I'll Tell You Tomorrow: Committing to Future Commitments

Andrew B. Choi*

First Draft: November 2022

This Draft: June 2024

Click here for latest version

Abstract

A principal wishes to promote an agent only if the state is good, and gradually receives private information about the state. The agent wants promotion but would rather leave than stay and fail promotion. The principal induces the agent to stay by committing today to tell the agent tomorrow about his chances of promotion the day after. The principal promotes the agent with some probability even after realizing early that the state is bad. The principal may commit not to lead the agent on. Our results apply to worker retention, relationship-specific investment, and forward guidance.

Keywords: higher-order commitment, rules versus discretion, dynamic mechanism design, dynamic information design

*Bocconi University, andybchoi@gmail.com. I am grateful to my advisor Philip Reny and my dissertation committee: Benjamin Brooks, Emir Kamenica, and Lars Stole for invaluable support and guidance. I also thank Yeon-Koo Che, Yong-Jae Choi, Francesc Dilmé, Alex Frankel, Takuma Habu, Marina Halac, Kevin He, Jinwoo Kim, Andreas Kleiner, Daniel Krähmer, Roger Myerson, Derek Neal, Salvatore Nunnari, Marco Ottaviani, Minseon Park, Doron Ravid, Dezsö Szalay, Christoph Schlom, Nancy Stokey, Tom Winberry, Kai Hao Yang, Yves Zenou, and various seminar participants for advice and suggestions.

1 Introduction

Consider a principal who will take action A or B two days from now. The principal could commit today to take action A. Alternatively, she could commit today that tomorrow, she will either commit to take action A the day after, or commit to take action B the day after; that is, she could commit to commit.

The U.S. Air Force provides a recent example. Because it faces stiff competition for its pilots from civilian airlines,[1] the Air Force has historically provided discretionary rewards – such as choice of assignment location or bonus pay – to pilots who stayed beyond their 10-year service requirement. However, this has not been very effective in retaining pilots because the Air Force did not tell a pilot until he reached the end of the 10-year requirement what, if any, rewards he would receive, whereas many pilots would decide a few years in advance that they would leave once they finished their requirement.[2] Realizing this, the Air Force announced in 2023 that, when a pilot has one to three years left in his service requirement, the Air Force will either commit to provide rewards or commit not to provide them. That is, the Air Force made a commitment about future commitments.[3]

Such higher-order commitments occur in many contexts. A firm may commit today to update a worker next year about his chance of promotion the year after. A regulator who wishes to provide forward guidance may commit to announce next month the amount of subsidy that she will provide to an industry next year. Alice may say to Bob on Wednesday, "I'm not sure if I can have dinner with you on Friday, but I'll tell you tomorrow." While commitments about future commitments are prevalent, little is known on why such commitments are made or how one should make them. In this paper, we argue that when a principal receives private information over time, committing to commit is an effective way for the principal to convince an agent to wait for the principal's decision. We characterize how a principal should optimally commit to commit, and we show that when it is difficult to convince the agent to wait, the principal should commit to sometimes do what the agent prefers regardless of what information the principal may receive in the future.

We consider a parsimonious principal-agent model with three periods and a binary

[1]In 2022, major U.S. airlines hired 25% – or 3,280 – of their new pilots from the military, contributing to the long-standing shortage of Air Force pilots. Training one fighter pilot can cost more than 10 million dollars (Mattock et al., 2019).

[2]See General David Allvin's testimony to Senate Committee on Armed Services, Subcommittee on Readiness and Management Support (`https://www.armed-services.senate.gov/hearings/to-receive-testimony-on-the-current-readiness-of-the-joint-force`). Relevant testimony starts at 01:06:19.

[3]See James M. Inhofe National Defense Authorization Act for Fiscal Year 2023, Section 604 (`https://www.congress.gov/bill/117th-congress/house-bill/7776/text`).

state. In periods 0 and 1, the agent decides whether to stay and continue interacting with the principal or to leave and take his outside option, which decreases every period. If the agent stays until period 2, the principal chooses whether to promote the agent. The agent values being promoted and does not care about the state, whereas the principal's payoff from promotion is positive in the good state and negative in the bad state. The state is initially unknown to both parties, and the principal privately updates her belief about the state over time. In period 1, she observes a signal that is correlated with the state; in period 2, she observes the state.

A mechanism takes as input the principal's report of her updated belief about the state in period 1 and period 2. In period 1, the mechanism outputs a recommendation to the agent, informing him about the likelihood of being promoted. In period 2, the mechanism decides whether the agent will be promoted. We interpret the mechanism as the principal's commitment to the agent, which constrains the principal's future communication and action but grants her some flexibility to respond to future information. Our goal is to find a mechanism that maximizes the principal's ex ante payoff, subject to the constraints that the agent stays in period 0 and obeys the mechanism's recommendation in period 1, and that the principal reports truthfully in periods 1 and 2.

Ideally, the principal would wait until period 2 and then promote the agent if and only if the state is good, but because waiting is costly for the agent, the principal must convince the agent to stay while the principal receives information. A simple way to do so is to commit to wait until period 2 and then promote the agent with a high enough probability even if the state is bad. However, we argue that the principal can convince more efficiently by committing to commit, that is, by committing in period 0 to the ways in which, in period 1, the principal can restrict her promotion decision and communicate this restriction to the agent.

To understand why, we first assume that the principal's information is contractible, which allows us to ignore the principal's incentive compatibility constraints. This describes environments in which the principal's information can be verified ex post, or environments with many heterogeneous agents. In this case, the principal's ex ante payoff is maximized by the following *contractible-optimal mechanism*. In period 1, if the posterior probability that the state is good conditional on the principal's signal is below a threshold, the mechanism informs the agent that he will fail promotion regardless of the realized state, and he leaves. If the conditional probability is above the threshold, the mechanism asks the agent to stay, promising to promote him in the good state and possibly also promising to promote him with positive probability in the bad state. Thus the contractible-optimal mechanism is parametrized by the period-1 threshold belief and the period-2 promotion probability in the bad state, which can vary independently of each other. We interpret this mechanism as a commitment about future commitments; the principal commits in period 0 that, in period 1, she will either commit not to promote the agent or commit to promote her with at least some probability. The agent stays in period 0 in anticipation of the update that the principal will provide in period 1. In period 1, if the principal says that he will not be promoted, he can then leave; otherwise, he learns that he is relatively likely to be promoted, and will stay. By sometimes committing in period 1 not to promote the agent, the mechanism reduces the agent's ex ante cost of staying in period 0.

Unfortunately, if the principal's information cannot be contracted on, the contractible-optimal mechanism is not incentive compatible for the principal. For example, suppose the mechanism never promotes the agent when the state turns out to be bad (which is the case when the agent's period-0 outside option is not too attractive). In period 1, no matter how pessimistic the principal is about the state, as long as there is the smallest possibility that the state may turn out to be good, the principal always prefers that the agent stays until period 2, as promotion only happens in the good state. Thus in period 1, even if the principal's belief that the state is good is below the threshold, she will prefer to report to the mechanism that her belief is above the threshold. Similar arguments can be used to show that the contractible-optimal mechanism is never incentive compatible except in a knife-edge case.

We thus study the optimal mechanism, which maximizes the principal's ex ante payoff while incentivizing the agent to stay and incentivizing the principal to report truthfully. We show that there always exists an optimal mechanism that can be written as a convex combination of a constant mechanism, which always promotes the agent regardless of the principal's beliefs, and three single-threshold mechanisms. Similarly to the contractible-optimal mechanism, a threshold mechanism is a commitment about commitments; it never promotes the agent if the principal reports that her posterior belief is below a threshold, and if the report is above the threshold, the agent is always promoted in the good state and promoted with a positive probability in the bad state. Unlike in the contractible-optimal mechanism, however, the threshold belief pins down the probability of promotion in the bad state through the principal's truth-telling incentives. In a threshold mechanism, the two parameters of the contractible-optimal mechanism are coupled into one.

The optimal mechanism always places a positive weight on at least one threshold mechanism. Because the principal's incentive compatibility constraints make it difficult for a threshold mechanism to provide a high ex ante payoff to the agent, when the value of the agent's initial outside option is high, the optimal mechanism must also place a positive weight on the constant mechanism that always promotes the agent, ignoring the principal's information. As a result, even when the principal knows for sure in period 1 that the state will be bad, the mechanism sometimes asks the agent to stay in period 1, and then promotes the agent with certainty in period 2. In the same optimal mechanism, if the principal has a more optimistic belief about the state in period 1, the mechanism may ask the agent to stay in period 1, but then sometimes refuse to promote him in period 2 if the bad state is realized. Therefore, conditional on having obeyed the recommendation to stay in period 1, the agent may be *less* likely to be promoted in period 2 when the principal in period 1 believed that the state was more likely to be good. Not only does the principal's past belief – which is payoff-irrelevant once the state is realized – affect the probability of promotion, but an optimistic belief makes promotion less likely.

This seemingly unnatural feature of the optimal mechanism might lead one to wonder how it could be implemented in real life. Here, the decomposition of the optimal mechanism provides an insight. A firm designing promotion rules can implement threshold mechanisms by committing to conduct a midterm and a final review of the worker's value to the firm. A mixture of threshold mechanisms and the mechanism that always promotes the agent can be implemented by reviewing the worker's value only a fraction of the time and otherwise promoting him by default. When a worker who faces such a randomized mechanism fails promotion in period 2, he may complain that the firm *led him on* by asking him to stay even though it knew he was unlikely to be promoted. We provide a sufficient condition for the worker's obedience to be robust to additional information. Under this condition, the worker's complaint is unwarranted because he would have chosen to stay even if he had known everything that the firm had known when he was recommended to stay.

