hands off the project and lets the agent decide whether to experiment or not.

In contrast to the main model, the agent has no flexibility until the cutoff belief is reached. I call this mechanism the reversed cutoff rule. Those types with low enough priors stop experimenting as soon as the cutoff is reached. Those with higher priors are not constrained and thus implement their preferred policies. If successes are conclusive, the principal simply sets up a lockup period during which the agent uses R regardless of the outcome. After the lockup period ends, the agent is free to experiment or not. If successes are inconclusive, the principal initially sets up a lockup length. Each time a success occurs, the lockup length is extended. The agent has no freedom until the lockup period ends.

Notably, the reserved cutoff rule is not time-consistent. Whenever those high types voluntarily stop experimenting, they perfectly reveal their types. The principal learns that these high types under-experiment compared with her optimal experimentation policy, tempted to command the agent to experiment further. Therefore, there is tension on the principal's ability to commit.

Given a direct mechanism $(w^1(\theta), w^0(\theta))$, let $U_\alpha(\theta)$ denote the payoff that the agent of type θ gets by maximizing over his report. As the optimal mechanism is truthful, $U_\alpha(\theta)$ equals $\theta\eta_\alpha w^1(\theta) - w^0(\theta)$ and the envelope condition implies that $U'_\alpha(\theta) = \eta_\alpha w^1(\theta)$. By integrating the envelope condition, one obtains the standard integral condition

$$\theta\eta_\alpha w^1(\theta) - w^0(\theta) = \overline{\theta}\eta_\alpha w^1(\overline{\theta}) - \overline{w}^0 - \int_\theta^{\overline{\theta}} \eta_\alpha w^1(\tilde{\theta})\mathrm{d}\tilde{\theta},$$

where $\overline{w}^0$ stands for $w^0(\overline{\theta})$. Substituting $w^0(\theta)$ and simplifying, I reduce the problem to finding a function $w^1 : \Theta \to [0,1]$ and a scalar $\overline{w}^0$ that solves

(OBJ-S) $$\max_{w^1, \overline{w}^0 \in \Phi^s} \left(\overline{\theta}\eta_\alpha w^1(\overline{\theta}) - \overline{w}^0 - \eta_\alpha \int_{\underline{\theta}}^{\overline{\theta}} w^1(\theta)G^s(\theta)\mathrm{d}\theta \right),$$

27

subject to

$$(13) \qquad \beta^{\mathrm{se}}(w^1(\theta)) \leq \theta \eta_\alpha w^1(\theta) + \int_\theta^{\overline{\theta}} \eta_\alpha w^1(\tilde{\theta}) \mathrm{d}\tilde{\theta} - \overline{\theta} \eta_\alpha w^1(\overline{\theta}) + \overline{w}^0, \ \ \forall \theta \in \Theta,$$

$$(14) \qquad \beta^{\mathrm{nw}}(w^1(\theta)) \geq \theta \eta_\alpha w^1(\theta) + \int_\theta^{\overline{\theta}} \eta_\alpha w^1(\tilde{\theta}) \mathrm{d}\tilde{\theta} - \overline{\theta} \eta_\alpha w^1(\overline{\theta}) + \overline{w}^0, \ \ \forall \theta \in \Theta,$$

where

$$\Phi^s \equiv \left\{ w^1, \overline{w}^0 \mid w^1 : \Theta \to [0,1], w^1 \text{ nondecreasing}; \overline{w}^0 \in [0,1] \right\},$$

$$G^s(\theta) = \frac{H(\theta)}{H(\overline{\theta})} + \left(1 - \frac{\eta_\rho}{\eta_\alpha}\right) \frac{\theta h(\theta)}{H(\overline{\theta})}, \text{ where } h(\theta) = \frac{f(\theta)}{1+\theta}, H(\theta) = \int_{\underline{\theta}}^\theta h(\tilde{\theta}) \mathrm{d}\tilde{\theta}.$$

I denote this problem by $\mathcal{P}^s$. Let θ_p^s be the highest value in Θ such that

$$(15) \qquad \int_{\underline{\theta}}^{\hat{\theta}} G^s(\theta) \mathrm{d}\theta \leq 0, \text{ for any } \hat{\theta} \leq \theta_p^s.$$

My next result shows that types with $\theta \leq \theta_p^s$ are pooled.

PROPOSITION 6 (Pooling on bottom):
An optimal contract $(w^{1*}, \overline{w}^{0*})$ satisfies $w^{1*}(\theta) = w^{1*}(\theta_p^s)$ for $\theta \leq \theta_p^s$. It is optimal for (13) or (14) to hold with equality at θ_p^s.

PROOF:
The contribution to (OBJ-S) from types with $\theta < \theta_p^s$ is $-\eta_\alpha \int_{\underline{\theta}}^{\theta_p^s} w^1(\theta) G^s(\theta) \mathrm{d}\theta$. Substituting $w^1(\theta) = w^1(\theta_p^s) - \int_\theta^{\theta_p^s} \mathrm{d}w^1$ and integrating by parts, I obtain
(16)
$$-\eta_\alpha \int_{\underline{\theta}}^{\theta_p^s} w^1(\theta) G^s(\theta) \mathrm{d}\theta = -\eta_\alpha w^1(\theta_p^s) \int_{\underline{\theta}}^{\theta_p^s} G^s(\theta) \mathrm{d}\theta + \eta_\alpha \int_{\underline{\theta}}^{\theta_p^s} \int_{\underline{\theta}}^\theta G^s(\tilde{\theta}) \mathrm{d}\tilde{\theta} \mathrm{d}w^1(\theta).$$

The first term only depends on $w^1(\theta_p^s)$. The second term depends on $dw^1(\theta)$ for all $\theta \in [\underline{\theta}, \theta_p^s]$. According to the definition of θ_p^s, the integrand of the second term, $\int_{\underline{\theta}}^\theta G^s(\tilde{\theta}) \mathrm{d}\tilde{\theta}$, is weakly negative for all $\theta \in [\underline{\theta}, \theta_p^s]$. Therefore, it is optimal to set $dw^1(\theta) = 0$ for all $\theta \in [\underline{\theta}, \theta_p^s]$. If $\theta_p^s = \overline{\theta}$, all types are pooled. The principal offers her preferred uninformed bundle, which is on the southeast boundary of Γ. If $\theta_p^s < \overline{\theta}$, the first term of (16) is zero as well because $\int_{\underline{\theta}}^{\theta_p^s} G^s(\theta) \mathrm{d}\theta = 0$. Adjusting $w^1(\theta_p^s)$ does not affect the objective function, so $w^1(\theta_p^s)$ can be decreased until either (13) or (14) binds. ∎

For the rest of this subsection, I focus on the more interesting case in which $\theta_p^s < \overline{\theta}$. I first define the reversed cutoff rule.

DEFINITION 2: *The reversed cutoff rule is the contract* (w^1, w^0) *such that*

$$(w^1(\theta), w^0(\theta)) = \begin{cases} (w^1_\alpha(\theta_p^s), w^0_\alpha(\theta_p^s)) & \text{if } \theta \leq \theta_p^s, \\ (w^1_\alpha(\theta), w^0_\alpha(\theta)) & \text{if } \theta > \theta_p^s. \end{cases}$$

I denote this rule by $(w^1_{\theta_p^s}, \overline{w}^0_{\theta_p^s})$. My next result gives a sufficient condition under which the reserved cutoff rule is optimal.

PROPOSITION 7 (Sufficiency-reversed cutoff rule):
The reversed cutoff rule $(w^1_{\theta_p^s}, \overline{w}^0_{\theta_p^s})$ *is optimal if* $G^s(\theta)$ *is nondecreasing when* $\theta \geq \theta_p^s$.

If $f(\theta)$ is differentiable, $G^s(\theta)$ being nondecreasing for $\theta \geq \theta_p^s$ is equivalent to requiring that

$$\frac{\theta f'(\theta)}{f(\theta)} \leq \frac{\eta_\alpha}{\eta_\rho - \eta_\alpha} - \frac{1}{1+\theta}, \quad \forall \theta \in [\theta_p^s, \overline{\theta}].$$

It is satisfied for any density f with $\theta f'/f$ bounded from above when η_ρ/η_α is sufficiently close to 1, or equivalently when two players' preferences are sufficiently aligned. The proof is relegated to Appendix.

C. Optimal contract with transfers

In this subsection, I examine the optimal contract when the principal can make transfers to the agent. From both theoretical and empirical point of view, it is important to know how the results change when transfers are allowed. I assume that the principal has full commitment power, that is, she can write a contract specifying both an experimentation policy π and a transfer scheme c at the outset of the game. I also assume that the agent is protected by limited liability so only non-negative transfers from the principal to the agent are allowed. An experimentation policy π is defined in the same way as before. A transfer scheme c offered by the principal is a non-negative, non-decreasing process $\{c_t\}_{t \geq 0}$, which may depend only on the history of events up to t, where c_t denotes the cumulative transfers the principal has made to the agent up to time t (inclusive).[16] Let Π^* denote the set of all possible policy and transfer scheme pairs. Under perfect commitment, the revelation principle applies. A direct mechanism specifies for each type an experimentation policy and transfer scheme pair. Again, it is difficult to solve for the optimal contract directly as the contract space Π^* is very large.

A methodological contribution of this paper is to represent a stochastic policy by a summary statistic, to characterize the range of all summary statistics and to reduce the dynamic contract problem to a static one. This method is not restricted to the delegation model but can be generalized to the case with transfers. Each policy and transfer pair, in terms of payoffs, can be represented by a summary statistic of four numbers. The first two numbers are the expected resource directed to the risky

---

[16]Formally, the number of successes achieved up to time t (inclusive) defines the point process $\{N_t\}_{t \geq 0}$. Let $\mathcal{F} := \{\mathcal{F}_t\}_{t \geq 0}$ denote the filtration generated by the process π and N_t. The process $\{c_t\}_{t \geq 0}$ is $\mathcal{F}$-adapted.

task conditional on state 1 and that conditional on state 0. The last two numbers are the total discounted expected transfer made to the agent conditional on state 1 and that conditional on state 0, summarizing the transfer payments that the agent expects to receive conditional on the state. Given that both players are risk-neutral, their payoffs can be written as a function of these four numbers. The range of these summary statistics as we vary across all possible policy and transfer scheme pairs can be treated as the feasible set when transfers are allowed. This feasible set is a subset of $\mathbf{R}_+^4$ and can be considered as the new contract space. This reduces the dynamic contracting problem to a static one. Given that the problem is now static, I utilize the optimal control methods (similar to those used by Krishna and Morgan (2008)) to solve for the optimal contract. The detailed analysis can be found in the Web Appendix. Here, I briefly introduce the results, focusing on when and how the transfers are made. (I assume a larger agent's return in this illustration.)

Under certain regularity conditions, when the bias η_α/η_ρ is relatively small the optimal contract consists of three segments, *i.e.*, $[\underline{\theta}, \theta^*]$, $[\theta^*, \theta_p]$ and $[\theta_p, \overline{\theta}]$. When the agent's type is below θ^*, the experimentation policy is between that optimal for the principal and that optimal for the agent. So, compared with the delegation problem, the low types stop earlier when transfers are allowed. In fact, the lowest type's experimentation policy is the socially optimal one and he receives the highest level of transfer payment. As the type increases, the experimentation policy shifts toward the agent's preferred one, with a corresponding decrease in the transfer payments. When $\theta \in [\theta^*, \theta_p]$, the experimentation policy that is preferred by the agent is offered and no transfers are made. It is as if the agent is free to choose any experimentation policy. For types above θ_p, the agent always chooses the bundle preferred by type θ_p. Effectively, there is pooling over $[\theta_p, \overline{\theta}]$. To sum up, the principal makes transfers to the low types, lets intermediate types do whatever they like, and pools the high types. When the bias is small, the only difference when transfers are allowed is that the low types are "bribed" to stop experimentation earlier than in the delegation model. When the bias η_α/η_ρ is large, the optimal contract consists of only two segments which are denoted $[\underline{\theta}, \tilde{\theta}_p]$ and $[\tilde{\theta}_p, \overline{\theta}]$. When $\theta \in [\underline{\theta}, \tilde{\theta}_p]$, the equilibrium policy is between the principal's preferred policy and the agent's preferred one. The contract bundle shifts toward the agent's preferred one as the type increases with a corresponding decrease in the transfers. When $\theta \in [\tilde{\theta}_p, \overline{\theta}]$, all types are pooled. The pooling bundle specifies a lower level of experimentation than what the principal prefers given the pooling segment $[\tilde{\theta}_p, \overline{\theta}]$. There is no segment in which the agent implements his preferred bundle.

VI.   Concluding remarks

This paper discusses how organizations can optimally manage innovative activities, particularly how much control right over resource allocation shall be left to the agent over time with the presence of misaligned preferences and hidden information. From this aspect, this paper contributes to the discussion of how to optimally allocate formal authority and real authority within organizations (Aghion and Tirole (1997)).

The optimal delegation rule requires the agent to achieve a success before the next deadline to keep the project alive. It is simple, time-consistent and already implemented in organizations such as Google. Google encourages its employees to come up with new ideas and build a prototype. If the initial result is satisfactory, Google makes it an official project, funds it and sets the next pair of goal and deadline. The goal must be met before the deadline to secure future funding. Successful products including Gmail, Google Earth and Google Maps survived all the deadlines. Needless to say, many more did not. For example, the once highly publicized and well-funded Google Wave was canceled in August 2010 as it failed to achieve the goal set by Google executives before then.

Besides in-house innovation, my results also apply to the government sector, which often experiments reforms in public policy. The constituents delegate reforms to politicians. Legislatures delegate policy-making to their standing committees. Throughout the process transfers are prohibited to prevent corruption. It has been concluded that every reform has consequences that cannot be fully known until it has been implemented (Strulovici (2010)). The constituents as well as the government learn the effects of a reform that gradually unfold. If a reform is thought to be a failure, it can be legislatively repealed, executively overturned or allowed to automatically expire with a sunset provision. Politicians hope to prolong the policy experimentation as they gain the most popularity from successful reforms that they initiated. My sliding deadline rule suggests that if politicians are better informed on policies, every reform should carry a sunset provision. They should be renewed only upon demonstrated successes.

REFERENCES

Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105(1): 1–29.

Alonso, Ricardo, and Niko Matouschek. 2008. "Optimal Delegation." *Review of Economic Studies*, 75(1): 259–293.

Amador, Manuel, and Kyle Bagwell. 2013. "The Theory of Optimal Delegation with an Application to Tariff Caps." *Econometrica*, 81(4): 1541–1599.

Amador, Manuel, Iván Werning, and George-Marios Angeletos. 2006. "Commitment vs. Flexibility." *Econometrica*, 74(2): 365–396.

Ambrus, Attila, and Georgy Egorov. 2013. "Comment on "Commitment vs. Flexibility"." *Econometrica*, 81(5): 2113–2124.

Aumann, Robert J. 1964. "Mixed and Behavior Strategies in Extensive Games." In *Advances in Game Theory*. 627–650. Princeton, NJ:Princeton University Press.

Bergemann, Dirk, and Ulrich Hege. 1998. "Venture Capital Financing, Moral Hazard, and Learning." *Journal of Banking and Finance*, 22(6): 703–735.

Bergemann, Dirk, and Ulrich Hege. 2005. "The Financing of Innovation: Learning and Stopping." *RAND Journal of Economics*, 36(4): 719–752.

Bolton, Patrick, and Christopher Harris. 1999. "Strategic Experimentation." *Econometrica*, 67(2): 349–374.

Coase, Ronald H. 1937. "The Nature of the Firm." *Economica*, 4: 386–405.

Cohen, Asaf, and Eilon Solan. 2013. "Bandit Problems with Lévy Processes." *Mathematics of Operations Research*, 38(1): 92–107.

Garfagnini, Umberto. 2011. "Delegated Experimentation." Working paper.

Gomes, Renato, Daniel Gottlieb, and Lucas Maestri. 2013. "Experimentation and Project Selection: Screening and Learning." Working paper.

Halac, Marina, Navin Kartik, and Qingmin Liu. 2013. "Optimal Contracts for Experimentation." Working paper.

Holmström, Bengt. 1977. "On Incentives and Control in Organizations." PhD diss. Stanford University.

Holmström, Bengt. 1984. "On the Theory of Delegation." In *Bayesian Models in Economic Theory*. 115–141. New York: North-Holland.

Hörner, Johannes, and Larry Samuelson. 2013. "Incentives for Experimenting Agents." *RAND Journal of Economics*, 44(4): 632–663.

Keller, Godfrey, and Sven Rady. 2010. "Strategic Experimentation with Poisson Bandits." *Theoretical Economics*, 5(2): 275–311.

Keller, Godfrey, and Sven Rady. forthcoming. "Breakdowns." *Theoretical Economics*.

Keller, Godfrey, Sven Rady, and Martin Cripps. 2005. "Strategic Experimentation with Exponential Bandits." *Econometrica*, 73(1): 39–68.

Krishna, Vijay, and John Morgan. 2008. "Contracting for Information under Imperfect Commitment." *RAND Journal of Economics*, 39(4): 905–925.

Luenberger, David G. 1996. *Optimization by Vector Space Method. Series in Decision and Control*, New York: Wiley.

Milgrom, Paul, and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica*, 70(2): 583–601.

Presman, Ernst L. 1990. "Poisson Version of the Two-Armed Bandit Problem with Discounting." *Theory of Probability and Its Applications*, 35: 307–317.

Seierstad, Atle, and Knut Sydsæter. 1987. *Optimal Control Theory with Economic Applications.* Elsevier Science.

Strulovici, Bruno. 2010. "Learning While Voting: Determinants of Collective Experimentation." *Econometrica*, 73(3): 933–971.

Williamson, Oliver E. 1975. *Markets and Hierarchies.* New York: Free Press.

<center>APPENDIX</center>

PROOF OF PROPOSITION 2:

I extend β^{se} to the real line in the following way: for some $\hat{w}^1$ in $(w^1_\alpha(\bar{\theta}), 1)$.[17]

$$\hat{\beta}(w^1) = \begin{cases} (\beta^{\mathrm{se}})'(0)w^1 & \text{if } w^1 \in (-\infty, 0), \\ \beta^{\mathrm{se}}(w^1) & \text{if } w^1 \in [0, \hat{w}^1], \\ \beta^{\mathrm{se}}(\hat{w}^1) + (\beta^{\mathrm{se}})'(\hat{w}^1)(w^1 - \hat{w}^1) & \text{if } w^1 \in (\hat{w}^1, \infty). \end{cases}$$

The function $\hat{\beta}$ is continuously differentiable, convex and lower than β^{se} on $[0,1]$. I then define a problem $\hat{\mathcal{P}}$ which differs from $\mathcal{P}$ in two aspects: (i) the constraint (10) is dropped; (ii) the constraint (9) is replaced with the following:

$$(A1) \qquad \theta\eta_\alpha w^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^\theta w^1(\tilde{\theta})\mathrm{d}\tilde{\theta} - \underline{\theta}\eta_\alpha w^1(\underline{\theta}) + \underline{w}^0 - \hat{\beta}(w^1(\theta)) \geq 0, \quad \forall \theta \in \Theta.$$

If $(w^1, \underline{w}^0)$ satisfies (9) and (10), it also satisfies (A1). Therefore, $\hat{\mathcal{P}}$ is a relaxation of $\mathcal{P}$. If the solution to $\hat{\mathcal{P}}$ is admissible, it is also the solution to $\mathcal{P}$.

Define the Lagrangian functional associated with $\hat{\mathcal{P}}$ as

$$\hat{L}(w^1, \underline{w}^0 \mid \Lambda) = \underline{\theta}\eta_\alpha w^1(\underline{\theta}) - \underline{w}^0 + \eta_\alpha \int_{\underline{\theta}}^{\bar{\theta}} w^1(\theta)G(\theta)\mathrm{d}\theta$$

$$+ \int_{\underline{\theta}}^{\bar{\theta}} \left(\theta\eta_\alpha w^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^\theta w^1(\tilde{\theta})\mathrm{d}\tilde{\theta} - \underline{\theta}\eta_\alpha w^1(\underline{\theta}) + \underline{w}^0 - \hat{\beta}(w^1(\theta)) \right) \mathrm{d}\Lambda,$$

where Λ is the Lagrange multiplier associated with (A1). Fixing a nondecreasing Λ, the Lagrangian is a concave functional on Φ because all terms in $\hat{L}(w^1, \underline{w}^0 \mid \Lambda)$ are linear in $(w^1, \underline{w}^0)$ except $\int_{\underline{\theta}}^{\bar{\theta}} -\hat{\beta}(w^1(\theta))\mathrm{d}\Lambda$ which is concave in w^1. Without loss of generality I set $\Lambda(\bar{\theta}) = 1$. Integrating the Lagrangian by parts yields

$$(A2) \qquad \hat{L}(w^1, \underline{w}^0 \mid \Lambda) = \left(\underline{\theta}\eta_\alpha w^1(\underline{\theta}) - \underline{w}^0 \right) \Lambda(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \left(\theta\eta_\alpha w^1(\theta) - \hat{\beta}(w^1(\theta)) \right) \mathrm{d}\Lambda$$

$$+ \eta_\alpha \int_{\underline{\theta}}^{\bar{\theta}} w^1(\theta) \left[\Lambda(\theta) - (1 - G(\theta)) \right] \mathrm{d}\theta.$$

---

[17] Such a $\hat{w}^1$ exists because $\bar{\theta}$ is finite and hence $w^1_\alpha(\bar{\theta})$ is bounded away from 1.

The following lemma provides a sufficient condition for $(\tilde{w}^1, \underline{\tilde{w}}^0) \in \Phi$ to solve $\hat{\mathcal{P}}$.

LEMMA 5 (Lagrangian—sufficiency):
A contract $(\tilde{w}^1, \underline{\tilde{w}}^0) \in \Phi$ solves $\hat{\mathcal{P}}$ if (A1) holds with equality and there exists a nondecreasing $\tilde{\Lambda}$ such that $\hat{L}(\tilde{w}^1, \underline{\tilde{w}}^0 | \tilde{\Lambda}) \geq \hat{L}(w^1, \underline{w}^0 | \tilde{\Lambda})$, $\forall (w^1, \underline{w}^0) \in \Phi$.

PROOF:
I first introduce the problem studied in Section 8.4 of Luenberger, 1996, p. 220: $\max_{x \in X} Q(x)$ subject to $x \in \Omega$ and $J(x) \in P$, where Ω is a subset of the vector space X, $Q : \Omega \to \mathbf{R}$ and $J : \Omega \to Z$; where Z is a normed vector space, and P is a nonempty positive cone in Z. To apply Theorem 1 in Luenberger, 1996, p. 220, set

(A3) $X = \{w^1, \underline{w}^0 \mid w^1 : \Theta \to \mathbf{R} \text{ and } \underline{w}^0 \in \mathbf{R}\}$, $\Omega = \Phi$,

$\quad Z = \{z \mid z : \Theta \to \mathbf{R} \text{ with } \sup\limits_{\theta \in \Theta} |z(\theta)| < \infty\}$, with the norm $\|z\| = \sup\limits_{\theta \in \Theta} |z(\theta)|$,

$\quad P = \{z \mid z \in Z \text{ and } z(\theta) \geq 0, \ \forall \theta \in \Theta\}$.

I let the objective in (OBJ) be Q and the left-hand side of (A1) be J. This result holds because the hypotheses of Theorem 1 in Luenberger, 1996, p. 220 are met. ∎
To apply Lemma 5 and show that $(\tilde{w}^1, \underline{\tilde{w}}^0)$ maximizes $\hat{L}(w^1, \underline{w}^0 | \tilde{\Lambda})$ for some $\tilde{\Lambda}$, I modify Lemma 1 in Luenberger, 1996, p. 227 which concerns the maximization of a concave functional in a convex cone. Note that Φ is not a convex cone, so Lemma 1 in Luenberger, 1996, p. 227 does not apply directly in the current setting.

LEMMA 6 (First-order conditions):
Let L be a concave functional on Ω, a convex subset of a vector space X. Take $\tilde{x} \in \Omega$. Suppose that the Gâteaux differentials $\partial L(\tilde{x}; x)$ and $\partial L(\tilde{x}; x - \tilde{x})$ exist for any $x \in \Omega$ and that $\partial L(\tilde{x}; x - \tilde{x}) = L(\tilde{x}; x) - L(\tilde{x}; \tilde{x})$.[18] A sufficient condition that $\tilde{x} \in \Omega$ maximizes L over Ω is that $\partial L(\tilde{x}; x) \leq 0$, $\forall x \in \Omega$ and $\partial L(\tilde{x}; \tilde{x}) = 0$.

PROOF:
For $x \in \Omega$ and $0 < \epsilon < 1$, the concavity of L implies that

$$L(\tilde{x} + \epsilon(x - \tilde{x})) \geq L(\tilde{x}) + \epsilon \left(L(x) - L(\tilde{x})\right)$$
$$\Longrightarrow L(x) - L(\tilde{x}) \leq \frac{1}{\epsilon}\left(L(\tilde{x} + \epsilon(x - \tilde{x})) - L(\tilde{x})\right).$$

As $\epsilon \to 0+$, the right-hand side of this equation tends toward $\partial L(\tilde{x}; x - \tilde{x})$. Therefore, I obtain
$$L(x) - L(\tilde{x}) \leq \partial L(\tilde{x}; x - \tilde{x}) = \partial L(\tilde{x}; x) - \partial L(\tilde{x}; \tilde{x}).$$

Given that $\partial L(\tilde{x}; \tilde{x}) = 0$ and $\partial L(\tilde{x}; x) \leq 0$ for all $x \in \Omega$, the sign of $\partial L(\tilde{x}; x - \tilde{x})$ is negative. Therefore, $L(x) \leq L(\tilde{x})$ for all $x \in \Omega$. ∎

---

[18]Let X be a vector space, Y a normed space and $D \subset X$. Given a transformation $T : D \to Y$, if for $\tilde{x} \in D$ and $x \in X$ the limit $\lim_{\epsilon \to 0}\left(T(\tilde{x} + \epsilon x) - T(\tilde{x})\right)/\epsilon$ exists, then it is called the Gâteaux differential at $\tilde{x}$ with direction x and is denoted $\partial T(\tilde{x}; x)$. If the limit exists for each $x \in X$, T is said to be Gâteaux differentiable at $\tilde{x}$.

Next, I prove Proposition 2 based on Lemma 5 and Lemma 6. To apply Lemma 6, let X and Ω be the same as in (A3). Fixing a nondecreasing Λ, the Lagrangian (A2) is a concave functional on Φ. By applying Lemma $A.1$ in Amador, Werning and Angeletos (2006), it is easy to verify that $\partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; w^1, \underline{w}^0 \mid \Lambda)$ and $\partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; (w^1, \underline{w}^0) - (w^1_{\theta_p}, \underline{w}^0_{\theta_p}) \mid \Lambda)$ exist for any $(w^1, \underline{w}^0) \in \Phi$.[19] The linearity condition is also satisfied

$$\partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; (w^1, \underline{w}^0) - (w^1_{\theta_p}, \underline{w}^0_{\theta_p}) \mid \Lambda)$$
$$= \partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; w^1, \underline{w}^0 \mid \Lambda) - \partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; w^1_{\theta_p}, \underline{w}^0_{\theta_p} \mid \Lambda).$$

So the hypotheses of Lemma 6 are met. The Gâteaux differential at $(w^1_{\theta_p}, \underline{w}^0_{\theta_p})$ is

$$\partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; w^1, \underline{w}^0 \mid \Lambda) = \left(\underline{\theta}\eta_\alpha w^1(\underline{\theta}) - \underline{w}^0\right)\Lambda(\underline{\theta}) + \eta_\alpha \int_{\theta_p}^{\bar{\theta}}(\theta - \theta_p)w^1(\theta)\mathrm{d}\Lambda$$

$$\text{(A4)} \qquad\qquad + \eta_\alpha \int_{\underline{\theta}}^{\bar{\theta}} w^1(\theta)\left[\Lambda(\theta) - (1 - G(\theta))\right]\mathrm{d}\theta, \quad \forall(w^1, \underline{w}^0) \in \Phi.$$

Next, I construct a nondecreasing multiplier $\tilde{\Lambda}$ in a similar manner as in Proposition 3 in Amador, Werning and Angeletos (2006), such that the first-order conditions $\partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; w^1, \underline{w}^0 \mid \tilde{\Lambda}) \leq 0$ and $\partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; w^1_{\theta_p}, \underline{w}^0_{\theta_p} \mid \tilde{\Lambda}) = 0$ are satisfied for any $(w^1, \underline{w}^0) \in \Phi$. Let $\tilde{\Lambda}(\underline{\theta}) = 0$, $\tilde{\Lambda}(\theta) = 1 - G(\theta)$ for $\theta \in (\underline{\theta}, \theta_p]$, and $\tilde{\Lambda}(\theta) = 1$ for $\theta \in (\theta_p, \bar{\theta}]$. The jump at $\underline{\theta}$ is upward since $1 - G(\underline{\theta})$ is nonnegative. The jump at θ_p is $G(\theta_p)$, which is nonnegative based on the definition of θ_p. Therefore, $\tilde{\Lambda}$ is nondecreasing. Substituting $\tilde{\Lambda}$ into (A4) yields

$$\partial\hat{L}(w^1_{\theta_p}, \underline{w}^0_{\theta_p}; w^1, \underline{w}^0 \mid \tilde{\Lambda}) = \eta_\alpha \int_{\theta_p}^{\bar{\theta}} w^1(\theta)G(\theta)\mathrm{d}\theta = \eta_\alpha \int_{\theta_p}^{\bar{\theta}}\left(\int_\theta^{\bar{\theta}} G(\tilde{\theta})\mathrm{d}\tilde{\theta}\right)\mathrm{d}w^1(\theta),$$

where the last equality follows by integrating by parts, which can be done given the monotonicity of w^1 and by the definition of θ_p. This Gâteaux differential is zero at $(w^1_{\theta_p}, \underline{w}^0_{\theta_p})$ and, by the definition of θ_p, it is nonpositive for all w^1 nondecreasing. It follows that the first-order conditions are satisfied for all $(w^1, \underline{w}^0) \in \Phi$. By Lemma 6, $(w^1_{\theta_p}, \underline{w}^0_{\theta_p})$ maximizes $\hat{L}(w^1, \underline{w}^0 \mid \tilde{\Lambda})$ over Φ. By Lemma 5, $(w^1_{\theta_p}, \underline{w}^0_{\theta_p})$ solves $\hat{\mathcal{P}}$. Because $(w^1_{\theta_p}, \underline{w}^0_{\theta_p})$ is admissible, it solves $\mathcal{P}$. ∎