The methodological contribution of this paper is to introduce a portable approach for studying mechanism design without transfers. By observing that the principal trades off the probability of promotion in the good state against the probability of promotion in the bad state, we show that the former pins down the latter via the envelope theorem. A caveat is that period-2 incentive compatibility, as well as the restriction that a probability cannot be greater than 1, imposes additional (infinite-dimensional) constraints on the probability of promotion in the bad state; we show that these constraints reduce to one additional linear inequality constraint. By Winkler (1988) and Bauer's maximum principle, the constraint adds an extra jump to the optimal allocation. Subsequent work has drawn upon the approach in this paper to analyze different economic problems.[4]

[4]For example, Dasgupta (2023) uses my approach to study the optimal design of knowledge-screening tests.

Related Literature We consider an important yet understudied question of how a principal should optimally commit to provide information about her future action. Our model combines dynamic mechanism design and dynamic information design, and our mechanism is an instance of the communication mechanism for multistage games introduced by Myerson (1986), whose revelation principle we appeal to. As in dynamic mechanism design, such as sequential screening (Baron and Besanko, 1984; Courty and Hao, 2000; Krähmer and Strausz, 2015) and rules versus discretion (Kydland and Prescott, 1977; Barro and Gordon, 1983; Athey et al., 2005; Halac and Yared, 2014, 2022), our mechanism makes an allocation decision (promotion of the agent) based on private information which is elicited from a player (the principal) over time.[5] In dynamic mechanism design, the principal typically does not need to communicate to the agent, since the agent does not take different actions in equilibrium. For example, in standard sequential screening models, the buyer may be allowed to walk away from the monopolist at some point during the game. However, because the buyer's outside option stays constant throughout his interaction with the monopolist, it is without loss to impose interim or ex post participation constraints and consider mechanisms in which the buyer always stays until the end of the game. Thus there is no reason for the monopolist to provide any information to the buyer. In contrast, in our model, the agent sometimes leaves in period 1, and the principal commits in period 0 to inform the agent in period 1 about whether he should stay or leave. This promise of future communication plays a crucial role in incentivizing the agent to stay in period 0.

Following the growth of the literature on information design (Kamenica and Gentzkow, 2011; Bergemann and Morris, 2013), there have been a series of papers that study how a principal commits to provide over time information about the state of the world, chosen exogenously by nature (Ely, 2017; Renault et al., 2017; Ely and Szydlowski, 2020; Orlov et al., 2020; Smolin, 2021; Bizzotto et al., 2021; Ball, 2022). In contrast, the principal in our model commits to provide information over time about her own future decision. One might think of our principal as solving an information design problem where the state of the world that the agent cares about is chosen by the principal.[6]

To solve for the optimal mechanism, we apply Proposition 2.1. of Winkler (1988). In our context, the proposition characterizes the extreme points of the feasible set of direct mechanisms. To apply Winkler's result, one must be able to write the problem as a linear

[5]In particular, rules versus discretion studies how a principal who expects to receive private information in the future optimally restricts her future decision. For example, Athey et al. (2005) considers an infinite repetition of two-period interactions between a principal and a continuum of agents in the context of monetary policy and shows that the optimal mechanism is a static upper bound on policy.

[6]A literature on the design of feedback in dynamic contests (Lizzeri et al., 2005; Yildirim, 2005; Ely et al., 2023) studies how providing interim feedback about performance can affect agents' choice of effort.

program with a finite number of moment constraints.[7] Although our problem initially involves infinitely many constraints, we show how these can be reduced to a finite number of constraints. Our model does not have transfers, but by observing that the principal can trade off the probability of promotion in the good state against the probability of promotion in the bad state, we are able to appeal to the envelope theorem approach used in standard screening problems (Riley and Zeckhauser, 1983; Myerson, 1981; Mussa and Rosen, 1978). Our principal reports her private information to the mechanism, as happens under an inscrutable mechanism in the informed principal problem (Myerson, 1983). The difference is that our mechanism is chosen before the principal receives private information with the goal of maximizing the principal's ex ante expected payoff.

The interpretation of our results speaks to the literature on worker retention. A firm benefits from retaining its workers because the workers possess, and choose how much to invest in, firm-specific human capital (Oi, 1962; Becker, 2009; Mortensen, 1978; Hashimoto and Yu, 1980; Hashimoto, 1981). We complement this literature by asking how a firm can optimally retain its worker via the prospect of promotion. Our results are also related to the literature on investment under uncertainty (Bernanke, 1983; Dixit et al., 1994). It has been argued that uncertainty of government policy can hinder firms' investment (Rodrik, 1991; Gulen and Ion, 2016). We show how a regulator can optimally incentivize investment by committing to reduce policy uncertainty over time.

2 Model

A principal (she) and an agent (he) interact over three periods. There is a state $\theta \in \{-1, 1\}$. The players share a common prior belief that the state is good ($\theta = 1$) with probability $\mu_0 \in (0, 1)$ and bad ($\theta = -1$) with probability $1 - \mu_0$.

Period 0 (Ex Ante) The agent chooses whether to participate in the interaction. If he chooses not to participate, the game ends, the principal receives a payoff of 0, and the agent receives his ex ante outside option that gives him a payoff of $c_0 > 0$. If the agent participates, the game proceeds to period 1.

Period 1 (Interim) First, the principal privately updates her belief that the state is good to $\mu \in [0, 1]$. We assume that μ is drawn according to a distribution $F \in \Delta[0, 1]$ that has a density f and satisfies $\mathbb{E}_F[\mu] = \mu_0$. We assume $f > 0$ and $\lim_{\mu \to 1} f(\mu) > 0$.

[7] For example, Winkler's result does not apply to the setting of Kleiner et al. (2021) because there are uncountably many majorization constraints.

The belief distribution F is commonly known to both the principal and the agent in period 0, but only the principal observes the realized μ.[8]

The principal then sends to the agent an arbitrary message, which may contain information about the principal's updated belief. Having observed the principal's message, the agent chooses whether to stay or leave. Let $A = \{\text{stay}, \text{leave}\}$ be the action space of the agent in period 1. If the agent leaves, the game ends, the principal receives 0, and the agent receives his interim outside option $c_1 \in (0, c_0)$. If the agent stays, the game proceeds to period 2.

Period 2 (Ex post) First, the principal privately observes the state θ. Then, the principal decides whether to promote the agent. If the agent is promoted, the agent receives $b > c_0$ and the principal receives θ. If the agent is not promoted, both players receive 0. We assume that the agent's ex ante outside option is high enough that incentivizing the agent is nontrivial for the principal, i.e. $c_0 > b\mu_0$.[9]

Mechanism We appeal to the revelation principal (Myerson, 1986) and restrict attention to direct mechanisms, which is a pair of functions $\sigma = (\sigma_1, \sigma_2)$. In period 1, the principal reports her belief to the mechanism. Given a report $\hat{\mu}$, with probability $\sigma_1(\hat{\mu})$, the mechanism asks the agent to stay. With probability $1 - \sigma_1(\hat{\mu})$, the agent is asked to leave. In period 2, the principal reports the state to the mechanism. The mechanism promotes the agent with probability $\sigma_2(\hat{\mu}, \tilde{a}, \hat{\theta})$ if the period-1 report was $\hat{\mu}$, the period-1 recommendation was $\tilde{a}$, the agent stayed in period 1, and the period-2 report is $\hat{\theta}$.[10]

Optimal Mechanism The principal's expected payoff from a mechanism σ is

$$\int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu, \text{stay}, 1) - (1 - \mu)\sigma_2(\mu, \text{stay}, -1))\, \mathrm{d}F(\mu)\,. \tag{1}$$

Our goal is to find a mechanism σ that maximizes the principal's expected payoff subject to the constraints that the agent participates and obeys recommendations and that the principal reports truthfully to the mechanism. Notice that, on path, the game never

[8] The distribution F can be generated by a signal $\pi : \{-1, 1\} \to \Delta[0, 1]$ defined by

$$\begin{aligned} f(s) &= \mu_0\pi(s|1) + (1 - \mu_0)\pi(s| - 1), &\quad \forall s \in [0, 1] \\ (1 - s)\mu_0\pi(s|1) &= s(1 - \mu_0)\pi(s| - 1) &\quad \forall s \in [0, 1]. \end{aligned}$$

[9] Otherwise, the agent participates in period 0 and stays in period 1 even if the principal does not send any meaningful messages in period 1 and promotes the agent in period 2 if and only if $\theta = 1$.

[10] Formally, a mechanism is a pair of Borel-measurable functions $\sigma_1 : [0, 1] \to [0, 1]$ and $\sigma_2 : [0, 1] \times A \times \{-1, 1\} \to [0, 1]$.

proceeds to period 2 if the agent is asked to leave in period 1. Moreover, if the mechanism never promotes the agent in period 2 whenever the agent disobeyed the recommendation to leave in period 1, the agent will always obey the recommendation to leave in period 1. Thus it is without loss of generality to set $\sigma_2(\mu, \text{leave}, \theta) = 0$ for all μ and θ and trivially satisfy the agent's obedience constraint after being asked to leave in period 1. To simplify notation, we write $\sigma_2(\mu, \theta) \coloneqq \sigma_2(\mu, \text{stay}, \theta)$.

The agent's obedience constraint after being recommended to stay in period 1 is

$$c_1 \le b \frac{\int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu, 1) + (1-\mu)\sigma_2(\mu, -1))\mathrm{d}F(\mu)}{\int_0^1 \sigma_1(\mu)\mathrm{d}F(\mu)}, \tag{2}$$

where the fraction on the right-hand side is the expected probability of being promoted conditional on obeying the recommendation to stay. The agent's ex ante individual rationality constraint is

$$c_0 \le b \int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu, 1) + (1-\mu)\sigma_2(\mu, -1))\, \mathrm{d}F(\mu) + c_1 \int_0^1 (1 - \sigma_1(\mu))\, \mathrm{d}F(\mu)\,, \tag{A-IR}$$

where the first integral on the right-hand side is the ex ante probability of being promoted, and the second integral is the ex ante probability of leaving at the interim stage. It is easy to see that obedience after being asked to stay is implied by individual rationality. Intuitively, in period 0, the agent knows he will obey if recommended to leave. If he is going to disobey when asked to stay, then he will always receive a payoff of c_1 from participating in the mechanism. However, by not participating in the first place, he receives $c_0 > c_1$. We can thus ignore the agent's obedience constraints.