PROOF OF PROPOSITION 3:

Let $\theta^*_\alpha = p^*_\alpha/(1 - p^*_\alpha)$ be the odds ratio at which the agent is indifferent between continuing and stopping. After operating R for $\delta > 0$ without success, the agent of type θ updates his odds ratio to $\theta e^{-\lambda^1\delta}$, referred to as his type at time δ. Let $\underline{\theta}_\delta = \max\{\underline{\theta}, \theta^*_\alpha e^{\lambda^1\delta}\}$. After a period of δ with no success, only those types above $\underline{\theta}_\delta$

[19]The observations made by Ambrus and Egorov (2013) do not apply to the results of Amador, Werning and Angeletos (2006) that my proof relies on.

remain. The principal's updated belief about the agent's type distribution, in terms of his type at time 0, is given by

$$
f_\delta^0(\theta) = \begin{cases} \dfrac{\left[1-p(\theta)\left(1-e^{-\lambda^1\delta}\right)\right]f(\theta)}{\int_{\underline{\theta}_\delta}^{\overline{\theta}}\left[1-p(\theta)\left(1-e^{-\lambda^1\delta}\right)\right]f(\theta)\mathrm{d}\theta} & \text{if } \theta \in [\underline{\theta}_\delta, \overline{\theta}], \\ 0 & \text{otherwise.} \end{cases}
$$

Here, $1 - p(\theta)(1 - e^{-\lambda^1\delta})$ is the probability that no success occurs from time 0 to δ conditional on the agent's type being θ at time 0. The principal's belief about the agent's type distribution, in terms of his type at time δ, is given by

$$
f_\delta(\theta) = \begin{cases} f_\delta^0(\theta e^{\lambda^1\delta})e^{\lambda^1\delta} & \text{if } \theta \in [\underline{\theta}_\delta e^{-\lambda^1\delta}, \overline{\theta}e^{-\lambda^1\delta}], \\ 0 & \text{otherwise.} \end{cases}
$$

I prove that continuing the cutoff rule is optimal by showing two things. First, given f_δ at time δ, the threshold of the pooling segment is $\theta_p e^{-\lambda^1\delta}$. Second, if Assumption 2 holds for $\theta \le \theta_p$ under f, it holds for $\theta \le \theta_p e^{-\lambda^1\delta}$ under f_δ. Given f_δ over $\theta \in [\underline{\theta}_\delta e^{-\lambda^1\delta}, \overline{\theta}e^{-\lambda^1\delta}]$, I define $h_\delta, H_\delta, G_\delta$ as follows:

$$
h_\delta(\theta) = \frac{f_\delta(\theta)}{1+\theta}, \quad \text{and} \quad H_\delta(\theta) = \int_{\underline{\theta}_\delta e^{-\lambda^1\delta}}^{\theta} h_\delta(\tilde{\theta})\mathrm{d}\tilde{\theta},
$$

$$
G_\delta(\theta) = \frac{H_\delta(\overline{\theta}e^{-\lambda^1\delta}) - H_\delta(\theta)}{H_\delta(\overline{\theta}e^{-\lambda^1\delta})} + \left(\frac{\eta_\rho}{\eta_\alpha} - 1\right)\theta\frac{h_\delta(\theta)}{H_\delta(\overline{\theta}e^{-\lambda^1\delta})}.
$$

Substituting $f_\delta(\theta) = f_\delta^0(\theta e^{\lambda^1\delta})e^{\lambda^1\delta}$ into $h_\delta(\theta)$ and simplifying, I obtain

$$
h_\delta(\theta) = \frac{f(\theta e^{\lambda^1\delta})e^{\lambda^1\delta}}{C(1+\theta e^{\lambda^1\delta})}, \quad \text{where } C = \int_{\underline{\theta}_\delta}^{\overline{\theta}}\left[1 - p(\theta)\left(1 - e^{-\lambda^1\delta}\right)\right]f(\theta)\mathrm{d}\theta.
$$

By simplifying and making a change of variables $z = e^{\lambda^1\delta}\theta$, I obtain the following

$$
\int_{\theta_p e^{-\lambda^1\delta}}^{\overline{\theta}e^{-\lambda^1\delta}}\left(\eta_\rho\theta - \eta_\alpha\theta_p e^{-\lambda^1\delta}\right)\frac{f_\delta(\theta)}{1+\theta}\mathrm{d}\theta = \frac{1}{Ce^{\lambda^1\delta}}\int_{\theta_p}^{\overline{\theta}}(\eta_\rho z - \eta_\alpha\theta_p)\frac{f(z)}{1+z}\mathrm{d}z.
$$

Therefore, if the threshold of the pooling segment given f is θ_p, then the threshold of the pooling segment at time δ is $\theta_p e^{-\lambda^1\delta}$. The condition required by Assumption 2 is that $1 - G_\delta(\theta)$ is nondecreasing for $\theta \le \theta_p e^{-\lambda^1\delta}$. The term $1 - G_\delta(\theta)$ can be written in terms of $f(\theta)$ by making a change of variables $z = e^{\lambda^1\delta}\theta$

$$
1 - G_\delta(\theta) = \frac{1}{C}\left[\int_{\underline{\theta}_\delta}^{z}\frac{f(\tilde{z})}{1+\tilde{z}}\mathrm{d}\tilde{z} + \left(\frac{\eta_\rho}{\eta_\alpha} - 1\right)z\frac{f(z)}{1+z}\right],
$$

36

If Assumption 2 holds for all $\theta \leq \theta_p$ given f, it holds for all $\theta \leq \theta_p e^{-\lambda^1 \delta}$ given f_δ. ∎

PROOF OF PROPOSITION 5:

Suppose that a success occurs at time t. Before the success, the type distribution is denoted f_{t-}. Without loss of generality, suppose that the support is $[\underline{\theta}, \overline{\theta}]$. Let $Q(\theta, dt)$ be the probability that a success occurs in an infinitesimal interval $[t, t+dt)$ given type θ: $Q(\theta, dt) = p(\theta)(1 - e^{-\lambda^1 dt}) + (1 - p(\theta))(1 - e^{-\lambda^0 dt})$. Given a success at time t, the principal's updated belief about the agent's type distribution, in terms of his type at time $t-$, is

$$f_t^*(\theta) = \lim_{dt \to 0} \frac{f_{t-}(\theta)Q(\theta, dt)}{\int_{\underline{\theta}}^{\overline{\theta}} f_{t-}(\theta)Q(\theta, dt)\mathrm{d}\theta} = \frac{f_{t-}(\theta)\left[p(\theta)\lambda^1 + (1 - p(\theta))\lambda^0\right]}{\int_{\underline{\theta}}^{\overline{\theta}} f_{t-}(\theta)\left[p(\theta)\lambda^1 + (1 - p(\theta))\lambda^0\right]\mathrm{d}\theta}.$$

After the success, the agent of type θ updates his odds ratio to $\theta\lambda^1/\lambda^0$. Therefore, the principal's belief about the agent's type distribution, in terms of his type at time t, is

$$f_t(\theta) = \begin{cases} f_t^*\left(\theta\lambda^0/\lambda^1\right)\lambda^0/\lambda^1 & \text{if } \theta \in \left[\underline{\theta}\lambda^1/\lambda^0, \overline{\theta}\lambda^1/\lambda^0\right], \\ 0 & \text{otherwise.} \end{cases}$$

Following the same argument as in Subsection A, I can show that (i) if the threshold of the pooling segment given f_{t-} is θ_p, the threshold of the pooling segment given f_t is $\theta_p\lambda^1/\lambda^0$; (ii) if Assumption 2 holds for $\theta \leq \theta_p$ given f_{t-}, it holds for $\theta \leq \theta_p\lambda^1/\lambda^0$ given f_t. ∎

PROOF OF PROPOSITION 7:

I define a problem $\hat{\mathcal{P}}^s$ which differs from $\mathcal{P}^s$ in two aspects: (i) the constraint (14) is dropped; and (ii) the constraint (13) is replaced with the following:

$$\theta\eta_\alpha w^1(\theta) + \int_\theta^{\overline{\theta}} \eta_\alpha w^1(\tilde{\theta})\mathrm{d}\tilde{\theta} - \overline{\theta}\eta_\alpha w^1(\overline{\theta}) + \overline{w}^0 - \hat{\beta}(w^1(\theta)) \geq 0, \quad \forall \theta \in \Theta.$$

Define the Lagrangian functional associated with $\hat{\mathcal{P}}^s$ as

$$\hat{L}^s(w^1, \overline{w}^0 \mid \Lambda) = \overline{\theta}\eta_\alpha w^1(\overline{\theta}) - \overline{w}^0 - \eta_\alpha \int_{\underline{\theta}}^{\overline{\theta}} w^1(\theta)G^s(\theta)\mathrm{d}\theta$$

$$+ \int_{\underline{\theta}}^{\overline{\theta}} \left(\theta\eta_\alpha w^1(\theta) + \int_\theta^{\overline{\theta}} \eta_\alpha w^1(\tilde{\theta})\mathrm{d}\tilde{\theta} - \overline{\theta}\eta_\alpha w^1(\overline{\theta}) + \overline{w}^0 - \hat{\beta}(w^1(\theta))\right)\mathrm{d}\Lambda.$$

Based on Lemma 5 and 6, it suffices to construct a nondecreasing Λ such that the first-order conditions $\partial\hat{L}^s(w^1_{\theta_p^s}, \overline{w}^0_{\theta_p^s}; w^1, \overline{w}^0 \mid \Lambda) \leq 0$ and $\partial\hat{L}^s(w^1_{\theta_p^s}, \overline{w}^0_{\theta_p^s}; w^1_{\theta_p^s}, \overline{w}^0_{\theta_p^s} \mid \Lambda) = 0$ are satisfied for any $(w^1, \overline{w}^0) \in \Phi^s$. Let $\Lambda(\theta) = 0$ for $\theta \in [\underline{\theta}, \theta_p^s)$, $\Lambda(\theta) = G^s(\theta)$ for $\theta \in [\theta_p^s, \overline{\theta})$, and $\Lambda(\overline{\theta}) = 1$. The jump at θ_p^s is nonnegative according to the definition of θ_p^s. The jump at $\overline{\theta}$ is nonnegative because $G^s(\theta) \leq 1$ for all θ. If $G^s(\theta)$ is nondecreasing for $\theta \in [\theta_p, \overline{\theta}]$, $\Lambda(\theta)$ is nondecreasing. ∎

37

B1. A special case: two types

In this section, I study the delegation problem with binary types, high type θ_h or low type θ_l. Let $q(\theta)$ denote the probability that the agent's type is θ. Formally, I solve for $(w^1, w^0) : \{\theta_l, \theta_h\} \rightarrow \Gamma$ such that

$$\max_{w^1, w^0} \sum_{\theta \in \{\theta_l, \theta_h\}} q(\theta) \left(\frac{\theta}{1+\theta} \eta_\rho w^1(\theta) - \frac{1}{1+\theta} w^0(\theta) \right),$$

$$\text{subject to} \quad \theta_l \eta_\alpha w^1(\theta_l) - w^0(\theta_l) \geq \theta_l \eta_\alpha w^1(\theta_h) - w^0(\theta_h),$$

$$\theta_h \eta_\alpha w^1(\theta_h) - w^0(\theta_h) \geq \theta_h \eta_\alpha w^1(\theta_l) - w^0(\theta_l).$$

For ease of exposition, I refer to the contract for the low (high) type agent as the low (high) type contract and the principal who believes to face the low (high) type agent as the low (high) type principal. Let $(w^{1*}(\theta_l), w^{0*}(\theta_l))$ and $(w^{1*}(\theta_h), w^{0*}(\theta_h))$ denote the equilibrium bundles. The optimum is characterized as follows.

PROPOSITION 8 (Two types):
Suppose that $(r + \lambda^1)\theta_l \eta_\rho / r > 1$. There exists a $b' \in (1, \theta_h/\theta_l)$ such that

1.1 If $\eta_\alpha/\eta_\rho \in [1, b']$, the principal's preferred bundles are implementable.

1.2 If $\eta_\alpha/\eta_\rho \in (b', \theta_h/\theta_l)$, separating is optimal. The low type contract is a stopping-time policy, the stopping time between $\tau_\rho(\theta_l)$ and $\tau_\alpha(\theta_l)$. The low type's IC constraint binds and the high type's does not.

1.3 If $\eta_\alpha/\eta_\rho \geq \theta_h/\theta_l$, pooling is optimal.

In all cases, the optimum can be attained using bundles on the boundary of Γ.

Without loss of generality, the presumption $(r + \lambda^1)\theta_l \eta_\rho / r > 1$ ensures that both the low type principal's preferred stopping time $\tau_\rho(\theta_l)$ and the high type principal's preferred stopping time $\tau_\rho(\theta_h)$ are strictly positive. The degenerate cases of $\tau_\rho(\theta_h) > \tau_\rho(\theta_l) = 0$ and $\tau_\rho(\theta_h) = \tau_\rho(\theta_l) = 0$ yield similar results to Proposition 8 and thus are omitted.

Proposition 8 describes the optimal contract as the bias level varies. According to result (*1.1*), if the bias is low enough, the principal simply offers her preferred policies given θ_l and θ_h. This is incentive compatible because even though the low type agent prefers longer experimentation than the low type principal, at a low bias level he still prefers the low type principal's preferred bundle instead of the high type principal's. Consequently the principal pays no informational rents. This result does not hold with a continuum of types. The principal's preferred bundles are two points on the southeast boundary of Γ with binary types, but they become an interval on the southeast boundary with a continuum of types in which case lower types are strictly better off mimicking higher types.

The result (1.2) corresponds to medium bias level. As the bias has increased, offering the principal's preferred policies is no longer incentive compatible. Instead, both the low type contract and the high type one deviate from the principal's preferred policies. The low type contract is always a stopping-time policy while the high type contract takes one of three possible forms: stopping-time, slack-after-success or delay policies.[20] One of the latter two forms is assigned as the high type contract if the agent's type is likely to be low and his bias is relatively large. All three forms are meant to impose a significant cost—excessive experimentation, constrained exploitation of success, or delay in experimentation—on the high type contract so as to deter the low type agent from misreporting. However the principal can more than offset the cost by effectively shortening the low type agent's experimentation. In the end, the low type agent over-experiments slightly and the high type contract deviates from the principal's preferred policy $(w_\rho^1(\theta_h), w_\rho^0(\theta_h))$ as well. One interesting observation is that the optimal contract can take a form other than a stopping-time policy.

If the bias is even higher, as shown by result (1.3), pooling is preferable. The condition $\eta_\alpha/\eta_\rho \geq \theta_h/\theta_l$ has an intuitive interpretation that the low type agent prefers to experiment longer than even the high type principal. The screening instruments utilized in result (1.2) impair the high type principal's payoff more than the low type agent's. As a result, the principal is better off offering her uninformed preferred bundle. Notably, for fixed types the prior probabilities of the types do not affect the pooling decision. Only the bias level does.