In period $t = 2$, the principal observes θ and reports $\hat{\theta}$. Given that her period-1 report was μ, her expected payoff is $\theta\sigma_2(\mu, \hat{\theta})$. If $\theta = 1$, incentive compatibility is equivalent to $\sigma_2(\mu, 1) \ge \sigma_2(\mu, -1)$. If $\theta = -1$, incentive compatibility is equivalent to $-\sigma_2(\mu, -1) \ge -\sigma_2(\mu, 1)$, which is again equivalent to $\sigma_2(\mu, 1) \ge \sigma_2(\mu, -1)$. Thus incentive compatibility in period 2 is given by

$$\sigma_2(\mu, 1) \ge \sigma_2(\mu, -1), \quad \forall \mu \in [0, 1]. \tag{P-IC$_2$}$$

This means that reporting the good state should always lead to a higher probability of promotion than reporting the bad state. Note that (P-IC$_2$) also rules out double deviations. Even if the principal falsely reports $\hat{\mu} \neq \mu$ in period 1, as long as (P-IC$_2$) holds, it is optimal for the principal to be truthful in $t = 2$. Hence the only remaining deviation for the principal is to misreport her belief in period 1 and then report the state truthfully in period 2. Such deviations are ruled out by the incentive compatibility constraints in period 1:

$$\sigma_1(\mu)(\mu\sigma_2(\mu,1)-(1-\mu)\sigma_2(\mu,-1)) \geq \sigma_1(\hat{\mu})(\mu\sigma_2(\hat{\mu},1)-(1-\mu)\sigma_2(\hat{\mu},-1)) \quad \forall \mu,\hat{\mu}\in[0,1]. \tag{P-IC$_1$}$$

A mechanism σ is optimal if it solves

$$\begin{aligned} &\max_{\sigma_1,\sigma_2} \int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu,1)-(1-\mu)\sigma_2(\mu,-1))\,\mathrm{d}F(\mu) \\ \text{s.t.}\quad & \text{A-IR, P-IC}_1,\ \text{P-IC}_2\,. \end{aligned}$$

Note that the principal's ex ante payoff from the optimal mechanism may be negative, in which case she will prefer to obtain a payoff of 0 by not inducing the agent to participate in the first place. Since it is straightforward to check whether the principal's ex ante payoff is positive, we restrict attention to mechanisms that satisfy A-IR.

Interpretation of the Model The leading interpretation of our model throughout the paper is that the principal is a firm who tries to retain the agent, who is the worker; see Section 7 for alternative interpretations. The state $\theta\in\{-1,1\}$ represents the firm's value for the worker. We assume that the firm, but not the worker, updates information about the worker's value. This would be the case if the worker's value depends on the demand for the firm's goods, which only the firm observes. Even if productivity is determined by the worker's innate ability, it may be that, by observing the worker, the firm acquires information about the worker's ability that the worker himself is unaware of. The firm's belief about the worker's value cannot be contracted on because it is subjective and not verifiable in a court of law. The firm does not receive any flow payoffs from employing the worker in periods 0 or 1. This would be the case, for example, if the worker is paid his marginal product until he is promoted.[11] The worker obtains a payoff of b from being promoted but has two outside offers. The first outside offer gives the worker a payoff of c_0 and disappears after period 0. The second outside offer is worth c_1 and disappears after period 1. The mechanism can be interpreted as a human resources (HR) policy which governs how the firm may communicate to or promote the worker.

Our assumption that utility is not transferable is adequate for studying environments where an employee's compensation is determined by their position, as is often the case in bureaucratic organizations. For example, the majority of civilian white-collar federal

[11] Alternatively, the worker may be a contractor who may or may not be hired for a project that starts in period 2. In each of periods 0 and 1, the contractor either waits for the possibility of being hired by the firm or leaves and commits himself to an alternative project that precludes him from working for the firm.

employees in the United States are paid according to the General Schedule, which determines the salary for employees in each grade and is set by Congress. The director of a government agency can decide who to employ and which grade its employees belong to, but cannot change the salary for each grade or introduce arbitrary bonus schemes. Similarly, although a manager may have the discretion to promote a worker, the wage for each position may be determined at the corporate level, or it may be regulated by laws that mandate equal pay for equal work.

3 Contractible Signals

3.1 Characterization

Let us first consider the mechanism that maximizes the principal's ex ante payoff subject only to the agent's individual rationality constraint, while ignoring the principal's incentive compatibility constraints. We call this the *contractible-optimal mechanism*, since this mechanism would be optimal for the principal if her signals were contractible, so that the mechanism could depend directly on the true signals rather than the principal's reports about the signals. This may be the case if, for instance, the firm's signals come from formal evaluations (of the worker or the firm), the result of which can be publicly verified ex post. As we discuss in Appendix B.2.2, the principal can also implement the contractible-optimal mechanism if there is a continuum of agents with ex post heterogeneous states, and the principal can commit to marginal distributions.

A contractible-optimal mechanism σ solves

$$\max_{\sigma_1,\sigma_2} \int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu,1) - (1-\mu)\sigma_2(\mu,-1))\,\mathrm{d}F(\mu)$$

$$\text{s.t.} \quad c_0 \le b\int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu,1) + (1-\mu)\sigma_2(\mu,-1))\,\mathrm{d}F(\mu) + c_1\int_0^1 (1-\sigma_1(\mu))\,\mathrm{d}F(\mu) \tag{A-IR}$$

Notice that, if $\sigma_2(\mu,1) < 1$, increasing $\sigma_2(\mu,1)$ increases the objective and relaxes A-IR. This is intuitive because when the state is good, both the principal and the agent prefer promotion. Thus we must have $\sigma_2(\mu,1) = 1$, and finding a contractible-optimal mechanism means choosing $\sigma_1(\mu)$ and $\sigma_2(\mu,-1)$ to solve

$$\max_{\sigma_1(\cdot),\,\sigma_2(\cdot,-1)} \int_0^1 \sigma_1(\mu)(\mu - (1-\mu)\sigma_2(\mu,-1))\,\mathrm{d}F(\mu)$$

$$\text{s.t.} \quad c_0 \le b\int_0^1 \sigma_1(\mu)(\mu + (1-\mu)\sigma_2(\mu,-1))\,\mathrm{d}F(\mu) + c_1\int_0^1 (1-\sigma_1(\mu))\,\mathrm{d}F(\mu)\,. \tag{A-IR}$$

The following lemma simplifies the problem.

Lemma 1 (Memoryless Promotion). *When the principal's signals are contractible, it is without loss to restrict attention to* $\sigma = (\sigma_1, \sigma_2)$ *such that* $\sigma_2(\mu, -1)$ *is constant in* μ.

Proof. Take any mechanism σ. By the intermediate value theorem, there exists $q \in [0,1]$ such that

$$\int_0^1 (1-\mu)\sigma_1(\mu)\sigma_2(\mu,-1)\mathrm{d}F(\mu) = q\int_0^1 (1-\mu)\sigma_1(\mu)\mathrm{d}F(\mu).$$

Let $\sigma_2'(\mu,-1) \coloneqq q$ for all $\mu \in [0,1]$. Clearly, both the objective and the right-hand side of A-IR take the same values under mechanisms (σ_1, σ_2) and (σ_1, σ_2'). □

Lemma 1 holds because a joint distribution of states and promotion decisions (but not interim beliefs) pins down the principal's and the agent's ex ante payoffs. The lemma fails when signals are not contractible, since the principal's interim payoffs, which appear in the principal's incentive compatibility constraints, do depend on her interim beliefs. Indeed, we will see in Sections 4 and 5 that in the optimal mechanism under non-contractible signals, the probability of promotion in period 2 in a given state may depend non-trivially on the principal's period-1 belief report, even after conditioning on the period-1 recommendation to the agent. In such mechanisms, in period 2, the principal knows more than the agent about the promotion probability in each state. Lemma 1 shows that such informational asymmetry is unnecessary if the principal's signals are contractible. Since σ_2 does not depend on the principal's belief conditional on recommendations, upon receiving the recommendation in period 1, the agent knows the exact probability of promotion that he will face in each state if he decides to stay.[12]

By Lemma 1, it is enough to choose $\sigma_1(\mu)$ and a constant $q_E \in [0,1]$, the probability of promoting the agent in the bad state, to solve

$$\begin{aligned} &\max_{\sigma_1, q_E} \int_0^1 \sigma_1(\mu)(\mu - (1-\mu)q_E)\,\mathrm{d}F(\mu) \\ \text{s.t.}\quad & c_0 \le b\int_0^1 \sigma_1(\mu)(\mu + (1-\mu)q_E)\,\mathrm{d}F(\mu) + c_1\int_0^1 (1-\sigma_1(\mu))\,\mathrm{d}F(\mu)\,. \qquad \text{(A-IR)} \end{aligned}$$

[12]Because the principal knows the state, it is always the case that the principal knows more than the agent about the promotion probability unconditional on the state. On the other hand, the agent does not know more than the principal, since although the principal does not observe the mechanism's period-1 recommendation to the agent, once period 2 ensues, the principal knows in equilibrium that the agent was asked to stay and obeyed.