I make two observations. First, the principal chooses to take advantage of the agent's private information unless the agent's bias is too large. This result also applies to the continuous type case. Second, the optimal contract can be tailored to the likelihood of the two types. For example, if the type is likely to be low, the principal designs the low type contract close to her low type bundle and purposefully makes the high type contract less attractive to the low type agent. Similarly, if the type is likely to be high, the principal starts with a high type contract close to her high type bundle without concerning about the low type's over-experimentation. This "type targeting", however, becomes irrelevant when the principal faces a continuum of types and has no incentives to target certain types.

PROOF OF PROPOSITION 8:

Let α_l (resp. α_h) denote the low (resp. high) type agent and ρ_l (resp. ρ_h) the low (resp. high) type principal. Suppose that $\theta_l \eta_\rho > r/(\lambda^1 + r)$. Both ρ_l's and ρ_h's preferred bundles lie in the interior of Γ^{st}. Given that $\theta_h > \theta_l$ and $\eta_\alpha > \eta_\rho$, the slopes of players' indifference curves are ranked as follows

$$\theta_h \eta_\alpha > \max\{\theta_h \eta_\rho, \theta_l \eta_\alpha\} \geq \min\{\theta_h \eta_\rho, \theta_l \eta_\alpha\} > \theta_l \eta_\rho.$$

---

[20]Here, I give an example in which the high type contract is a slack-after-success policy. Parameters are $\eta_\alpha = 6, \eta_\rho = 1, \theta_l = 3/2, \theta_h = 19, r = \lambda^1 = 1$. The agent's type is low with probability 2/3. The optimum is $(w^{1*}(\theta_h), w^{0*}(\theta_h)) \approx (0.98, 1)$ and $(w^{1*}(\theta_l), w^{0*}(\theta_l)) \approx (0.96, 0.79)$.

Let ICL and ICH denote α_l's and α_h's IC constraints. Let I_{α_l} denote α_l's indifference curves. If α_l prefers $(w_\rho^1(\theta_l), w_\rho^0(\theta_l))$ to $(w_\rho^1(\theta_h), w_\rho^0(\theta_h))$, the optimum is

$$\{(w_\rho^1(\theta_l), w_\rho^0(\theta_l)), (w_\rho^1(\theta_h), w_\rho^0(\theta_h))\}.$$

This is true when the slope of the line connecting $(w_\rho^1(\theta_l), w_\rho^0(\theta_l))$ and $(w_\rho^1(\theta_h), w_\rho^0(\theta_h))$ is greater than $\theta_l \eta_\alpha$. This condition is satisfied when η_α / η_ρ is bounded from above by

$$b' \equiv \frac{\theta_h(\lambda^1 + r)\left(\theta_h^{\frac{r}{\lambda^1}} - \theta_l^{\frac{r}{\lambda^1}}\right)}{r\left(\theta_h^{\frac{r+\lambda^1}{\lambda^1}} - \theta_l^{\frac{r+\lambda^1}{\lambda^1}}\right)}.$$

If this condition does not hold, at least one IC constraint binds. I explain how to find the optimal bundles.

1) ICL binds. Suppose not. It must be the case that ICH binds and that the principal offers two distinct bundles $(w^1(\theta_l), w^0(\theta_l)) < (w^1(\theta_h), w^0(\theta_h))$ which lie on the same indifference curve of α_h. Given that $\theta_h \eta_\alpha > \max\{\theta_h \eta_\rho, \theta_l \eta_\rho\}$, both ρ_h and ρ_l strictly prefer $(w^1(\theta_l), w^0(\theta_l))$ to $(w^1(\theta_h), w^0(\theta_h))$. The principal is strictly better off by offering a pooling bundle $(w^1(\theta_l), w^0(\theta_l))$. A contradiction. Hence, ICL binds.

2) If $\theta_h \eta_\rho < \theta_l \eta_\alpha$, the optimum is pooling. Suppose not. Suppose that the principal offers two distinct bundles $(w^1(\theta_l), w^0(\theta_l)) < (w^1(\theta_h), w^0(\theta_h))$ which are on the same indifference curve of α_l. Given that $\theta_l \eta_\rho < \theta_h \eta_\rho < \theta_l \eta_\alpha$, α_l's indifference curves are steeper than ρ_h's and ρ_l's. Both ρ_h and ρ_l strictly prefer $(w^1(\theta_l), w^0(\theta_l))$ to $(w^1(\theta_h), w^0(\theta_h))$. The principal is strictly better off by offering a pooling bundle $(w^1(\theta_l), w^0(\theta_l))$. A contradiction. If $\theta_h \eta_\rho = \theta_l \eta_\alpha$, ρ_h has the same indifference curves as α_l. If $\{(w^1(\theta_l), w^0(\theta_l)), (w^1(\theta_h), w^0(\theta_h))\}$ is optimal, it is optimal for the principal to offer a pooling contract $(w^1(\theta_l), w^0(\theta_l))$.

3) If $\theta_h \eta_\rho > \theta_l \eta_\alpha$, the optimum are on the boundary of Γ. Suppose not. Suppose that $(w^1(\theta_l), w^0(\theta_l))$ or $(w^1(\theta_h), w^0(\theta_h))$ is in the interior. The indifference curve of α_l going through $(w^1(\theta_l), w^0(\theta_l))$ intersects the boundary at $(\tilde{w}^1(\theta_l), \tilde{w}^0(\theta_l))$ and $(\tilde{w}^1(\theta_h), \tilde{w}^0(\theta_h))$ such that $\tilde{w}^1(\theta_l) < \tilde{w}^1(\theta_h)$. Given that $\theta_h \eta_\rho > \theta_l \eta_\alpha > \theta_l \eta_\rho$, ρ_h prefers $(\tilde{w}^1(\theta_h), \tilde{w}^0(\theta_h))$ to $(w^1(\theta_h), w^0(\theta_h))$ and ρ_l prefers $(\tilde{w}^1(\theta_l), \tilde{w}^0(\theta_l))$ to $(w^1(\theta_l), w^0(\theta_l))$. The principal is strictly better off by offering $(\tilde{w}^1(\theta_l), \tilde{w}^0(\theta_l))$ and $(\tilde{w}^1(\theta_h), \tilde{w}^0(\theta_h))$. Therefore, the optimal bundles are on the boundary. The problem is reduced to locate the low type agent's indifference curve on which $(w^{1*}(\theta_l), w^{0*}(\theta_l))$ and $(w^{1*}(\theta_h), w^{0*}(\theta_h))$ lie. This indifference curve must be between the indifference curves of α_l which go through $(w_\rho^1(\theta_l), w_\rho^0(\theta_l))$ and $(w_\rho^1(\theta_h), w_\rho^0(\theta_h))$.

■

In this subsection, I first discuss the economic content of Assumption 2. To do so, I consider a delegation problem in which only bundles on Γ^{st}, the southeast boundary of Γ, are considered. This restriction reduces the action space to a one-dimensional line segment and allows me to compare Assumption 2 with the condition identified in Alonso and Matouschek (2008). Then I show that no cut-off rule is optimal if Assumption 2 fails.

Recall that Γ^{st} is characterized by the function $\beta^{\text{se}}(w^1) = 1-(1-w^1)^{r/(r+\lambda^1)}, \forall w^1 \in [0,1]$. The function β^{se} is twice continuously differentiable. The derivative $(\beta^{\text{se}})'(w^1)$ strictly increases in w^1 and approaches infinity as w^1 approaches 1. I identify an element $(w^1, \beta^{\text{se}}(w^1)) \in \Gamma^{\text{st}}$ with the derivative $(\beta^{\text{se}})'(w^1)$ at that point. The set of possible derivatives is denoted $Y = [(\beta^{\text{se}})'(0), \infty]$. Since there is a one-to-one mapping between Γ^{st} and Y, I let Y be the action space and refer to $y \in Y$ as an action. The principal simply assigns a non-empty subset of Y as the delegation set. Let $n(y) = ((\beta^{\text{se}})')^{-1}(y)$ be the inverse of the mapping from w^1 to the derivative $(\beta^{\text{se}})'(w^1)$.

Player i's preferred action given θ is $y_i(\theta) = \eta_i\theta$. Player i's payoff given type θ and action y is denoted

$$V_i(\theta, y) = \eta_i \frac{\theta}{1+\theta} n(y) - \frac{1}{1+\theta}\beta^{\text{se}}(n(y)).$$

I first solve the principal's preferred action if she believes that the agent's type is below θ. The principal chooses $y \in Y$ to maximize $\int_{\underline{\theta}}^{\theta} V_\rho(\tilde{\theta}, y) f(\tilde{\theta})\mathrm{d}\tilde{\theta}$. The maximum is achieved by choosing action

$$\eta_\rho \frac{\int_{\underline{\theta}}^{\theta} \tilde{\theta}h(\tilde{\theta})\mathrm{d}\tilde{\theta}}{H(\theta)}.$$

Following Alonso and Matouschek (2008), I define the backward bias for a given type θ as

$$T(\theta) \equiv \frac{H(\theta)}{H(\bar{\theta})}\left(\eta_\alpha\theta - \eta_\rho\frac{\int_{\underline{\theta}}^{\theta} \tilde{\theta}h(\tilde{\theta})\mathrm{d}\tilde{\theta}}{H(\theta)}\right).$$

Here, $T(\theta)$ measures the difference between the agent's preferred action given θ and the principal's preferred action if she believes that the type is below θ. It is easy to verify that $T(\theta) = \eta_\alpha \int_{\underline{\theta}}^{\theta}(1 - G(\tilde{\theta}))\mathrm{d}\tilde{\theta}$.

Assumption 2 is equivalent to requiring that the backward bias is convex when $\theta \le \theta_p$. This is the condition that Alonso and Matouschek (2008) find for the interval delegation set to be optimal in their setting. Intuitively, when this condition holds, the principal finds it optimal to fill in the "holes" in the delegate set. I shall emphasize that this is not a proof of the optimality of the cutoff rule, because considering only bundles on the southeast boundary might be restrictive. For example, I have

shown that with two types the optimal contract involves bundles which are not on the southeast boundary for certain parameter values. With a continuum of types, there exist examples such that the principal is strictly better off by offering policies other than stopping-time policies. By using Lagrangian methods, I prove that the cutoff rule is indeed optimal under Assumption 2. In my setting, the principal's preferred bundle is not more sensitive than the agent's to the agent's private information and Assumption 2 ensures that the type distribution is sufficiently smooth so the principal has no particular interest to screen some types. Hence, the interval delegation set is optimal.

The discussion so far suggests that Assumption 2 is also necessary for the cutoff rule to be optimal. In the rest of this subsection, I show that no x_p-cutoff contract is optimal for any $x_p \in \Theta$ if Assumption 2 does not hold. The x_p-cutoff contract is defined as $(w^1(\theta), w^0(\theta)) = (w_\alpha^1(\theta), w_\alpha^0(\theta))$ for $\theta < x_p$ and $(w^1(\theta), w^0(\theta)) = (w_\alpha^1(x_p), w_\alpha^0(x_p))$ for $\theta \geq x_p$. The x_p-cutoff contract is denoted $(w_{x_p}^1, \underline{w}_{x_p}^0)$.

Define the Lagrangian functional associated with $\mathcal{P}$ as

(B1) $L(w^1, \underline{w}^0 \mid \Lambda^{\mathrm{se}}, \Lambda^{\mathrm{nw}}) = \underline{\theta}\eta_\alpha w^1(\underline{\theta}) - w^0(\underline{\theta}) + \eta_\alpha \int_{\underline{\theta}}^{\overline{\theta}} w^1(\theta)G(\theta)\mathrm{d}\theta$

$+ \int_{\underline{\theta}}^{\overline{\theta}} \left(\theta\eta_\alpha w^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^{\theta} w^1(\tilde{\theta})\mathrm{d}\tilde{\theta} - \underline{\theta}\eta_\alpha w^1(\underline{\theta}) + \underline{w}^0 - \beta^{\mathrm{se}}(w^1(\theta)) \right) \mathrm{d}\Lambda^{\mathrm{se}}$

$+ \int_{\underline{\theta}}^{\overline{\theta}} \left[\beta^{\mathrm{nw}}(w^1(\theta)) - \left(\theta\eta_\alpha w^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^{\theta} w^1(\tilde{\theta})\mathrm{d}\tilde{\theta} - \underline{\theta}\eta_\alpha w^1(\underline{\theta}) + \underline{w}^0 \right) \right] d\Lambda^{\mathrm{nw}},$

where the function $\Lambda^{\mathrm{se}}, \Lambda^{\mathrm{nw}}$ are the Lagrange multiplier associated with constraints (9) and (10). I first show that if $(w_{x_p}^1, \underline{w}_{x_p}^0)$ is optimal for some x_p, there must exist some Lagrange multipliers $\tilde{\Lambda}^{\mathrm{se}}, \tilde{\Lambda}^{\mathrm{nw}}$ such that $L(w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}}, \tilde{\Lambda}^{\mathrm{nw}})$ is maximized at $(w_{x_p}^1, \underline{w}_{x_p}^0)$. Since any x_p-cutoff contract is continuous, I can restrict attention to the set of continuous contracts

$\hat{\Phi} \equiv \left\{ w^1, \underline{w}^0 \mid w^1 : \Theta \to [0, 1], w^1 \text{ nondecreasing and continuous}; \underline{w}^0 \in [0, 1] \right\}.$

LEMMA 7 (Lagrangian—necessity):
If $(w_{x_p}^1, \underline{w}_{x_p}^0)$ solves $\mathcal{P}$, there exist nondecreasing functions $\tilde{\Lambda}^{se}, \tilde{\Lambda}^{nw} : \Theta \to \mathbf{R}$ such that
$$L(w_{x_p}^1, \underline{w}_{x_p}^0 \mid \tilde{\Lambda}^{se}, \tilde{\Lambda}^{nw}) \geq L(w^1, \underline{w}^0 \mid \tilde{\Lambda}^{se}, \tilde{\Lambda}^{nw}), \;\; \forall(w^1, \underline{w}^0) \in \hat{\Phi}.$$

Furthermore, it is the case that

(B2)
$$0 = \int_{\underline{\theta}}^{\overline{\theta}} \left(\theta \eta_\alpha w_{x_p}^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^{\theta} w_{x_p}^1(\tilde{\theta}) \, d\tilde{\theta} - \underline{\theta} \eta_\alpha w_{x_p}^1(\underline{\theta}) + \underline{w}_{x_p}^0 - \beta^{se}(w_{x_p}^1(\theta)) \right) d\tilde{\Lambda}^{se}$$

$$+ \int_{\underline{\theta}}^{\overline{\theta}} \left[\beta^{nw}(w_{x_p}^1(\theta)) - \left(\theta \eta_\alpha w_{x_p}^1(\theta) - \eta_\alpha \int_{\underline{\theta}}^{\theta} w_{x_p}^1(\tilde{\theta}) \, d\tilde{\theta} - \underline{\theta} \eta_\alpha w_{x_p}^1(\underline{\theta}) + \underline{w}_{x_p}^0 \right) \right] d\tilde{\Lambda}^{nw}.$$

PROOF:

I first introduce the problem studied in section 8.4 of Luenberger, 1996, p. 217: $\max_{x \in X} Q(x)$ subject to $x \in \Omega$ and $J(x) \in P$, where Ω is a convex subset of the vector space X, $Q : \Omega \to \mathbf{R}$ and $J : \Omega \to Z$ are both concave; where Z is a normed vector space, and P is a nonempty positive cone in Z. To apply Theorem 1 in Luenberger, 1996, p. 217, set

$$X = \{w^1, \underline{w}^0 \mid \underline{w} \in \mathbf{R} \text{ and } w^1 : \Theta \to \mathbf{R}\},$$
$$\Omega = \hat{\Phi},$$
$$Z = \{z \mid z : \Theta \to \mathbf{R}^2 \text{ with } \sup_{\theta \in \Theta} \|z(\theta)\| < \infty\},$$
$$\text{with the norm} \|z\| = \sup_{\theta \in \Theta} \|z(\theta)\|,$$
$$P = \{z \mid z \in Z \text{ and } z(\theta) \geq (0,0), \ \forall \theta \in \Theta\}.$$

I let the objective function in (OBJ) be Q and the left-hand side of (9) and (10) be defined as J. It is easy to verify that both Q and J are concave. This result holds because the hypotheses of Theorem 1 in Luenberger, 1996, p. 217 are met. ∎

My next result shows that no x_p-cutoff contract is optimal if Assumption 2 fails.