Let us define

$$\begin{aligned} \hat{c} &:= b\int_{c_1/2b}^{1} \mu \mathrm{d}F(\mu) + c_1 F(c_1/2b) \\ \check{c} &:= b\,(1 - F(c_1/2b)) + c_1 F(c_1/2b). \end{aligned} \tag{3}$$

$\hat{c}$ is the ex ante payoff of the agent if he stays in period 1 if and only if the principal's belief is above $c_1/2b$ and is promoted in period 2 if and only if the state is good. $\check{c}$ is the ex ante payoff of the agent if he stays in period 1 if and only if the principal's belief is above $c_1/2b$ and is always promoted in period 2. Clearly, we have $b\mu_0 < \hat{c} < \check{c} < b$. The following proposition characterizes the contractible-optimal mechanism.

Proposition 1 (Contractible-Optimal Mechanism)**.** *Fix $b, c_1 \in \mathbb{R}_+$ and $F \in \Delta[0,1]$. The following are true:*

(i) Suppose $c_0 \in (b\mu_0, \hat{c})$. Then, there exists a mechanism (q_E, μ_E) with $q_E = 0$ and $\mu_E \in (0, c_1/2b)$ that is contractible-optimal. μ_E is unique and strictly and continuously increasing in c_0.

(ii) Suppose $c_0 \in [\hat{c}, \check{c}]$. Then, there exists a mechanism (q_E, μ_E) with $q_E > 0$ and $\mu_E = c_1/2b$ that is contractible-optimal. q_E is unique and strictly and continuously increasing in c_0.

(iii) If $c_0 \in (\check{c}, b)$, then there exists a mechanism (q_E, μ_E) with $q_E = 1$ and $\mu_E \in (0, c_1/2b)$ that is contractible-optimal. μ_E is unique and strictly and continuously decreasing in c_0.

Proof. See Appendix A.1. □

Figure 1 depicts the contractible-optimal mechanism for different values of the agent's ex ante outside option c_0. Note that we have drawn $\sigma_2(\mu, -1)$ only for the values of μ such that $\sigma_1(\mu) > 0$, as $\sigma_2(\mu, -1)$ is irrelevant if $\sigma_1(\mu) = 0$.

Proposition 1 describes how the contractible-optimal mechanism changes as we increase c_0, starting from $b\mu_0$. First, the threshold belief μ_E increases up to $c_1/2b$ (case (*i*)), then q_E increases from 0 to 1 (case (*ii*)), and finally, μ_E decreases back to 0 (case (*iii*)). In case (*i*), raising μ_E benefits the agent because as long as $q_E = 0$ and $\mu \leq c_1/b$, the agent would prefer to leave in period zero rather than wait until period one. A higher μ_E hurts the principal because she would rather have the agent stay, given that she promotes him in period 2 if and only if the state is good. In case (*iii*), lowering μ_E benefits the agent because he will be promoted whenever he obeys the recommendation to stay. A

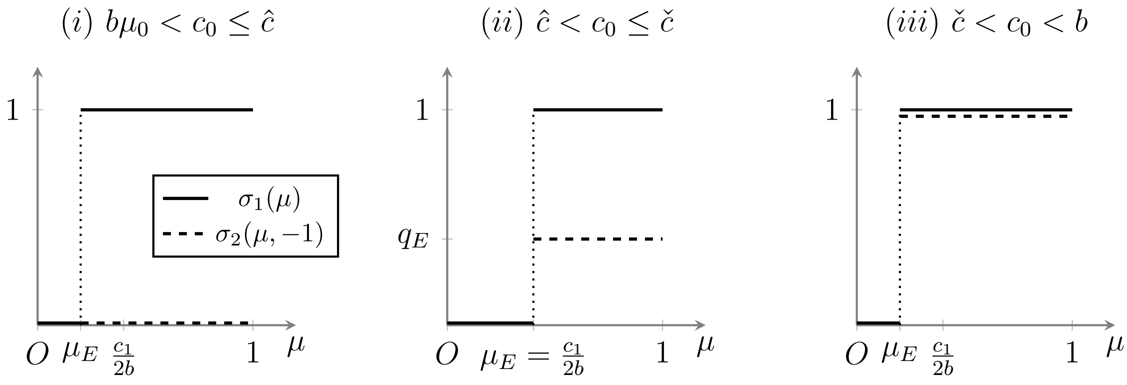

Figure 1: Contractible-Optimal Mechanism

lower μ_E hurts the principal because $\mu_E \leq c_1/2b$ implies that the principal with a belief $\mu \leq \mu_E$ would rather have the agent leave than promote him for sure.

A naive principal may have considered a mechanism that does not communicate to the agent at the interim stage (or, equivalently, always asks him to stay) and makes the promotion decision after the state is realized – for example, always promoting if $\theta = 1$, and promoting with some probability if $\theta = -1$. Proposition 1 says that the principal can do better by using her interim information.[13] In particular, the optimal way to exploit the interim information is to ask the agent to leave when the principal's interim belief is low. This reduces the agent's ex ante opportunity cost of participating in the mechanism because he is able to leave in period 1 and obtain c_1 when the state is unlikely to be good. The principal and the agent would forgo the potential benefit from promotion in the good state, but this is not too costly ex ante because the agent leaves only when the interim belief is low.

We interpret a mechanism that features interim communication as a commitment about future commitments. For example, in panel (ii) of Figure 1, the "stay" recommendation can be seen as a commitment to promote with a probability of at least q_E, and the "leave" recommendation is a commitment not to promote. In period 0, the principal commits to choose one out of these two commitments in period 1.

3.2 Violation of Principal's Incentive Compatibility

When signals are not contractible, the contractible-optimal mechanism characterized by Proposition 1 is almost never incentive compatible for the principal. In case (i), if the principal's interim belief μ is below the threshold μ_E, reporting her belief truthfully gives her an interim payoff of 0. However, she can obtain a strictly positive interim payoff by

[13] Any mechanism that ignores information in period 1 must have $\mu_E = 0$. Proposition 1 shows that such a mechanism can never be contractible-optimal.

misreporting that her belief is above μ_E, inducing the agent to stay, and then reporting truthfully in period 2. In case (*iii*), when the principal's belief is $\mu \in (\mu_E, 1/2)$, it is profitable for her to report $\mu < \mu_E$. Similarly, in case (*ii*), the principal can always profitably misreport her belief except in the knife-edge case where $q_E = \mu_E/(1-\mu_E)$.[14]

In order to incentivize the principal to report her beliefs truthfully, a mechanism must distort the recommendation and promotion decisions away from the contractible-optimal mechanism. The next two sections explore how to optimally introduce such distortions.

4 Optimal Mechanism

4.1 Simplifying the Problem

Recall from Section 2 that an optimal mechanism solves

$$\begin{aligned} &\max_{\sigma_1,\sigma_2} \int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu,1) - (1-\mu)\sigma_2(\mu,-1))\,\mathrm{d}F(\mu) \\ \text{s.t.}\quad &\text{A-IR, P-IC}_1\text{, P-IC}_2\,. \end{aligned}$$

To make the problem linear in the mechanism, we introduce the following change of variables: $\sigma^+(\mu) := \sigma_1(\mu)\sigma_2(\mu,1)$ and $\sigma^-(\mu) := \sigma_1(\mu)\sigma_2(\mu,-1)$. In words, $\sigma^+(\mu)$ is the equilibrium ex ante probability that the agent will be promoted when the principal reports μ and 1. $\sigma^-(\mu)$ is the equilibrium ex ante probability that the agent will be promoted when the principal reports μ and -1. We thus choose three functions, σ_1, σ^+, and σ^-, each mapping $[0,1]$ into $[0,1]$, to solve

$$\max_{\sigma_1,\sigma^+,\sigma^-} \int_0^1 (\mu\sigma^+(\mu) - (1-\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu)$$

$$\text{s.t.}\quad c_0 \le b\int_0^1 (\mu\sigma^+(\mu) + (1-\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu) + c_1\int_0^1 (1-\sigma_1(\mu))\,\mathrm{d}F(\mu) \qquad \text{(A-IR)}$$

$$\mu\sigma^+(\mu) - (1-\mu)\sigma^-(\mu) \ge \mu\sigma^+(\mu') - (1-\mu)\sigma^-(\mu') \quad \forall \mu,\mu' \in [0,1] \qquad (\text{P-IC}_1)$$

$$\sigma^+(\mu) \ge \sigma^-(\mu) \quad \forall\mu \in [0,1] \qquad (\text{P-IC}_2)$$

$$\sigma_1(\mu) \ge \max\{\sigma^+(\mu), \sigma^-(\mu)\} \quad \forall \mu \in [0,1]. \qquad \text{(F)}$$

The feasibility constraint (F) ensures that $(\sigma_1, \sigma^+, \sigma^-)$ corresponds to a mechanism (σ_1, σ_2) with $\sigma_2 \le 1$.[15]

[14] The value of c_0 such that this is true will be defined as $\tilde{c}$ when we describe the optimal mechanism in Proposition 3.

[15] An astute reader might observe that we are no longer requiring $\sigma_2(\mu,1) \ge \sigma_2(\mu,-1)$ when $\sigma_1(\mu) = 0$. This is without loss since the conditional probabilities $\sigma_2(\mu,1)$ and $\sigma_2(\mu,-1)$ do not matter if the mechanism never recommends "stay" given belief μ.

Lemma 2. *It is without loss of optimality to set $\sigma_1(\mu) = \sigma^+(\mu)$, that is, $\sigma_2(\mu, 1) = 1$.*

Proof. Consider the following perturbation: whenever $\sigma_1(\mu) > \sigma^+(\mu)$, we hold $\sigma^+(\mu)$ and $\sigma^-(\mu)$ fixed and reduce $\sigma_1(\mu)$ to $\sigma^+(\mu)$. This relaxes the A-IR constraint, does not violate the other constraints, and does not affect the objective function.[16] $\square$

Lemma 2 implies that if the agent obeyed the recommendation to stay in period 1, and the good state is realized in period 2, the mechanism should promote the agent with probability 1 regardless of μ. Intuitively, if the mechanism sometimes recommends "stay" but does not always promote the agent even in the good state, it would be more efficient to have the agent leave more often in period 1.