PROPOSITION 9:
If Assumption 2 does not hold, then no x_p-cutoff contract is optimal for any $x_p \in \Theta$.

PROOF:

The proof proceeds by contradiction. Suppose that $(w_{x_p}^1, \underline{w}_{x_p}^0)$ is optimal for some $x_p \in \Theta$. According to Lemma 7, there exist nondecreasing $\tilde{\Lambda}^{se}, \tilde{\Lambda}^{nw}$ such that the Lagrangian (B1) is maximized at $(w_{x_p}^1, \underline{w}_{x_p}^0)$ and (B2) holds. This implies that $\tilde{\Lambda}^{nw}$ is constant so the integral related to $\tilde{\Lambda}^{nw}$ can be dropped. Without loss of generality I set $\tilde{\Lambda}^{se}(\overline{\theta}) = 1$. Integrating the Lagrangian by parts yields

$$L(w^1, \underline{w}^0 \mid \tilde{\Lambda}^{se}) = \left(\underline{\theta} \eta_\alpha w^1(\underline{\theta}) - \underline{w}^0 \right) \tilde{\Lambda}^{se}(\underline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} \left(\theta \eta_\alpha w^1(\theta) - \beta^{se}(w^1(\theta)) \right) d\tilde{\Lambda}^{se}$$

$$+ \eta_\alpha \int_{\underline{\theta}}^{\overline{\theta}} w^1(\theta) \left[\tilde{\Lambda}^{se}(\theta) - (1 - G(\theta)) \right] d\theta.$$

43

Then, I establish the necessary first-order conditions for $L(w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}})$ to be maximized at x_p-cutoff rule and show that they cannot be satisfied if Assumption 2 fails.

Let $a, b \in \Theta$ be such that $a < b < \theta_p$ and $1 - G(a) > 1 - G(b)$ (so Assumption 2 does not hold). It is easy to verify that the Gâteaux differential $\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}})$ exists for any $(w^1, \underline{w}^0) \in \hat{\Phi}$. I want to show that a necessary condition that $(w^1_{x_p}, \underline{w}^0_{x_p})$ maximizes $L(w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}})$ over $\hat{\Phi}$ is that

(B3) $$\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}}) \leq 0, \ \forall (w^1, \underline{w}^0) \in \hat{\Phi},$$

(B4) $$\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; w^1_{x_p}, \underline{w}^0_{x_p} \mid \tilde{\Lambda}^{\mathrm{se}}) = 0.$$

If $(w^1_{x_p}, \underline{w}^0_{x_p})$ maximizes $L(w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}})$, then for any $(w^1, \underline{w}^0) \in \hat{\Phi}$, it must be true that

$$\frac{d}{d\epsilon} L((w^1_{x_p}, \underline{w}^0_{x_p}) + \epsilon((w^1, \underline{w}^0) - (w^1_{x_p}, \underline{w}^0_{x_p})) \mid \tilde{\Lambda}^{\mathrm{se}}) \bigg|_{\epsilon=0} \leq 0.$$

Hence, $\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; (w^1, \underline{w}^0) - (w^1_{x_p}, \underline{w}^0_{x_p}) \mid \tilde{\Lambda}^{\mathrm{se}}) \leq 0$. Setting $(w^1, \underline{w}^0) = (w^1_{x_p}, \underline{w}^0_{x_p})/2 \in \hat{\Phi}$ yields $\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; w^1_{x_p}, \underline{w}^0_{x_p} \mid \tilde{\Lambda}^{\mathrm{se}}) \geq 0$. By the definition of $(w^1_{x_p}, \underline{w}^0_{x_p})$, there exists $\epsilon > 0$ sufficiently small such that $(1 + \epsilon)(w^1_{x_p}, \underline{w}^0_{x_p}) \in \hat{\Phi}$. Setting $(w^1, \underline{w}^0)$ to be $(1 + \epsilon)(w^1_{x_p}, \underline{w}^0_{x_p})$ yields $\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; w^1_{x_p}, \underline{w}^0_{x_p} \mid \tilde{\Lambda}^{\mathrm{se}}) \leq 0$. Together, (B3) and (B4) obtain.

The last step is to show that there exists no $\tilde{\Lambda}^{\mathrm{se}}$ that satisfies the first-order conditions (B3) and (B4). Here, I use the same approach as in the proof of Proposition 4 in Amador, Werning and Angeletos (2006). The Gâteaux differential $\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}})$ is similar to (A4) with θ_p replaced by x_p. Conditions (B3) and (B4) imply that $\tilde{\Lambda}^{\mathrm{se}}(\underline{\theta}) = 0$. Integrating the Gâteaux differential by parts yields

(B5) $$\partial L(w^1_{x_p}, \underline{w}^0_{x_p}; w^1, \underline{w}^0 \mid \tilde{\Lambda}^{\mathrm{se}}) = \chi(\underline{\theta}) w^1(\underline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} \chi(\theta) \mathrm{d}w^1(\theta),$$

with

$$\chi(\theta) \equiv \eta_\alpha \int_\theta^{\overline{\theta}} \left[\tilde{\Lambda}^{\mathrm{se}}(\tilde{\theta}) - (1 - G(\tilde{\theta})) \right] \mathrm{d}\tilde{\theta} + \eta_\alpha \int_{\max\{x_p, \theta\}}^{\overline{\theta}} (\tilde{\theta} - x_p) \mathrm{d}\tilde{\Lambda}^{\mathrm{se}}(\tilde{\theta}).$$

By condition (B3), it follows that $\chi(\theta) \leq 0$ for all θ. Condition (B4) implies that $\chi(\theta) = 0$ for $\theta \in [\underline{\theta}, x_p]$. It follows that $\tilde{\Lambda}^{\mathrm{se}}(\theta) = 1 - G(\theta)$ for all $\theta \in (\underline{\theta}, x_p]$. This implies that $x_p \leq b$ otherwise the associated multiplier $\tilde{\Lambda}^{\mathrm{se}}$ would be decreasing. Integrating by parts the second term of $\chi(\theta)$, I obtain

$$\chi(\theta) = \int_\theta^{\overline{\theta}} G(\tilde{\theta}) \mathrm{d}\tilde{\theta} + (\theta - x_p)(1 - \tilde{\Lambda}^{\mathrm{se}}(\theta)), \ \forall \theta \geq x_p.$$

44

By definition of θ_p, there must exist a $\theta \in [x_p, \theta_p)$ such that the first term is strictly positive; since $\tilde{\Lambda}^{\mathrm{se}}(\theta) \leq 1$, the second term is nonnegative. Hence $\chi(\theta) > 0$, contradicting the necessary conditions. ∎

B3. Lévy processes and Lévy bandits

Here, I extend the analysis to the more general Lévy bandits (Cohen and Solan (2013)). The risky task's payoff is driven by a Lévy process whose Lévy triplet depends on an unknown binary state. In what follows, I start with a reminder about Lévy processes and Lévy bandits. Then, I show that the optimality of the cutoff rule and its time consistency property generalize to Lévy bandits.

A Lévy process $L = (L(t))_{t \geq 0}$ is a continuous-time stochastic process that (i) starts at the origin: $L(0) = 0$; (ii) admits càdlàg modification;[21] (iii) has stationary independent increments. Examples of Lévy processes include a Brownian motion, a Poisson process, and a compound Poisson process.

Let (Ω, P) be the underlying probability space. For every Borel measurable set $A \in \mathcal{B}(\mathbf{R} \setminus \{0\})$, and every $t \geq 0$, let the Poisson random measure $N(t, A)$ be the number of jumps of L in the time interval $[0, t]$ with jump size in A: $N(t, A) = \#\{0 \leq s \leq t \mid \Delta L(s) \equiv L(s) - L(s-) \in A\}$. The measure ν defined by

$$\nu(A) \equiv \mathbf{E}[N(1, A)] = \int N(1, A)(\omega) \mathrm{d}P(\omega).$$

is called the Lévy measure of the process L.

I focus on Lévy processes that have finite expectation for each t. For a fixed Lévy process L, there exists a constant $\mu \in \mathbf{R}$, a Brownian motion $\sigma Z(t)$ with standard deviation $\sigma \geq 0$, and an independent Poisson random measure $N_\nu(t, dh)$ with the associated Lévy measure ν such that, for each $t \geq 0$, the Lévy-Itō decomposition of $L(t)$ is

$$L(t) = \mu t + \sigma Z(t) + \int_{\mathbf{R} \setminus \{0\}} h \tilde{N}_\nu(t, \mathrm{d}h),$$

where $\tilde{N}_\nu(t, A) \equiv N_\nu(t, A) - t\nu(A)$ is the compensated Poisson random measure.[22] Hence, a Lévy process L is characterized by a triplet $\langle \mu, \sigma, \nu \rangle$.

The agent operates a two-armed bandit in continuous time, with a safe task S that yields a known flow payoff s_i to player i, and a risky task R whose payoff, depending on an unknown state $x \in \{0, 1\}$, is given by the process L^x. For ease of exposition, I assume that both players derive the same payoff from R but different payoffs from S. For a fixed state x, L^x is a Lévy process characterized by the triplet $\langle \mu^x, \sigma^x, \nu^x \rangle$. For an arbitrary prior p that the state is 1, I denote by P_p the probability measure over space of realized paths.

I keep the same assumptions (A1–A6) on the Lévy processes L^x as in Cohen and

---

[21] It is continuous from the right and has limits from the left.

[22] Consider a set $A \in \mathcal{B}(\mathbf{R} \setminus \{0\})$ and a function $f : \mathbf{R} \to \mathbf{R}$. The integral with respect to a Poisson random measure $N(t, A)$ is defined as $\int_A f(h) N(t, \mathrm{d}h) = \sum_{s \leq t} f(\Delta L(s)) \mathbb{1}_A(\Delta(L(s)))$.

Solan (2013) and modify A5 to ensure that both players prefer to use R in state 1 and S in state 0. That is, $\mu^1 > s_i > \mu^0$, for $i \in \{\alpha, \rho\}$.[23] Let $\eta_i = (\mu^1 - s_i)/(s_i - \mu^0)$ denote player i's net gain from the experimentation. I assume that the agent gains more from the experimentation, *i.e.*, $\eta_\alpha > \eta_\rho$.[24]

A (pure) allocation policy is a non-anticipative stochastic process $\pi = \{\pi_t\}_{t \geq 0}$. Here, $\pi_t \in [0, 1]$ (resp. $1 - \pi_t$) may be interpreted as the fraction of the resource in the interval $[t, t + dt)$ that is devoted to R (resp. S), which may depend only on the history of events up to t.[25] The space of all policies, including randomized ones, is denoted Π. (See Footnote 9.)

Player i's payoff given a policy $\pi \in \Pi$ and a prior belief $p \in [0, 1]$ that the state is 1 is

$$U_i(\pi, p) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt} \left[\mathrm{d}L^x \left(\int_0^t \pi_s ds \right) + (1 - \pi_t) s_i \mathrm{d}t \right] \;\middle|\; \pi, p \right].$$

Over an interval $[t, t + dt)$, if the fraction π_t of the resource is allocated to R, the expected payoff increment to player i conditional on x is $[(1 - \pi_t)s_i + \pi_t\mu^x]dt$. By the Law of Iterated Expectations, player i's payoff can be written as the discounted sum of the expected payoff increments

$$U_i(\pi, p) = \mathbf{E}\left[\int_0^\infty re^{-rt} \left[\pi_t\mu^x + (1 - \pi_t)s_i \right] \mathrm{d}t \;\middle|\; \pi, p \right].$$

For a fixed policy π, I define $\mathbf{w}^1(\pi)$ and $\mathbf{w}^0(\pi)$ as follows:

$$\mathbf{w}^1(\pi) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt}\pi_t \mathrm{d}t \;\middle|\; \pi, 1 \right] \text{ and } \mathbf{w}^0(\pi) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt}\pi_t \mathrm{d}t \;\middle|\; \pi, 0 \right].$$

Then, player i's payoff can be written as

$$U_i(\pi, p) = p\left(\mu^1 - s_i\right)\mathbf{w}^1(\pi) + (1 - p)\left(\mu^0 - s_i\right)\mathbf{w}^0(\pi) + s_i.$$

Let Γ denote the image of the mapping $(\mathbf{w}^1, \mathbf{w}^0) : \Pi \to [0, 1]^2$, referred to as the feasible set. The following lemma characterizes the southeast boundary of Γ.

---

[23]The assumptions are (A1) $\mathbf{E}[(L^x)^2(1)] = (\mu^x)^2 + (\sigma^x)^2 + \int h^2 \nu^x(\mathrm{d}h) < \infty$; (A2) $\sigma^1 = \sigma^0$; (A3) $|\nu^1(\mathbf{R} \setminus \{0\}) - \nu^0(\mathbf{R} \setminus \{0\})| < \infty$; (A4) $|\int h(\nu^1(\mathrm{d}h) - \nu^0(\mathrm{d}h))| < \infty$; (A5) $\mu^0 < s_\alpha < s_\rho < \mu^1$; (A6) For every $A \in \mathcal{B}(\mathbf{R} \setminus \{0\})$, $\nu^0(A) < \nu^1(A)$. Assumption (A1) states that both L^1 and L^0 have finite quadratic variation. It follows that both have finite expectation. Assumptions (A2) to (A4) ensure that players cannot distinguish between the two states in any infinitesimal time. Assumption (A5) states that the expected payoff rate of R is higher than that of S in state 1 and lower in state 0. The last assumption (A6) requires that jumps of any size h, both positive or negative, occur more often in state 1 than in state 0. Consequently, jumps always provide good news, and increase the posterior belief of state 1.

[24]The results generalize to the case in which, for a fixed state x, the drift term of the Lévy process L^x differs for the principal and the agent, as long as the relation $\eta_\alpha > \eta_\rho$ holds.

[25]Suppose the process L is a Lévy process L^1 with probability $p \in (0, 1)$ and L^0 with probability $1 - p$. Let $\mathcal{F}_s^L$ be the sigma-algebra generated by the process $(L(t))_{t \leq s}$. Then it is required that the process π satisfies that $\{\int_0^t \pi_s ds \leq t'\} \in \mathcal{F}_{t'}^L$, for any $t, t' \in [0, \infty)$.