Our goal now is to find $\sigma^+(\mu)$ and $\sigma^-(\mu)$ that solve the following maximization problem, which we denote $(\mathcal{P})$.

$$\max_{\sigma^+,\sigma^-} \int_0^1 (\mu\sigma^+(\mu) - (1-\mu)\sigma^-(\mu))\, \mathrm{d}F(\mu)$$

$$\text{s.t.} \quad c_0 \le b\int_0^1 (\mu\sigma^+(\mu) + (1-\mu)\sigma^-(\mu))\, \mathrm{d}F(\mu) + c_1 \int_0^1 (1-\sigma^+(\mu))\, \mathrm{d}F(\mu) \tag{A-IR}$$

$$\mu\sigma^+(\mu) - (1-\mu)\sigma^-(\mu) \ge \mu\sigma^+(\mu') - (1-\mu)\sigma^-(\mu') \quad \forall \mu, \mu' \in [0,1] \tag{P-IC$_1$}$$

$$\sigma^+(\mu) \ge \sigma^-(\mu) \quad \forall \mu \in [0,1]. \tag{P-IC$_2$}$$

The following lemma shows that an optimal mechanism (almost) always recommends the agent to stay if the principal's interim belief is (close enough to) 1.

Lemma 3. *Any optimal mechanism must have $\lim_{\mu\to 1} \sigma^+(\mu) = \sigma^+(1) = 1$.*

Proof. We first show $\sigma^+(1) = 1$. Suppose to the contrary that there exists an optimal mechanism σ with $\sigma^+(1) < 1$. Consider the mechanism $\tilde{\sigma}$ defined by

$$\tilde{\sigma}^+(\mu) = \begin{cases} \sigma^+(\mu) & \text{if } \mu < \frac{1}{2} \\ \sigma^+(\mu) + 1 - \sigma^+(1) & \text{if } \mu \ge \frac{1}{2}, \end{cases}$$

$$\tilde{\sigma}^-(\mu) = \begin{cases} \sigma^-(\mu) & \text{if } \mu < \frac{1}{2} \\ \sigma^-(\mu) + 1 - \sigma^+(1) & \text{if } \mu \ge \frac{1}{2}. \end{cases}$$

The mechanism $\tilde{\sigma}$ is well-defined because P-IC$_1$, evaluated at $\mu = 1$, implies $\sigma^+(1) \ge \sigma^+(\mu)$ for any $\mu \in [0,1]$. It is straightforward to check that $\tilde{\sigma}$ satisfies all the constraints

[16] Note that the perturbation involves reducing $\sigma_1(\mu)$ while increasing both $\sigma_2(\mu, 1)$ and $\sigma_2(\mu, -1)$. In finding the contractible-optimal mechanism, we used a different perturbation argument: whenever $\sigma_2(\mu, 1) < 1$, increase it to 1 while holding $\sigma_1(\mu)$ and $\sigma_2(\mu, -1)$ fixed. This is no longer adequate because it may violate P-IC$_1$.

of $(\mathcal{P})$ and gives the principal a strictly higher ex ante payoff. Thus σ could not have been optimal.

Next, we argue that $\lim_{\mu \to 1} \sigma^+(\mu) = \sigma^+(1)$. Consider the constraint P-IC$_1$. Letting $\mu' = 1$ and taking $\limsup_{\mu \to 1}$ and $\liminf_{\mu \to 1}$ on both sides of the inequality gives us

$$\limsup_{\mu \to 1} \sigma^+(\mu) \geq \sigma^+(1)$$
$$\liminf_{\mu \to 1} \sigma^+(\mu) \geq \sigma^+(1).$$

Since $\sigma^+(\mu) \leq \sigma^+(1)$, we may conclude that $\lim_{\mu \to 1} \sigma^+(\mu) = \sigma^+(1)$. □

4.2 Characterization

We now construct the optimal mechanism by using two simple mechanisms. The first simple mechanism always promotes the agent, regardless of the interim belief μ or the state θ, and is clearly incentive compatible for the principal.

Definition 1. Let $\sigma^+ : [0,1] \to [0,1]$ and $\sigma^- : [0,1] \to [0,1]$. The pair (σ^+, σ^-) is an *always-promote mechanism* if $\sigma^+(\mu) = \sigma^-(\mu) = 1$ for all $\mu \in [0,1]$.

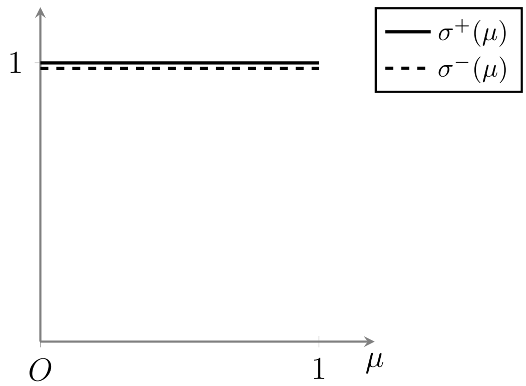

Figure 2: Always-Promote Mechanism

One may recall that, in Figure 1, we plotted $\sigma_1(\mu)$ and $\sigma_2(\mu, -1)$ to depict the contractible-optimal mechanism. In this section, we instead plot $\sigma^+(\mu)$ and $\sigma^-(\mu)$. However, the figures can be directly compared for the following reasons. First, we have shown that $\sigma_1(\mu) = \sigma^+(\mu)$ for all μ. Second, in the contractible-optimal mechanism, either we have $\sigma_1(\mu) = 1$, so that $\sigma_2(\mu, -1) = \sigma^-(\mu)$, or we have $\sigma_1(\mu) = 0$, so that $\sigma_2(\mu, -1)$ is meaningless.

The second simple mechanism asks the agent to leave if the principal's interim belief is sufficiently low. As in the contractible-optimal mechanism, this helps to induce the agent to participate because the agent prefers to receive c_1 for sure rather than receive b with a sufficiently low probability. However, to satisfy the principal's interim incentive compatibility constraints, the mechanism must, upon keeping the agent in period 1, sometimes promote him even in the bad state.

Definition 2. Let $\sigma^+ : [0,1] \to [0,1]$ and $\sigma^- : [0,1] \to \mathbb{R}_+$. The pair (σ^+, σ^-) is a *threshold mechanism* if there exists $\mu^* \in [0,1)$ such that

$$\sigma^+(\mu) = \begin{cases} 0 & \text{if } \mu < \mu^* \\ 1 & \text{if } \mu \geq \mu^* \end{cases}$$

$$\sigma^-(\mu) = \begin{cases} 0 & \text{if } \mu < \mu^* \\ q := \frac{\mu^*}{1-\mu^*} & \text{if } \mu \geq \mu^*. \end{cases}$$

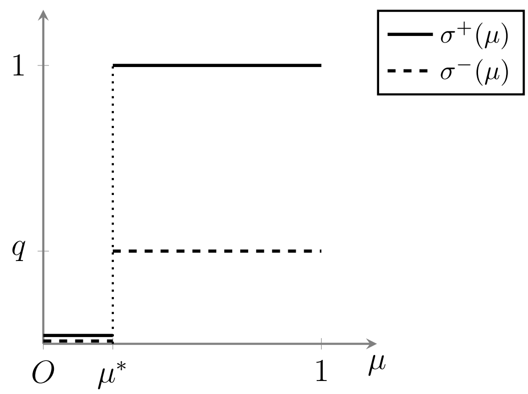

Figure 3: A Threshold Mechanism

If $\mu^* > \frac{1}{2}$, then $q > 1$, so that the threshold mechanism is not actually a feasible mechanism; we might even have referred to the pair (σ^+, σ^-) satisfying the conditions of Definition 2 as a "threshold pre-mechanism". Later in this section, we will construct an optimal mechanism by taking a convex combination (an operation we define shortly) over multiple threshold mechanisms. The resulting convex combination must be a feasible mechanism, but the individual threshold mechanisms need not be.

A threshold mechanism is parametrized by the threshold μ^*. In period 1, the principal chooses from a menu consisting of two options. If the principal reports a pessimistic belief $\mu < \mu^*$, the mechanism tells the agent, "I will never promote you, so please leave". If the principal reports an optimistic belief $\mu \geq \mu^*$, the mechanism tells the agent, "I will promote you with probability at least $q = \frac{\mu^*}{1-\mu^*}$, so please stay". Note that the principal's interim payoff from reporting that her belief is above μ^* is increasing in her true belief. Therefore, the mechanism is incentive compatible for the principal if and only if the principal is indifferent between her two options when her belief is equal to the threshold μ^*. At μ^*, the probability of promotion the worker in the good state jumps up by 1, while the probability of promotion in the bad state jumps up by q. Since the state is good with probability μ^*, for the principal to be indifferent, it must be that $\mu^* = (1-\mu^*)q$. In contrast to the contractible-optimal mechanism (Proposition 1), where the principal could flexibly choose both μ_E and q_E, here μ^* and q are coupled, and there is only one degree of freedom. To incentivize the principal to truthfully reveal whether her belief is below or above the threshold, it must be that reporting an optimistic belief above the threshold forces the principal to promote the agent in the bad state with a probability which is pinned down by the threshold.

Given a family of mechanisms (σ_i^+, σ_i^-), $i = 1, \ldots, I$, we may define a new mechanism (σ^+, σ^-) by taking a convex combination: $\sigma^+ = \sum_{i=1}^I k_i \sigma_i^+$ and $\sigma^- = \sum_{i=1}^I k_i \sigma_i^-$, where $k_i \in [0, 1]$ for each i, and $\sum_{i=1}^I k_i = 1$. Note that P-IC$_1$ and P-IC$_2$ are preserved under convex combination. Figure 4 illustrates a convex combination of a threshold mechanism and the always-promote mechanism.