LEMMA 8:   *There exists $a^* > 0$ such that the southeast boundary of Γ is given by*

$$\{(w^1, w^0) \mid w^0 = 1 - (1 - w^1)^{a^*/(1+a^*)}, w^1 \in [0, 1]\}$$

PROOF:

Based on the proof of Lemma 2, I want to show that the maximum in (5) is achieved by either a lower-cutoff policy when $p_1 \geq 0, p_2 \leq 0$. If $p_1 \geq 0, p_2 \geq 0$ ($p_1 \leq 0, p_2 \leq 0$), $\max_{w \in \Gamma}(p_1, p_2) \cdot w$ is achieved by the policy which directs all resources to R (S). If $p_1 > 0, p_2 < 0$, according to Cohen and Solan (2013), $\max_{w \in \Gamma}(p_1, p_2) \cdot w$ is achieved by a lower-cutoff Markov policy which directs all resource to R if the posterior belief is above the cutoff and to S if below. The cutoff belief, denoted p^*, satisfies the equation $p^*/(1 - p^*) = a^*/(1 + a^*)$, where a^* is the positive root of Equation 6.1 in Cohen and Solan (2013). Let $K(p) \equiv \max_{w \in \Gamma}(p_1, p_2) \cdot w$. If $|p_1|/(|p_1| + |p_2|) \leq p^*$, $K(p)$ equals zero. If $|p_1|/(|p_1| + |p_2|) > p^*$, I obtain $K(p) = -p_2 \left(\frac{-p_2 a^*}{p_1(1+a^*)} \right)^{a^*} /(a^* + 1) + p_1 + p_2$. It is easy to verify that the functional form of the southeast boundary is

$$\beta^{\text{se}}(w^1) = 1 - (1 - w^1)^{\frac{a^*}{a^*+1}}, w^1 \in [0, 1].$$

∎

Given Lemma 8, the proof of Proposition 2, which only relies on the properties of the southeast boundary of the feasible set, applies directly to the current setting. Therefore, the cutoff rule is optimal.

PROPOSITION 10 (Lévy bandits—sufficiency):
The cutoff rule is optimal if Assumption 2 holds.

For every prior $p \in [0, 1]$ that the state is 1, the probability measure P_p satisfies $P_p = pP_1 + (1-p)P_0$. An important auxiliary process is the Radon-Nikodym density, given by

$$\psi_t \equiv \frac{d(P_0 \mid \mathcal{F}_{K(t)})}{d(P_1 \mid \mathcal{F}_{K(t)})}, \quad \text{where } K(t) = \int_0^t \pi_s \mathrm{d}s \text{ and } t \in [0, \infty).$$

According to Lemma 1 in Cohen and Solan (2013), if the prior belief is p, the posterior belief at time t is given by

$$p_t = \frac{p}{p + (1-p)\psi_t}.$$

The agent of type θ updates his belief about the state. He assigns odds ratio θ/ψ_t to the state being 1, referred to as his type at time t. Let $\underline{\theta}_t = \max\{\underline{\theta}, \theta_\alpha^* \psi_t\}$. Recall that θ_α^* denotes the odds ratio at which the agent is indifferent between continuing and stopping. At time t, only those types above $\underline{\theta}_t$ remain. The principal's updated belief about the agent's type distribution, in terms of his type at time 0, is given by

the density function

$$f_t^0(\theta) = \begin{cases} \dfrac{[p(\theta)+(1-p(\theta))\psi_t]f(\theta)}{\int_{\underline{\theta}_t}^{\overline{\theta}}[p(\theta)+(1-p(\theta))\psi_t]f(\theta)\mathrm{d}\theta} & \text{if } \theta \in [\underline{\theta}_t,\overline{\theta}], \\ 0 & \text{otherwise.} \end{cases}$$

The principal's belief about the agent's type distribution, in terms of his type at time t, is given by the density function

$$f_t(\theta) = \begin{cases} f_t^0(\theta\psi_t)\psi_t & \text{if } \theta \in [\underline{\theta}_t/\psi_t,\overline{\theta}/\psi_t], \\ 0 & \text{otherwise.} \end{cases}$$

I prove that continuing the cutoff rule is optimal by showing two things. First, given the distribution f_t at time t, the threshold of the top pooling segment is θ_p/ψ_t. Second, if Assumption 2 holds for $\theta \leq \theta_p$ under distribution f, then it holds for $\theta \leq \theta_p/\psi_t$ under f_t. The detailed proof is similar to that of Proposition 5 (see Appendix) and hence omitted.

PROPOSITION 11 (Lévy bandits—time consistency):
If Assumption 2 holds, the cutoff rule is time-consistent.

<div align="center">

B4. Optimal contract with transfers

</div>

<div align="center">

THE SET-UP

</div>

In this section, I discuss the optimal contract when the principal can make transfers to the agent. I assume that the principal has full commitment power, that is, she can write a contract specifying both an experimentation policy π and a transfer scheme c at the outset of the game. I also assume that the agent is protected by limited liability so only non-negative transfers from the principal to the agent are allowed. An experimentation policy π is defined in the same way as before. A transfer scheme c offered by the principal is a non-negative, non-decreasing process $\{c_t\}_{t\geq 0}$, which may depend only on the history of events up to t, where c_t denotes the cumulative transfers the principal has made to the agent up to, and including, time t.[26] Let Π^* denote the set of all possible policy and transfer scheme pairs.

For any policy and transfer scheme pair (π,c) and any prior p, the principal's and the agent's payoffs are respectively

$$U_\alpha(\pi,c,p) = \mathbf{E}\left[\int_0^\infty re^{-rt}\left[(1-\pi_t)s_\alpha + \pi_t\lambda^\omega h_\alpha\right]\mathrm{d}t + \int_0^\infty re^{-rt}\mathrm{d}c_t \;\middle|\; \pi,c,p\right]$$

$$U_\rho(\pi,c,p) = \mathbf{E}\left[\int_0^\infty re^{-rt}\left[(1-\pi_t)s_\rho + \pi_t\lambda^\omega h_\rho\right]\mathrm{d}t - \int_0^\infty re^{-rt}\mathrm{d}c_t \;\middle|\; \pi,c,p\right].$$

---

[26]Formally, the number of successes achieved up to, and including, time t defines the point process $\{N_t\}_{t\geq 0}$. Let $\mathcal{F} := \{\mathcal{F}_t\}_{t\geq 0}$ denote the filtration generated by the process π and N_t. The process $\{c_t\}_{t\geq 0}$ is $\mathcal{F}$-adapted.

For a given policy and transfer scheme pair (π, c), I define $\mathbf{t}^1(\pi, c)$ and $\mathbf{t}^0(\pi, c)$ as follows:
(B6)

$$\mathbf{t}^1(\pi, c) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt}\mathrm{d}c_t \,\Big|\, \pi, c, 1\right] \text{ and } \mathbf{t}^0(\pi, c) \equiv \mathbf{E}\left[\int_0^\infty re^{-rt}\mathrm{d}c_t \,\Big|\, \pi, c, 0\right].$$

I refer to $\mathbf{t}^1(\pi, c)$ (resp. $\mathbf{t}^0(\pi, c)$) as the expected transfer in state 1 (resp. state 0).

LEMMA 9 (A policy and transfer scheme pair as four numbers):
For a given policy and transfer scheme pair $(\pi, c) \in \Pi^*$ and a given prior $p \in [0, 1]$, the principal's and the agent's payoffs can be written as

$$
\begin{aligned}
U_\alpha(\pi, c, p) =&\, p\left[\left(\lambda^1 h_\alpha - s_\alpha\right)\mathbf{w}^1(\pi) + \mathbf{t}^1(\pi, c)\right] \\
&+ (1-p)\left[\left(\lambda^0 h_\alpha - s_\alpha\right)\mathbf{w}^0(\pi) + \mathbf{t}^0(\pi, c)\right] + s_\alpha \\
U_\rho(\pi, c, p) =&\, p\left[\left(\lambda^1 h_\rho - s_\rho\right)\mathbf{w}^1(\pi) - \mathbf{t}^1(\pi, c)\right] \\
&+ (1-p)\left[\left(\lambda^0 h_\rho - s_\rho\right)\mathbf{w}^0(\pi) - \mathbf{t}^0(\pi, c)\right] + s_\rho.
\end{aligned}
$$

PROOF:
The agent's payoff given $(\pi, c) \in \Pi^*$ and prior $p \in [0, 1]$ is

$$
\begin{aligned}
U_\alpha(\pi, c, p) =&\, p\mathbf{E}\left[\int_0^\infty re^{-rt}\pi_t\left(\lambda^1 h_\alpha - s_\alpha\right)\mathrm{d}t + \int_0^\infty re^{-rt}\mathrm{d}c_t \,\Big|\, \pi, c, 1\right] \\
&+ (1-p)\mathbf{E}\left[\int_0^\infty re^{-rt}\pi_t\left(\lambda^0 h_\alpha - s_\alpha\right)\mathrm{d}t + \int_0^\infty re^{-rt}\mathrm{d}c_t \,\Big|\, \pi, c, 0\right] + s_\alpha \\
=&\, p\left[\left(\lambda^1 h_\alpha - s_\alpha\right)\mathbf{w}^1(\pi) + \mathbf{t}^1(\pi, c)\right] \\
&+ (1-p)\left[\left(\lambda^0 h_\alpha - s_\alpha\right)\mathbf{w}^0(\pi) + \mathbf{t}^0(\pi, c)\right] + s_\alpha.
\end{aligned}
$$

The principal's payoff can be rewritten similarly. ∎

Lemma 9 shows that all payoffs from implementing (π, c) can be written in terms of its expected resource and expected transfer pairs. Instead of working with a generic policy/transfer scheme pair, it is without loss of generality to focus on its expected resource and expected transfer pairs. The image of the mapping $(\mathbf{w}^1, \mathbf{w}^0, \mathbf{t}^1, \mathbf{t}^0)$: $\Pi^* \to [0, 1]^2 \times [0, \infty) \times [0, \infty)$ can be interpreted as the new contract space when transfers are allowed. The following lemma characterizes this contract space.

LEMMA 10: *The image of the mapping $(\mathbf{w}^1, \mathbf{w}^0, \mathbf{t}^1, \mathbf{t}^0)$: $\Pi^* \to [0, 1]^2 \times [0, \infty) \times [0, \infty)$, denoted Γ^*, satisfies the following condition*

$$\mathrm{int}(\Gamma \times [0, \infty)^2) \subset \Gamma^* \subset \Gamma \times [0, \infty)^2.$$

PROOF:
The relation $\Gamma^* \subset \Gamma \times [0, \infty)^2$ is obviously true. Hence, I only need to show that $\mathrm{int}(\Gamma \times [0, \infty)^2) \subset \Gamma^*$. Given that Γ^* is a convex set, I only need to show that Γ^* is a dense set of $\mathrm{int}(\Gamma \times [0, \infty)^2)$: For any $(w^1, w^0, t^1, t^0) \in \mathrm{int}(\Gamma \times [0, \infty)^2)$

49

and any $\epsilon > 0$, there exists (π, c) such that the Euclidean distance $\|(w^1, w^0, t^1, t^0) - (\mathbf{w}^1, \mathbf{w}^0, \mathbf{t}^1, \mathbf{t}^0)(\pi, c)\|$ is below ϵ. Pick any point $(w^1, w^0) \in \text{int}(\Gamma)$, there exists a bundle $(\tilde{w}^1, \tilde{w}^0) \in \text{int}(\Gamma)$ and a small number Δ such that $(w^1, w^0) = (1 - \Delta)(\tilde{w}^1, \tilde{w}^0) + \Delta(1, 1)$. The policy is as follows. With probability $1 - \Delta$, the agent implements a policy that is mapped to $(\tilde{w}^1, \tilde{w}^0)$. With probability Δ, the agent implements the policy that is mapped to $(1, 1)$. In the latter case, the agent allocates the unit resource to R all the time. Transfers only occur in the latter case. Here, I construct a transfer scheme such that the expected transfer is arbitrarily close to (t^1, t^0). Let p_t denote the posterior belief that $\omega = 1$. Given that all the resource is directed to R, p_t converges in probability to 1 conditional on state 1 and p_t converges in probability to 0 conditional on state 0. This implies that $\forall \tilde{\epsilon} > 0$, $\exists \tilde{t}$ such that for all $t \geq \tilde{t}$, I have $\Pr(|p_t - 1| > \tilde{\epsilon} \mid \omega = 1) < \tilde{\epsilon}$ and $\Pr(|p_t - 0| > \tilde{\epsilon} \mid \omega = 0) < \tilde{\epsilon}$. The transfer scheme is to make a transfer of size $t^1/(\Delta r e^{-r\tilde{t}})$ at time $\tilde{t}$ if $p_{\tilde{t}} > 1 - \tilde{\epsilon}$ and make a transfer of size $t^0/(\Delta r e^{-r\tilde{t}})$ if $p_{\tilde{t}} < \tilde{\epsilon}$. The expected transfer conditional on state 1 is

$$\Delta r e^{-r\tilde{t}} \left[\Pr(p_{\tilde{t}} > 1 - \tilde{\epsilon} \mid \omega = 1) \frac{t^1}{\Delta r e^{-r\tilde{t}}} + \Pr(p_{\tilde{t}} < \tilde{\epsilon} \mid \omega = 1) \frac{t^0}{\Delta r e^{-r\tilde{t}}} \right]$$

$$= \Pr(p_{\tilde{t}} > 1 - \tilde{\epsilon} \mid \omega = 1) t^1 + \Pr(p_{\tilde{t}} < \tilde{\epsilon} \mid \omega = 1) t^0.$$

Given that $1 - \tilde{\epsilon} < \Pr(p_{\tilde{t}} > 1 - \tilde{\epsilon} \mid \omega = 1) \leq 1$ and $0 \leq \Pr(p_{\tilde{t}} < \tilde{\epsilon} \mid \omega = 1) < \tilde{\epsilon}$, the expected transfer conditional on state 1 is in the interval $(t^1 - \tilde{\epsilon} t^1, t^1 + \tilde{\epsilon} t^0)$. Similarly, the expected transfer conditional on state 0 is in the interval $(t^0 - \tilde{\epsilon} t^0, t^0 + \tilde{\epsilon} t^1)$. As $\tilde{\epsilon}$ approaches zero, the constructed transfer scheme is arbitrarily close to (t^1, t^0). ∎

Lemma 10 says that any $(w^1, w^0, t^1, t^0) \in \Gamma \times [0, \infty) \times [0, \infty)$ is virtually implementable: for all $\epsilon > 0$, there exist a (π, c) such that $(\mathbf{w}^1, \mathbf{w}^0, \mathbf{t}^1, \mathbf{t}^0)(\pi, c)$ is ϵ-close to (w^1, w^0, t^1, t^0). To proceed, I treat the set $\Gamma \times [0, \infty) \times [0, \infty)$, the closure of Γ^*, as the contract space.