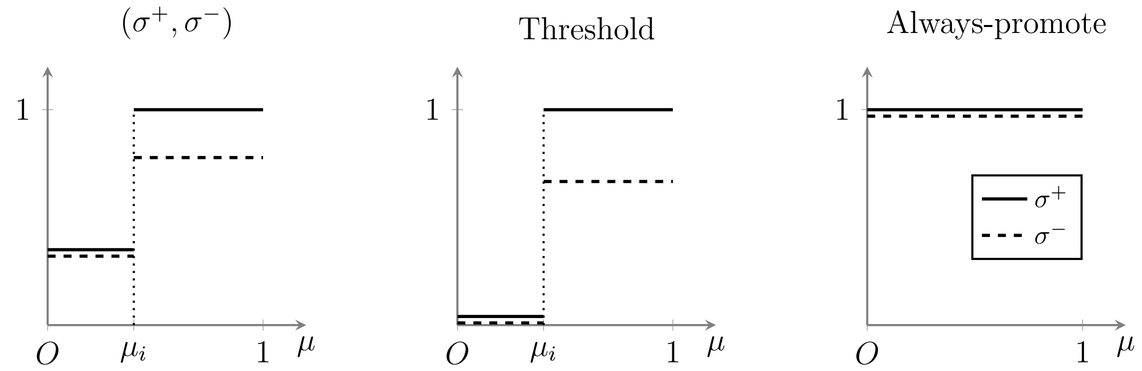

Figure 4: (σ^+, σ^-) is a convex combination of threshold and always-promote.

The following theorem characterizes the optimal mechanism.

Theorem 1. *There exists an optimal mechanism that is a convex combination of the always-promote mechanism and at most three distinct threshold mechanisms.*

Proof (sketch). See Appendix A.2 for a formal proof; here, we provide a sketch. Recall from Section 4.1 that our problem $(\mathcal{P})$ is

$$\max_{\sigma^+,\sigma^-} \int_0^1 (\mu\sigma^+(\mu) - (1-\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu)$$

$$\text{s.t.} \quad c_0 \le b\int_0^1 (\mu\sigma^+(\mu) + (1-\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu) + c_1\int_0^1 (1-\sigma^+(\mu))\,\mathrm{d}F(\mu) \tag{A-IR}$$

$$\mu\sigma^+(\mu) - (1-\mu)\sigma^-(\mu) \ge \mu\sigma^+(\mu') - (1-\mu)\sigma^-(\mu') \quad \forall \mu, \mu' \in [0,1] \tag{P-IC$_1$}$$

$$\sigma^+(\mu) \ge \sigma^-(\mu) \quad \forall \mu \in [0,1]. \tag{P-IC$_2$}$$

Because P-IC$_1$ implies that $\sigma^+(\mu) - \sigma^-(\mu)$ is non-decreasing when $\mu \le 1/2$ and non-increasing when $\mu \ge 1/2$, P-IC$_2$ is equivalent to $\sigma^+(0) \ge \sigma^-(0)$ and $\sigma^+(1) \ge \sigma^-(1)$. P-IC$_1$ contains an infinite number of inequality constraints, but we may define $\phi(\mu) := \sigma^+(\mu) + \sigma^-(\mu)$ and rewrite P-IC$_1$ as

$$\mu\phi(\mu) - \sigma^-(\mu) \ge \mu\phi(\mu') - \sigma^-(\mu') \quad \forall \mu, \mu' \in [0,1].$$

This is reminiscent of the incentive compatibility constraints in a standard monopolist screening problem with a single good if we think of μ as value, ϕ as allocation, and σ^- as transfer. Although the principal cannot pay money to the agent, σ^- serves as a way of transferring utility from the principal to the agent. The envelope theorem implies that P-IC$_1$ holds if and only if

$$\sigma^-(\mu) = \sigma^-(0) + \mu\phi(\mu) - \int_0^\mu \phi(x)\,\mathrm{d}x \quad \forall \mu \in [0,1]$$

$$\phi(\mu) \text{ is non-decreasing.}$$

Thus ϕ pins down $\sigma^-(\mu)$ up to a constant, and our problem can be reduced to that of choosing ϕ and $\sigma^-(0) \in [0,1]$. Fixing $\sigma^-(0) \in [0,1]$, we have a constrained problem of finding a non-decreasing function $\phi : [0,1] \to [2\sigma^-(0), \sigma^-(1)+1]$ that maximizes a linear objective subject to two linear inequality constraints, A-IR and $\phi(1) \ge 2\sigma^-(1)$. By the Bauer maximum principle, there exists ϕ that solves the constrained problem and is an extreme point of the feasible set of the constrained problem.

Let E be the set of non-decreasing functions on $[0,1]$ that take on at most two values, $2\sigma^-(0)$ and $\sigma^-(1)+1$. It is well known that E is the set of extreme points of the set of non-decreasing functions from $[0,1]$ to $[2\sigma^-(0), \sigma^-(1)+1]$. By Proposition 2.1. in Winkler (1988), any extreme point of the feasible set of the constrained problem is a convex combination of at most three elements of E. We may therefore restrict attention to functions $\phi : [0,1] \to [2\sigma^-(0), \sigma^-(1)+1]$ that are non-decreasing step functions with at most three discontinuities. This means that we may also restrict σ^+ and σ^- to be non-decreasing step functions with at most three shared discontinuities. Such a mechanism (σ^+, σ^-) is a convex combination of at most three threshold mechanisms and the always-promote mechanism, where the weight on the always-promote mechanism is $\sigma^-(0)$. $\square$

Corollary 1 follows from Theorem 1 and explicitly describes the outcome of the convex combination. Figure 5 depicts a generic form of the optimal mechanism.

Corollary 1. *There exists an optimal mechanism $\sigma = (\sigma^+, \sigma^-)$ that satisfies the following conditions:*

1. *$\sigma^+(\mu)$ is a non-decreasing step function taking values in $\{p, p_1, p_2, 1\}$, where $0 \leq p \leq p_1 \leq p_2 \leq 1$.*
2. *$\sigma^-(\mu)$ is a non-decreasing step function taking values in $\{p, q_1, q_2, 1\}$, where $p \leq q_1 \leq q_2 \leq 1$*
3. *$\sigma^+(\mu) \geq \sigma^-(\mu)$ for all $\mu \in [0, 1]$.*
4. *σ^+ and σ^- share the same points of discontinuity.*
5. *$\sigma^+(1) = 1$.*

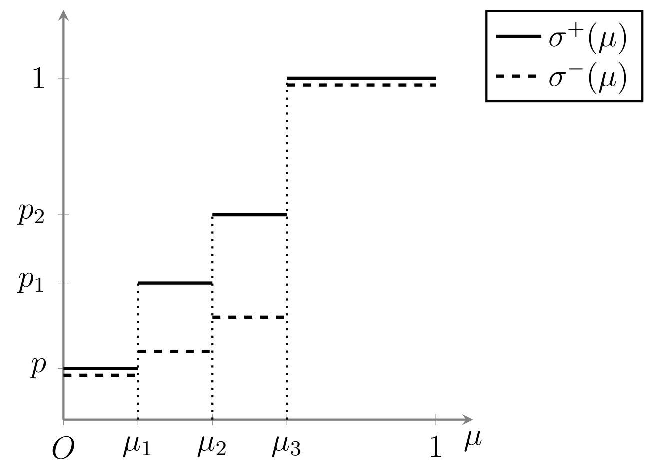

Figure 5: Optimal Mechanism

Given an arbitrary belief distribution F, Theorem 1 reduces the problem of finding an optimal mechanism to a finite-dimensional one. It is without loss of optimality for the recommendation and promotion probabilities to be constant in each interval of interim beliefs, and there need be at most four such intervals. Consequently, it is also without loss of optimality for the mechanism to have the principal only report which interval her belief belongs to. For example, a firm may evaluate worker's productivity with a letter grade of A, B, C, or D. Even if the firm receives more information about the worker than can be conveyed by the letter grades, such a coarse grading scheme performs just as well as any finer grading scheme would.[17]

In light of Theorem 1, let us identify each of the three threshold mechanisms with the pair (q_i, μ_i) for $i = 1, 2, 3$. Let p be the weight placed on the always-promote mechanism. Define

$$\tilde{c} := b \int_{c_1/2b}^{1} \left(\mu + (1 - \mu) \frac{c_1}{2b - c_1} \right) \mathrm{d}F(\mu) + c_1 F(c_1/2b).$$

The next proposition describes how the optimal mechanism depends on the agent's ex ante outside option.

Proposition 2 (Comparative Statics). *There exists $\bar{c} \in [\tilde{c}, b)$ such that the following statements are true.*

(i) If $c_0 \in (b\mu_0, \bar{c}]$, then there exists an optimal mechanism that is a convex combination of three threshold mechanisms with thresholds $\mu_i \in [0, 1)$ for $i = 1, 2, 3$.

(ii) Suppose $c_0, c_0' \in (b\mu_0, \bar{c}]$ with $c_0 > c_0'$. Let $\{\mu_i\}_{i=1,2,3}$ be the three thresholds of an optimal mechanism given c_0, and $\{\mu_i'\}_{i=1,2,3}$ the three thresholds for an optimal mechanism given c_0'. Then, it cannot be that $\mu_i < \mu_j'$ for all $i, j \in \{1, 2, 3\}$.

(iii) There exist $\bar{\mu}_1, \bar{\mu}_2, \bar{\mu}_3 \in (0, 1]$ such that, for each $c_0 \in (\bar{c}, b]$, we can find an optimal mechanism that is a convex combination of the always-promote mechanism and three threshold mechanisms with thresholds $\bar{\mu}_1, \bar{\mu}_2, \bar{\mu}_3$. The weight $p \in (0, 1]$ placed on the always-promote mechanism is unique and is strictly and continuously increasing in c_0.