Based on Lemma 9, I can write players' payoffs as functions of (w^1, w^0, t^1, t^0). To simplify exposition, I assume that $s_\alpha - \lambda^0 h_\alpha = s_\rho - \lambda^0 h_\rho$. The method illustrated below can be easily adjusted to solve for the optimal contract when $s_\alpha - \lambda^0 h_\alpha \neq s_\rho - \lambda^0 h_\rho$. Without loss of generality, I further assume that $s_\alpha - \lambda^0 h_\alpha = 1$. The principal's and the agent's payoffs given (w^1, w^0, t^1, t^0) and type θ are then respectively

$$\frac{\theta}{1 + \theta} \left(\eta_\rho w^1 - t^1 \right) - \frac{1}{1 + \theta} \left(w^0 + t^0 \right) \text{ and } \frac{\theta}{1 + \theta} \left(\eta_\alpha w^1 + t^1 \right) - \frac{1}{1 + \theta} \left(w^0 - t^0 \right).$$

Based on Lemma 9 and 10, I reformulate the contract problem. The principal simply offers a direct mechanism $(w^1, w^0, t^1, t^0) : \Theta \to \Gamma \times [0, \infty) \times [0, \infty)$, called a contract,

such that

$$\max_{w^1,w^0,t^1,t^0} \int_\Theta \left(\frac{\theta}{1+\theta} \left(\eta_\rho w^1(\theta) - t^1(\theta) \right) - \frac{1}{1+\theta} \left(w^0(\theta) + t^0(\theta) \right) \right) \mathrm{d}F(\theta),$$

$$\text{s.t.} \quad \theta \left(\eta_\alpha w^1(\theta) + t^1(\theta) \right) - w^0(\theta) + t^0(\theta) \geq$$
$$\theta \left(\eta_\alpha w^1(\theta') + t^1(\theta') \right) - w^0(\theta') + t^0(\theta'), \quad \forall \theta, \theta' \in \Theta.$$

Given a direct mechanism $(w^1(\theta), w^0(\theta), t^1(\theta), t^0(\theta))$, let $U_\alpha(\theta)$ denote the payoff that the agent of type θ gets by maximizing over his report, *i.e.*, $U_\alpha(\theta) = \max_{\theta' \in \Theta} \{ \theta \left(\eta_\alpha w^1(\theta') + t^1(\theta') \right) - w^0(\theta') + t^0(\theta') \}$. As the optimal mechanism is truthful, $U_\alpha(\theta)$ equals $\theta(\eta_\alpha w^1(\theta) + t^1(\theta)) - w^0(\theta) + t^0(\theta)$ and the envelope condition implies that $U'_\alpha(\theta) = \eta_\alpha w^1(\theta) + t^1(\theta)$. Incentive compatibility of (w^1, w^0, t^1, t^0) requires that, for all θ

(B7) $$\left(t^0 \right)' (\theta) = -\theta \left(\eta_\alpha \left(w^1 \right)' (\theta) + \left(t^1 \right)' (\theta) \right) + \left(w^0 \right)' (\theta),$$

whenever differentiable, or in integral form,

$$t^0(\theta) = U_\alpha(\bar{\theta}) - \int_\theta^{\bar{\theta}} \left(\eta_\alpha w^1(s) + t^1(s) \right) \mathrm{d}s - \theta \left(\eta_\alpha w^1(\theta) + t^1(\theta) \right) + w^0(\theta).$$

Incentive compatibility also requires $\eta_\alpha w^1 + t^1$ to be a nondecreasing function of θ. Thus, (B7) and the monotonicity of $\eta_\alpha w^1 + t^1$ are necessary for incentive compatibility. As is standard, these two conditions are also sufficient.

The principal's payoff for a fixed θ is denoted $U_\rho(\theta)$

$$U_\rho(\theta) = \frac{\theta}{1+\theta} \left(\eta_\rho w^1(\theta) - t^1(\theta) \right) - \frac{1}{1+\theta} \left(w^0(\theta) + t^0(\theta) \right).$$

The principal's problem is to maximize $\int_\Theta U_\rho(\theta) \mathrm{d}F$ subject to (i) (B7) and the monotonicity of $\eta_\alpha w^1 + t^1$; (ii) the feasibility constraint $(w^1(\theta), w^0(\theta)) \in \Gamma, \forall \theta \in \Theta$; and (iii) the limited liability constraint (LL constraint, hereafter) $t^1(\theta), t^0(\theta) \geq 0, \forall \theta \in \Theta$. I denote this problem by $\mathcal{P}$. Substituting $t^0(\theta)$ into the objective, I rewrite $\int_\Theta U_\rho(\theta) \mathrm{d}F$ as

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[\frac{(\eta_\alpha + \eta_\rho)\theta w^1(\theta)}{1+\theta} - \frac{2w^0(\theta)}{1+\theta} + \frac{\left(\eta_\alpha w^1(\theta) + t^1(\theta) \right) H(\theta)}{f(\theta)} \right] f(\theta) \mathrm{d}\theta - U_\alpha(\bar{\theta}) H(\bar{\theta}),$$

where $h(\theta) = \frac{f(\theta)}{1+\theta}$ and $H(\theta) = \int_{\underline{\theta}}^\theta h(s) \mathrm{d}s.$

I then define a relaxed problem $\mathcal{P}'$ which differs from $\mathcal{P}$ in two aspects: (i) the monotonicity of $\eta_\alpha w^1 + t^1$ is dropped; and (ii) the feasibility constraint is replaced with $w^0(\theta) \geq \beta^{\mathrm{se}}(w^1(\theta)), \forall \theta \in \Theta$, where $\beta^{\mathrm{se}}(\cdot)$ characterizes the southeast boundary of Γ. If the solution to $\mathcal{P}'$ is monotone and satisfies the feasibility constraint, it is

51

also the solution to $\mathcal{P}$. The problem $\mathcal{P}'$ can be transformed into a control problem with the state $\boldsymbol{s} = (w^1, w^0, t^1, t^0)$ and the control $\boldsymbol{y} = (y^1, y^0, y_t^1)$. The associated costate is $\boldsymbol{\gamma} = (\gamma^1, \gamma^0, \gamma_t^1, \gamma_t^0)$. The law of motion is

(B8) $\qquad (w^1)' = y^1, \, (w^0)' = y^0, \, (t^1)' = y_t^1, \, (t^0)' = y^0 - \theta \left(\eta_\alpha y^1 + y_t^1 \right).$

For problem $\mathcal{P}'$, I define a Lagrangian

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{s}, \boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \theta) = & \left[\frac{(\eta_\alpha + \eta_\rho)\theta w^1(\theta)}{1+\theta} - \frac{2w^0(\theta)}{1+\theta} + \frac{\left(\eta_\alpha w^1(\theta) + t^1(\theta) \right) H(\theta)}{f(\theta)} \right] f(\theta) \\
& + \gamma^1(\theta) y^1(\theta) + \gamma^0(\theta) y^0(\theta) + \gamma_t^0(\theta) \left[y^0(\theta) - \theta \left(\eta_\alpha y^1(\theta) + y_t^1(\theta) \right) \right] \\
& + \gamma_t^1(\theta) y_t^1(\theta) + \mu_t^1(\theta) t^1(\theta) + \mu_t^0 t^0 + \mu(\theta) \left[w^0(\theta) - \beta^{\mathrm{se}}(w^1(\theta)) \right],
\end{aligned}
$$

where $\boldsymbol{\mu} = (\mu_t^1, \mu_t^0, \mu)$ are multipliers associated with the LL and feasibility constraints. From now on, the dependence of $(\boldsymbol{s}, \boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\mu}, f, h, H)$ on θ is omitted when no confusion arises. Given any θ, the control maximizes the Lagrangian. The first-order conditions are,

(B9) $\qquad \dfrac{\partial \mathcal{L}}{\partial y^1} = \gamma^1 - \eta_\alpha \theta \gamma_t^0 = 0, \quad \dfrac{\partial \mathcal{L}}{\partial y^0} = \gamma^0 + \gamma_t^0 = 0, \quad \dfrac{\partial \mathcal{L}}{\partial y_t^1} = \gamma_t^1 - \theta \gamma_t^0 = 0.$

The costate variables are continuous and have piecewise-continuous derivatives,

(B10)
$$
\begin{aligned}
\dot{\gamma}^1 &= -\frac{\partial \mathcal{L}}{\partial w^1} = -\left[\frac{\theta}{1+\theta}(\eta_\alpha + \eta_\rho) + \frac{\eta_\alpha H}{f} \right] f + \mu \, (\beta^{\mathrm{se}})' \, (w^1), \\
\dot{\gamma}^0 &= -\frac{\partial \mathcal{L}}{\partial w^0} = \frac{2}{1+\theta} f - \mu, \quad \dot{\gamma}_t^1 = -\frac{\partial \mathcal{L}}{\partial t^1} = -H - \mu_t^1, \quad \dot{\gamma}_t^0 = -\frac{\partial \mathcal{L}}{\partial t^0} = -\mu_t^0.
\end{aligned}
$$

This is a problem with free endpoint and a scrap value function $\Phi(\bar{\theta}) = -U_\alpha(\bar{\theta}) H(\bar{\theta})$. Therefore, the costate variables must satisfy the following boundary conditions,

(B11)
$$
\begin{aligned}
(\gamma^1(\underline{\theta}), \gamma^0(\underline{\theta}), \gamma_t^1(\underline{\theta}), \gamma_t^0(\underline{\theta})) &= (0, 0, 0, 0), \\
(\gamma^1(\bar{\theta}), \gamma^0(\bar{\theta}), \gamma_t^1(\bar{\theta}), \gamma_t^0(\bar{\theta})) &= \left(\frac{\partial \Phi}{\partial w^1}, \frac{\partial \Phi}{\partial w^0}, \frac{\partial \Phi}{\partial t^1}, \frac{\partial \Phi}{\partial t^0} \right) \Bigg|_{\theta = \bar{\theta}} \\
&= \left(-\eta_\alpha \bar{\theta} H(\bar{\theta}), H(\bar{\theta}), -\bar{\theta} H(\bar{\theta}), -H(\bar{\theta}) \right).
\end{aligned}
$$

Also, the following slackness conditions must be satisfied,

(B12)
$$
\begin{aligned}
&\mu_t^1 \geq 0, t^1 \geq 0, \mu_t^1 t^1 = 0; \quad \mu_t^0 \geq 0, t^0 \geq 0, \mu_t^0 t^0 = 0; \\
&\mu \geq 0, w^0 - \beta^{\mathrm{se}}(w^1) \geq 0, \mu \left(w^0 - \beta^{\mathrm{se}}(w^1) \right) = 0.
\end{aligned}
$$

LEMMA 11 (Necessity and sufficiency):
  Let $\boldsymbol{y}^*$ be the optimal control and $\boldsymbol{s}^*$ the corresponding trajectory. Then there

exist costate variables $\boldsymbol{\gamma}^*$ and multipliers $\boldsymbol{\mu}^*$ such that (B8)–(B12) *are satisfied. Conversely,* (B8)–(B12) *are also sufficient since the Lagrangian is concave in* $(\boldsymbol{s}, \boldsymbol{y})$.

Based on the sufficiency part of Lemma 11, I only need to construct $(\boldsymbol{s}, \boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\mu})$ such that the conditions (B8)–(B12) are satisfied. In what follows, I first describe the optimal contract and then prove its optimality.

<center>OPTIMAL CONTRACT: DESCRIPTION</center>

I identify a bundle (w^1, w^0) on the southeast boundary of Γ with the derivative $(\beta^{\mathrm{se}})'(w^1)$ at that point. If the optimal contract only involves bundles on the southeast boundary, the trajectory $(\beta^{\mathrm{se}})'(w^1(\theta))$ uniquely determines the trajectory $(w^1(\theta), w^0(\theta))$ and vice versa. Figure B1 illustrates three important trajectories of $(\beta^{\mathrm{se}})'(w^1(\theta))$ which determine the optimal contract under certain regularity conditions. The x axis is the agent's type. The y axis indicates the slope of the tangent line at a certain bundle on the southeast boundary. The thick-dashed line (labeled T_2) corresponds to the slope (or the bundle) preferred by the agent for any given θ. The thin-dashed line (labeled T_3) shows the bundle preferred by the principal if she believes that the agent's type is above θ. The thin-solid line (labeled T_1) is the bundle given by the following equation

(B13) $$(\beta^{\mathrm{se}})'(w^1(\theta)) = \frac{(\eta_\alpha + \eta_\rho)\theta}{2 - \frac{(1+\theta)H}{\theta f} + \frac{1+\theta}{\theta^2 f} \int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s}.$$

Loosely speaking, it is the bundle that the principal would offer if the LL constraint were not bound. Besides these three trajectories, the dotted line shows the bundle preferred by the principal for any given θ. Let θ^* denote the type at which T_1 and T_2 intersects and θ_p the type at which T_2 and T_3 intersects. Equations (B14) and (B15) gives the formal definition of θ^* and θ_p. It is easy to verify that $\theta^* > \underline{\theta}$ and $\theta_p < \overline{\theta}$. Moreover, θ^* increases and θ_p decreases in η_α/η_ρ.

(B14) $$\theta^* := \sup\left\{\hat{\theta} \in \Theta : \frac{(\eta_\alpha + \eta_\rho)\theta}{2 - \frac{(1+\theta)H}{\theta f} + \frac{1+\theta}{\theta^2 f} \int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s} < \eta_\alpha \theta, \ \forall \theta \leq \hat{\theta}\right\},$$

(B15) $$\theta_p := \inf\left\{\hat{\theta} \in \Theta : \frac{\eta_\rho \int_{\theta}^{\overline{\theta}} sh(s)\mathrm{d}s}{H(\overline{\theta}) - H(\theta)} \leq \eta_\alpha \theta, \ \forall \theta \geq \hat{\theta}\right\}.$$

When the bias η_α/η_ρ is small, $\theta^* < \theta_p$. The optimal contract (the thick-solid line) consists of three separate segments, *i.e.*, $[\underline{\theta}, \theta^*]$, $[\theta^*, \theta_p]$, and $[\theta_p, \overline{\theta}]$. (See Figure B1.) When the agent's type is below θ^*, the equilibrium allocation is given by (B13), which lies between that optimal for the principal $((\beta^{\mathrm{se}})'(w^1(\theta)) = \eta_\rho\theta)$ and that optimal for the agent $((\beta^{\mathrm{se}})'(w^1(\theta)) = \eta_\alpha\theta)$. As θ increase, the contract bundle shifts toward the agent's preferred bundle, with a corresponding decrease in the transfer payments. When $\theta \in [\theta^*, \theta_p]$, the bundle that is preferred by the agent

<center>53</center>

is offered and no transfers are made. It is as if the agent is free to choose any experimentation policy. For types above θ_p, the agent always chooses the bundle preferred by type θ_p. There is, effectively, pooling over $[\theta_p, \overline{\theta}]$.