Proof. See Appendix B.4. □

When the agent's ex ante outside option is low ($c_0 \le \bar{c}$), for the principal to incentivize the agent to participate, using threshold mechanisms is cheaper than using the always-promote mechanism. To see why, consider the ratio at which each of the always-promote mechanism and the threshold mechanism transfers utility from the principal to the agent, relative to the principal's most preferred mechanism, $\sigma^+ \equiv 1$ and $\sigma^- \equiv 0$. Since promotion in the bad state gives the agent b and costs the principal -1, the always-promote mechanism transfers the principal's utility to the agent at a rate of b. The threshold

[17] On the other hand, it is not without loss for the principal to *receive* a coarse signal in period 1. As we show in Appendix B.2, such a coarsening of the information structure relaxes the principal's incentive constraints in period 1 and may increase the principal's optimal ex ante payoff.

mechanism also sometimes promotes the agent in the bad state, transferring utility at a rate of b. However, the threshold mechanism also asks the agent to leave if $\mu < \mu^*$. When μ^* is close to 0, this is a very efficient transfer of utility, since the agent becomes better off by taking the interim outside option of c_1 for sure rather than likely failing promotion and receiving 0, while the principal does not lose much because she was unlikely to promote him even if he stayed. As a result, for low values of c_0, threshold mechanisms dominate the always-promote mechanism.

Once c_0 is above $\bar{c}$, the thresholds remain constant at $\bar{\mu}_i$, and the weight p on the always-promote mechanism increases in tandem with c_0. Intuitively, if c_0 is very high, the only way to meet the agent's participation constraint is to promote him with ex ante probability close to 1, but threshold mechanisms cannot do this because the promotion probability in the bad state can be increased only by decreasing the probability that the agent stays at the interim stage. Thus, when c_0 is sufficiently high, it becomes optimal to place a positive weight p on the always-promote mechanism. Since increasing p transfers utility from the principal to the agent at a constant rate of b, once c_0 is high enough that $p > 0$ is optimal, for any higher value of c_0, it is optimal to increase p while holding the thresholds μ_i-s fixed.

4.3 Optimal Mechanism: Single Threshold Case

We now present a condition that guarantees the existence of an optimal mechanism that is a convex combination of the always-promote mechanism and a single, rather than three, threshold mechanism. For $\lambda \geq 0$, define

$$\begin{aligned} T(\mu^*, \lambda) &:= \int_{\mu^*}^{1} \left(\mu - (1-\mu)\frac{\mu^*}{1-\mu^*} \right) \mathrm{d}F(\mu) \\ &\quad + \lambda \left(b \int_{\mu^*}^{1} \left(\mu + (1-\mu)\frac{\mu^*}{1-\mu^*}) \right) \mathrm{d}F(\mu) + c_1 F(\mu^*) \right). \\ \lambda_0 &:= \frac{\int_0^1 (1-\mu)\mathrm{d}F(\mu)}{b \int_0^1 (1-\mu)\mathrm{d}F(\mu) + f(0)c_1}. \end{aligned}$$

$T(\mu^*, \lambda)$ is the sum of the principal's and the agent's payoffs from a threshold mechanism, where the agent's payoff is weighted by λ.

Consider the following condition:

$$T(\mu^*, \lambda) \text{ is strictly concave in } \mu^* \text{ for any } \lambda \geq \lambda_0. \tag{4}$$

Intuitively, condition (4) holds if the density f of the interim belief distribution is sufficiently flat. For instance, suppose f is differentiable, and define

$$\underline{f} := \min\{f(\mu) \mid \mu \in [0,1]\}.$$

Condition (4) holds if f satisfies

$$0 \leq f'(\mu) \leq \frac{2b}{3b - c_1}\underline{f}, \quad \forall \mu \in [0,1].$$

In particular, condition (4) always holds if the belief distribution F is Uniform.[18]

Theorem 2. *Suppose condition* (4) *holds. Then, there exists an optimal mechanism that is a convex combination of the always-promote mechanism and a single threshold mechanism.*

Proof. See Appendix B.5. □

The following proposition is an analogue of Proposition 2 for when (4) holds.

Proposition 3 (Comparative Statics)**.** *Suppose condition* (4) *holds. Then, there exist* $\bar{c} \in [\tilde{c}, \check{c})$ *and* $\bar{\mu} \in (c_1/2b, 1/2]$ *such that:*

(*i*) *If* $c_0 \in (b\mu_0, \bar{c}]$*, there exists a unique threshold mechanism that is optimal. The threshold* μ^* *satisfies* $\mu^* \leq \bar{\mu}$ *and is strictly increasing in* c_0*.*

(*ii*) *If* $c_0 \in (\bar{c}, b)$*, there exists an optimal mechanism that is a convex combination of the threshold mechanism with threshold* $\mu^* = \bar{\mu}$ *and the always-promote mechanism. The weight* $p \in (0,1]$ *placed on the always-promote mechanism is unique and is strictly and continuously increasing in* c_0*.*

Proof. See Appendix B.7. □

The two cases of Proposition 3 are depicted in Figure 6. When the agent's ex ante outside option is low ($c_0 \leq \bar{c}$), the optimal mechanism is a threshold mechanism, and μ^* increases as c_0 increases. Once c_0 is above $\bar{c}$, μ^* stays fixed at $\bar{\mu}$, and the weight p on the always-promote mechanism increases.

[18] For an alternative sufficient condition for (4) that does not require f to be monotone, see Appendix B.6.

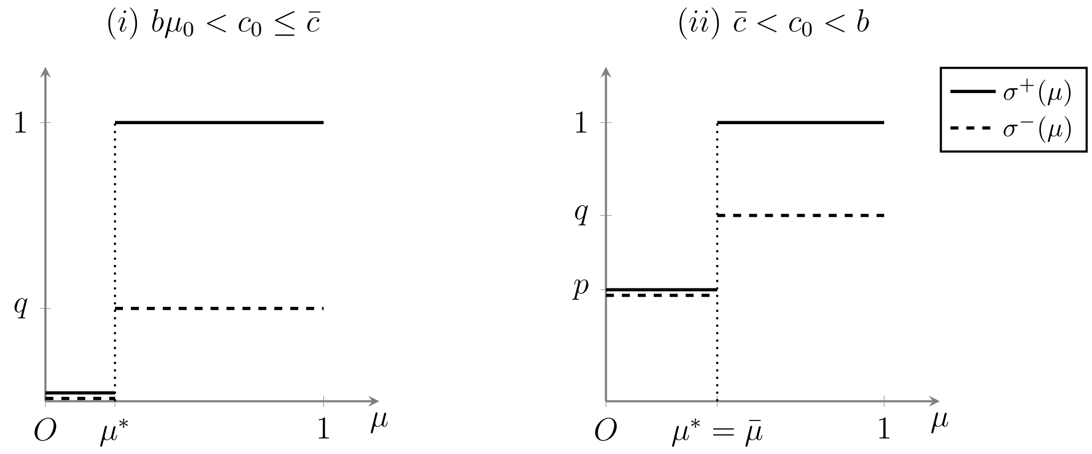

Figure 6: Optimal Mechanism When (4) Holds

5 Properties of the Optimal Mechanism

5.1 Committing to Commit

By Proposition 2, regardless of the parameter values or the interim belief distribution F, it is always optimal to place a positive weight on at least one threshold mechanism with a strictly positive threshold $\mu^* > 0$.[19] Like the contractible-optimal mechanism, we interpret a threshold mechanism, or a mechanism which places weight on a threshold mechanism, as a commitment about future commitments – in period 0, the principal chooses a menu of interim commitments from which she is allowed to choose in period 1.

As was the case for the contractible-optimal mechanism, committing to commit incentivizes the agent to participate in the mechanism by reducing his ex ante opportunity cost of participation. Unlike in the contractible-optimal mechanism, incentive compatibility requires that in order to ask the agent to leave when the principal's interim belief is low, when the principal's interim belief is high, the mechanism must sometimes promote the agent even if the state turns out to be bad. Proposition 2 tells us that committing to commit remains valuable to the principal despite this friction. It is always optimal for the principal to commit in period 0 to provide information in period 1 about her decision in period 2.

It is crucial that the mechanism in period 1 not only restricts the period-2 promotion decision, but also communicates this restriction to the agent. In our environment, the value of committing to commit comes entirely from aiding the agent's decision whether to stay or leave in period 1. Indeed, if we were to assume that the principal were unable to send messages to the agent in period 1, she would have no reason to make any decision

[19]Even if no weight is placed on the always-promote mechanism (case (i)), not all thresholds can equal 0, as otherwise the agent's participation constraint would be violated.

in period 1 with only partial information about the state.[20]

The following observation illustrates the importance of committing to commit: it is even possible for the optimal mechanism to induce a lower ex ante probability of promotion compared to the principal's most preferred mechanism, $\sigma^+(\mu) = 1$ and $\sigma^-(\mu) = 0$.[21] In such cases, the agent's benefit from being able to make a better interim decision outweighs his loss from a lower ex ante probability of promotion.

We do not claim that the optimal mechanism characterized by Proposition 2 is the unique optimal mechanism. However, we argue in Section B.1 that both committing to commit and ignoring information (which we discuss next) are necessary features of any optimal mechanism.

5.2 Ignoring Information

When the agent's ex ante outside option c_0 is sufficiently high, the optimal mechanism places a positive weight $p > 0$ on the always-promote mechanism. That is, with probability p, the mechanism promotes the agent regardless of the principal's reports in period 1 or 2; to make optimal use of the principal's signals, the mechanism commits to sometimes ignore them. As a result, the mechanism sometimes asks the agent to stay in period 1 and promotes him in period 2 even when the principal already knows in period 1 that the state is bad, i.e. $\mu = 0$.[22] On the other hand, the contractible-optimal mechanism always asks the agent to leave if the principal's interim belief is sufficiently low ($\mu \leq \mu_E$). Thus the ignoring of information is a distortion that is caused by the interaction of the principal's incentive compatibility constraints and the agent's participation constraint.