When the bias η_α/η_ρ is large, $\theta^* > \theta_p$. The optimal contract (the thick-solid line) consists of only two segments which are denoted $[\underline{\theta}, \tilde{\theta}_p]$ and $[\tilde{\theta}_p, \overline{\theta}]$. (See Figure B2.) When $\theta \in [\underline{\theta}, \tilde{\theta}_p]$, the equilibrium allocation is between the principal's preferred bundle and the agent's preferred one. The contract bundle shifts toward the agent's preferred one as the type increases with a corresponding decrease in the transfers. When $\theta \in [\tilde{\theta}_p, \overline{\theta}]$, all types are pooled. The pooling bundle specifies a lower level of experimentation than what the principal prefers given the pooling segment $[\tilde{\theta}_p, \overline{\theta}]$. There is no segment in which the agent implements his preferred bundle.
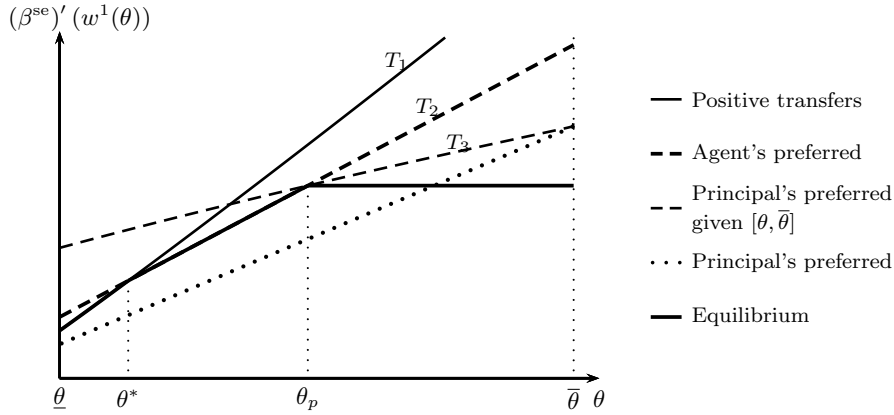

FIGURE B1. EQUILIBRIUM ALLOCATION: THREE SEGMENTS

Note: Parameters are $\eta_\alpha = 1, \eta_\rho = 4/5, \underline{\theta} = 1, \overline{\theta} = 3, f(\theta) = 1/2$.

One immediate observation is that the most pessimistic type's policy is socially optimal. The contract for type $\underline{\theta}$ is chosen such that $(\beta^{\text{se}})'(w^1(\underline{\theta})) = (\eta_\alpha + \eta_\rho)\underline{\theta}/2$. A key feature is that transfers only occur in state 1. The intuition can be seen by examining a simple example with three types $\theta_l < \theta_m < \theta_h$. To prevent the medium type from mimicking the high type, the principal compensates the medium type by promising him a positive transfer. This transfer promise makes the medium type's contract more attractive to the low type. To make the medium type's transfer less attractive to the low type, the principal concentrates all the payments in state 1 as the low type is less confident that the state is 1. Whenever the principal promises type θ a positive transfer, she makes type θ's contract more attractive to a lower type, say $\theta' < \theta$. As type θ' is less confident that the state is 1 than type θ, type θ' does not value transfers in state 1 as much as type θ is. Therefore, the most efficient way to make transfers is to condition on state being 1.
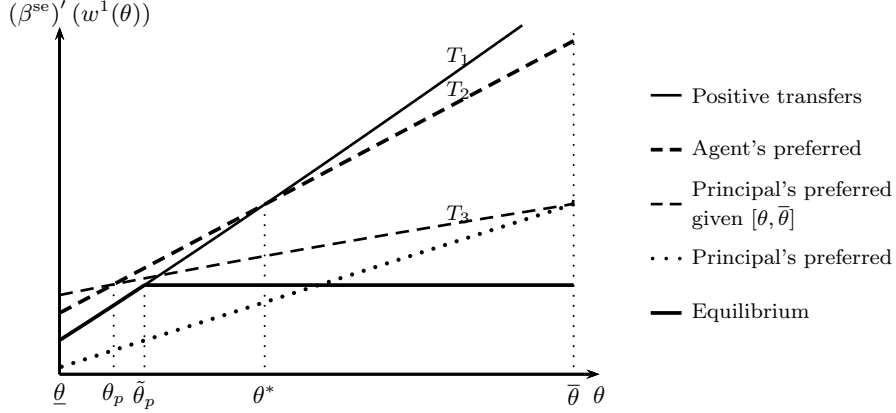
54

FIGURE B2. EQUILIBRIUM ALLOCATION: TWO SEGMENTS

Note: Parameters are $\eta_\alpha = 1, \eta_\rho = 3/5, \underline{\theta} = 1, \overline{\theta} = 3, f(\theta) = 1/2$.

OPTIMAL CONTRACT: PROOF

I start with the case when the bias is small and the contract has three segments. The proof is constructive. I first determine the trajectory of the costate γ^0, which pins down $\gamma^1, \gamma_t^1, \gamma_t^0$ according to (B9). Then I determine the trajectories of $\mu, \mu_t^1, \mu_t^0, w^1$ based on (B10). The trajectories of w^0, t^1, t^0 then follow.

On the interval $[\underline{\theta}, \theta^*]$, $t^1 > 0$ and $t^0 = 0$, so the LL constraint $t^1 \geq 0$ does not bind. Therefore, I have $\mu_t^1 = 0$ and $\dot{\gamma}_t^1 = -H$. Combined with the boundary condition, this implies that $\gamma_t^1 = -\int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s$. From (B9), we know that $\gamma^0 = \int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s/\theta$ and $\gamma^1 = -\eta_\alpha \int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s$. Substituting $\dot{\gamma}^1$ and $\dot{\gamma}^0$ into (B10), I have

$$\mu = \frac{2f}{1+\theta} - \frac{H(\theta)}{\theta} + \frac{\int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s}{\theta^2} \quad \text{and} \quad (\beta^{\mathrm{se}})'(w^1(\theta)) = \frac{(\eta_\alpha + \eta_\rho)\theta}{2 - \frac{(1+\theta)H}{\theta f} + \frac{1+\theta}{\theta^2 f}\int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s}.$$

Since $\mu_t^0 = -\dot{\gamma}_t^0 = \dot{\gamma}^0$, I have $\mu_t^0 = -\int_{\underline{\theta}}^{\theta} H(s)\mathrm{d}s/(\theta^2) + H(\theta)/\theta$, which is always positive.

On the interval $[\theta^*, \theta_p]$, the type θ is assigned his most preferred bundle. Transfers t^1 and t^0 both equal zero. Therefore, I have $(\beta^{\mathrm{se}})'(w^1(\theta)) = \eta_\alpha \theta$. From (B9), we know that $\gamma^0 = -\dot{\gamma}^1/\eta_\alpha - \theta\dot{\gamma}^0$. Substituting $\dot{\gamma}^1$, $\dot{\gamma}^0$ and $(\beta^{\mathrm{se}})'(w^1(\theta)) = \eta_\alpha\theta$, I have

(B16)
$$\gamma^0 = H - \frac{\theta}{1+\theta}\frac{\eta_\alpha - \eta_\rho}{\eta_\alpha}f.$$

55

Combining (B10) and (B16), I have

$$\mu = \frac{f}{1+\theta}\left(1 + \frac{\eta_\alpha - \eta_\rho}{\eta_\alpha}\frac{1}{1+\theta} + \frac{\eta_\alpha - \eta_\rho}{\eta_\alpha}\frac{f'}{f}\theta\right)$$

$$\mu_t^0 = \frac{f}{1+\theta}\left(1 - \frac{\eta_\alpha - \eta_\rho}{\eta_\alpha}\frac{1}{1+\theta} - \frac{\eta_\alpha - \eta_\rho}{\eta_\alpha}\frac{f'}{f}\theta\right)$$

$$\mu_t^1 = \frac{\theta}{1+\theta}f\left(1 - \frac{\eta_\alpha - \eta_\rho}{\eta_\alpha}\frac{2+\theta}{1+\theta} - \frac{\eta_\alpha - \eta_\rho}{\eta_\alpha}\frac{f'}{f}\theta\right).$$

The multipliers μ, μ_t^0 and μ_t^1 have to be weakly positive, which requires that

(B17) $$\frac{\eta_\alpha}{\eta_\alpha - \eta_\rho} \geq -\frac{1}{1+\theta} - \frac{f'}{f}\theta, \quad \forall \theta \in [\theta^*, \theta_p]$$

(B18) $$\frac{\eta_\alpha}{\eta_\alpha - \eta_\rho} \geq \frac{2+\theta}{1+\theta} + \frac{f'}{f}\theta, \quad \forall \theta \in [\theta^*, \theta_p].$$

Note that Assumption (B17) is the same as the main assumption in the delegation case.

On the interval $[\theta_p, \bar\theta]$, all types choose the same bundle, the one preferred by type θ_p. Transfers t^1 and t^0 both equal zero. The threshold of the pooling segment θ_p satisfies the following condition,

$$(\beta^{\text{se}})'\left(w^1(\theta_p)\right) = \theta_p\eta_\alpha = \frac{\eta_\rho \int_{\theta_p}^{\bar\theta} \theta h(\theta)\mathrm{d}\theta}{H(\bar\theta) - H(\theta_p)}.$$

I first check that the boundary condition $\gamma^1(\bar\theta) = -\eta_\alpha H(\bar\theta)\bar\theta$ is satisfied. Over the interval $[\theta_p, \bar\theta]$, I have

$$\dot\gamma^1 = -\left[\frac{\theta}{1+\theta}(\eta_\alpha + \eta_\rho)f + \eta_\alpha H\right] + \mu\left(\beta^{\text{se}}\right)'\left(w^1(\theta_p)\right).$$

Given the definition of θ_p, it is easy to verify that

$$(\beta^{\text{se}})'\left(w^1(\theta_p)\right)\int_{\theta_p}^{\bar\theta} \mu\mathrm{d}\theta = (\beta^{\text{se}})'\left(w^1(\theta_p)\right)\int_{\theta_p}^{\bar\theta}\left(\frac{2f}{1+\theta} - \dot\gamma^0\right)\mathrm{d}\theta$$

$$= \eta_\rho\int_{\theta_p}^{\bar\theta}\theta h(\theta)\mathrm{d}\theta - (\eta_\alpha - \eta_\rho)\frac{\theta_p^2}{1+\theta_p}f(\theta_p).$$

56

Therefore, I have

$$\gamma^1(\bar{\theta}) - \gamma^1(\theta_p) = -\eta_\alpha \int_{\theta_p}^{\bar{\theta}} \theta h(\theta) d\theta - \eta_\alpha \int_{\theta_p}^{\bar{\theta}} H(\theta) d\theta - (\eta_\alpha - \eta_\rho) \frac{\theta_p^2}{1 + \theta_p} f(\theta_p)$$

$$= -H(\bar{\theta}) \bar{\theta} \eta_\alpha - \gamma^1(\theta_p).$$

Therefore, the boundary condition $\gamma^1(\bar{\theta}) = -\eta_\alpha H(\bar{\theta}) \bar{\theta}$ is satisfied. The slackness condition $\mu \geq 0$ and $\mu_t^0 \geq 0$ requires that $0 \leq \dot{\gamma}^0 \leq 2f/(1 + \theta)$. This is equivalent to the condition that $0 \leq \gamma^0(\bar{\theta}) - \gamma^0(\theta_p) \leq 2\left(H(\bar{\theta}) - H(\theta_p)\right)$, which is satisfied iff

(B19)
$$\frac{\eta_\alpha}{\eta_\alpha - \eta_\rho} \geq \frac{\frac{\theta_p}{1 + \theta_p} f(\theta_p)}{H(\bar{\theta}) - H(\theta_p)}.$$

The slackness condition $\mu_t^1 \geq 0$ requires that $\dot{\gamma}_t^1 \leq -H$. This is equivalent to the condition that

$$\gamma_t^1(\bar{\theta}) - \gamma_t^1(\theta_p) = -\bar{\theta} H(\bar{\theta}) + \theta_p H(\theta_p) - \frac{\theta_p^2}{1 + \theta_p} \frac{\eta_\alpha - \eta_\rho}{\eta_\alpha} f(\theta_p) \leq -\int_{\theta_p}^{\bar{\theta}} H(s) ds,$$

which is always satisfied.

To sum, if Assumptions (B17), (B18) and (B19) hold, the constructed trajectory solves $\mathcal{P}'$. If the trajectory w^1 defined by (B13) is weakly increasing over $[\underline{\theta}, \theta^*]$, the monotonicity of $\eta_\alpha w^1 + t^1$ is satisfied.[27] Therefore, the constructed trajectory solves $\mathcal{P}$ as well.

When the bias is large and the contract has two segments, the proof is similar to the previous case. So, I mainly explain how to pin down the threshold $\tilde{\theta}_p$. When $\theta \in [\underline{\theta}, \tilde{\theta}_p]$, the LL constraint $t^1 \geq 0$ does not bind. The costate is derived in the same way as in the previous case when $\theta \in [\underline{\theta}, \theta^*]$. This implies that $\gamma^1(\tilde{\theta}_p) = -\eta_\alpha \int_{\underline{\theta}}^{\tilde{\theta}_p} H(s) ds$. On the other hand, the threshold $\tilde{\theta}_p$ is chosen so that the boundary condition $\gamma^1(\bar{\theta}) = -\eta_\alpha H(\bar{\theta}) \bar{\theta}$ is satisfied. This means that $\int_{\tilde{\theta}_p}^{\bar{\theta}} \dot{\gamma}^1 d\theta = -\eta_\alpha H(\bar{\theta}) \bar{\theta} + \eta_\alpha \int_{\underline{\theta}}^{\tilde{\theta}_p} H(s) ds$. Substituting $\dot{\gamma}^1$ and simplifying, I obtain that $(\beta^{\text{se}})' (w^1(\tilde{\theta}_p))$ must satisfy the following condition

(B20)
$$(\beta^{\text{se}})' (w^1(\tilde{\theta}_p)) = \frac{\eta_\rho \int_{\tilde{\theta}_p}^{\bar{\theta}} \theta h(\theta) d\theta - \eta_\alpha \tilde{\theta}_p H(\tilde{\theta}_p) + \eta_\alpha \int_{\underline{\theta}}^{\tilde{\theta}_p} H(s) ds}{H(\bar{\theta}) - 2H(\tilde{\theta}_p) + \frac{\int_{\underline{\theta}}^{\tilde{\theta}_p} H(s) ds}{\tilde{\theta}_p}}$$

At the same time, $\tilde{\theta}_p$ must also satisfy (B13). Equation (B13) and (B20) determines

---

[27] Given that t^0 is constantly zero, I have $\eta_\alpha \left(w^1\right)' (\theta) + \left(t^1\right)' (\theta) = \left(w^0\right)' (\theta)/\theta$. Therefore, the monotonicity of w^0 implies the monotonicity of $\eta_\alpha w^1 + t^1$. Given that only boundary bundles (w^1, w^0) are assigned, the monotonicity of w^1 suffices.

the threshold $\tilde{\theta}_p$. Since $(\beta^{\mathrm{se}})'(w^1(\tilde{\theta}_p)) < \eta_\alpha \tilde{\theta}_p$, (B20) implies that

$$(\beta^{\mathrm{se}})'(w^1(\tilde{\theta}_p)) < \frac{\eta_\rho \int_{\tilde{\theta}_p}^{\overline{\theta}} \theta h(\theta)\mathrm{d}\theta}{H(\overline{\theta}) - H(\tilde{\theta}_p)}.$$

This shows that over the pooling region $[\tilde{\theta}_p, \overline{\theta}]$ the agent is asked to implement a bundle with less experimentation than what the principal prefers given that $\theta \in [\tilde{\theta}_p, \overline{\theta}]$.