5.3 Memory

Consider $\sigma_2(\mu, \theta)$, which is the probability of promotion in period 2 conditional on the agent having obeyed the interim recommendation to stay and conditional on the state

[20] There may be other environments in which the principal benefits from committing to restrict her decision based on interim information, even if the restriction is kept hidden from the agent. Such a commitment can relax the principal's incentive compatibility constraints and may allow the principal to implement outcomes (i.e. distributions over decisions conditional on each state) that she could not implement without commitment.

[21] This is the case, for example, if $b = 10$, $c_0 = 8.975$, $c_1 = 8$, and $f(\mu) = (-1/10)(\mu - 1/2) + 1$. It can be verified that the threshold mechanism with $\mu^* = 1/2$ is optimal. In this mechanism, the ex ante promotion probability is $\int_{0.5}^{1} f(\mu)\mathrm{d}\mu = 0.4875$. In the principal's most preferred mechanism, the agent is promoted with probability $\int_0^1 \mu f(\mu)\mathrm{d}\mu \approx 0.4917$.

[22] Note that the probability that the agent stays in period 1 is constant at p on the interval $[0, \mu_1)$. Thus there exist $\mu, \mu' \in [0, \mu_1)$ such that $\mu < \mu'$, and such that the agent sometimes stays when the belief is μ and sometimes leaves when the belief is μ'. It would be more efficient to leave more often μ and stay more often at μ', but such Pareto improvements are not achievable because of the principal's incentive constraints.

being θ. In the optimal mechanism, does $\sigma_2(\mu, \theta)$ depend on μ? One might have answered in the negative, since μ is payoff-irrelevant given θ, but our characterization of the optimal mechanism says otherwise. Although we know from Lemma 2 that $\sigma(\mu, 1)$ is constant at 1, $\sigma(\mu, -1)$ can depend non-trivially on μ – for instance, if condition (4) holds and c_0 is large (case (*ii*) of Proposition 3). An interpretation is that, when making the final promotion decision, a firm's HR department must consider not only the firm's final assessment of a worker, but must also retrieve its record of the firm's past assessment. This is in contrast to the contractible-optimal mechanism, which does not need to remember the period-1 belief report in making the period-2 promotion decision (Lemma 1).

6 Implementation

6.1 Midterm Review

How would a firm deciding whether to promote a worker implement the optimal mechanism? When $\sigma_2(\mu, -1)$ is decreasing in μ, if the worker turns out to be unproductive in period 2, he is *less* likely to be promoted when the firm had a higher belief about his value in the past. One might worry that this feature makes implementation difficult, but there exists a natural implementation that mirrors how the mechanism is constructed. It follows from Theorem 1 that the optimal mechanism is a convex combination of 1) a convex combination of at most three threshold mechanisms and 2) the always-promote mechanism. The firm implements this by randomizing between two different promotion schemes – with probability $1-p$, the firm implements the convex combination of threshold mechanisms, and with probability p, the firm implements the always-promote mechanism.

Although our definition of a threshold mechanism allows for the possibility that a threshold mechanism is not a feasible mechanism by itself, the convex combination of all threshold mechanisms that constitute the optimal mechanism must be a feasible mechanism, as otherwise the optimal mechanism would not be feasible either. Moreover, the conditional probability $\sigma_2(\mu, -1)$ of promoting the worker in the bad state is increasing in μ under any convex combination of threshold mechanisms. To see this, consider the convex combination of three threshold mechanisms with thresholds μ_i and weights $k_i \in (0, 1)$, for $i = 1, 2, 3$ and $\sum_{i=1}^{3} k_i = 1$. Suppose $\mu_1 \leq \mu_2 \leq \mu_3$. Then $\sigma_2(\mu, -1) = \sigma^-(\mu)/\sigma^+(\mu)$ is given by

$$\sigma_2(\mu,-1)=\begin{cases}0 & \text{if } \mu\in[0,\mu_1)\\ \dfrac{\mu_1}{1-\mu_1} & \text{if } \mu\in[\mu_1,\mu_2)\\ \dfrac{1}{k_1+k_2}\left(k_1\dfrac{\mu_1}{1-\mu_1}+k_2\dfrac{\mu_2}{1-\mu_2}\right) & \text{if } \mu\in[\mu_2,\mu_3)\\ k_1\dfrac{\mu_1}{1-\mu_1}+k_2\dfrac{\mu_2}{1-\mu_2}+k_3\dfrac{\mu_3}{1-\mu_3} & \text{if } \mu\in[\mu_3,1],\end{cases}$$

which is clearly increasing in μ on $[0,1]$.[23]

Therefore, we may interpret the convex combination of threshold mechanisms as a promotion scheme that consists of a midterm review and a final review. The review may be an assessment of the worker's qualities, or it may be an evaluation of business opportunities that determine the firm's demand for the worker. The midterm review takes place in period 1, and the probability that the worker passes the midterm review is a non-decreasing step function of the firm's belief which is formed during the review. The worker is asked to stay with the firm if he passes the midterm; otherwise, he is no longer considered for promotion and is asked to leave. If the worker obeys the recommendation to stay, he is reviewed again in period 2. During this final review, the firm observes whether the worker is productive. If the worker is productive, he is promoted with certainty. Even if he is not productive, he is promoted with a probability which is increasing in the firm's belief during the midterm review.

The interpretation of the always-promote mechanism is straightforward – the firm simply promotes the worker without a review. Therefore, if the firm cannot contract on its future information, and the worker has a high ex ante outside option, the firm commits to review the worker only some of the time. With probability $1-p$, a review takes place, and the worker is promoted if he passes both a midterm and a final review. With probability p, the worker is not reviewed and is promoted by default. Although there is no exogenous cost in reviewing, the firm sometimes chooses not to review due to strategic concerns.

If the firm employs many workers, a probabilistic promotion threshold can be interpreted as a commitment to promote at least a certain fraction of workers. Under this interpretation, it becomes important whether the state pertains to the firm or the workers. If the state represents the workers' individual abilities and is ex post heterogeneous across workers, the firm may be able to implement the contractible-optimal mechanism. This is not the case if the state represents firm-side uncertainty, so that the workers are

[23] If $\mu\in[0,\mu_1)$, we have $\sigma_1(\mu)=\sigma^+(\mu)=0$, so $\sigma_2(\mu,-1)$ is irrelevant and may be set to zero.

homogeneous ex post.[24] Alternatively, the firm could implement randomness by having the low-productivity worker complete a so-called performance improvement plan, the outcome of which depends mostly on luck rather than the worker's productivity.

6.2 Robust Obedience

Can the worker know whether he is subject to a review or not?[25] Since the worker's participation constraint binds in the optimal mechanism, when deciding whether to stay or leave in period 0, the worker should not know whether he will be reviewed in the future; otherwise, the worker will leave if he knows that he will be reviewed. Although the worker's interim obedience constraint does not bind in the optimal mechanism, if the worker who is asked to stay learns that he was subject to and passed the midterm review, his conditional probability of being promoted decreases, and this may lead him to disobey the recommendation and leave the firm. If this is the case, to implement the optimal mechanism, the firm must ensure that, even after the midterm review takes place, the worker does not know whether the he has been reviewed. For example, if the review consists of an interview, the firm may need to nominally interview the worker even if he will be promoted by default.

However, the threshold mechanism may be sufficiently attractive that the worker's interim obedience is robust to additional information that he may have. To see this, consider a threshold mechanism with a threshold of $\mu^* > c_1/2b$. Under this mechanism, the worker's interim expected payoff after being recommended to stay in period 1 is

$$\begin{aligned}
&\frac{b}{1-F(\mu^*)}\int_{\mu^*}^{1}\left(\mu+(1-\mu)\frac{\mu^*}{1-\mu^*}\right)\mathrm{d}F(\mu) \qquad (5)\\
>&\frac{b}{1-F(\mu^*)}\int_{\mu^*}^{1}\left(\mu^*+(1-\mu^*)\frac{\mu^*}{1-\mu^*}\right)\mathrm{d}F(\mu)\\
=&2b\mu^*\\
>&c_1.
\end{aligned}$$

(5) shows that if a worker knows that he is playing a threshold mechanism with $\mu^* > c_1/2b$, and he has been recommended to stay, he should obey. The same would hold if the worker is playing a convex combination of such threshold mechanisms. Therefore, if the optimal mechanism is a convex combination of the always promote mechanism and threshold mechanisms with thresholds all greater than $c_1/2b$, the worker obeys the interim

[24] See Appendix B.2.2 for a comparison of these two cases.

[25] The convex combination of threshold mechanisms is by construction incentive compatible for the firm, so the firm would report truthfully even if it knew whether the worker is being reviewed.

recommendation to stay even if he learns that he is playing the convex combination of the threshold mechanisms.

In fact, (5) implies an even stronger result. Suppose the worker is asked to stay in period 1 and know that he is subject to reviews. Furthermore, suppose the worker becomes aware that he barely passed the midterm review. That is, the firm's belief was $\bar{\mu}$, which is the lowest possible belief under which he is asked to stay. Even then, the worker will be willing to stay. In other words, the firm *does not lead the worker on.* The firm is sometimes more pessimistic than the worker about the probability of promotion, but even if the firm honestly shared all of its information, the worker would still choose to stay when asked to do so.

Proposition 4 (No Leading On)**.** *Suppose σ is an optimal mechanism that is a convex combination of the always promote mechanism and threshold mechanisms. Suppose all thresholds satisfy $\mu_i > c_1/2b$. Then, the agent obeys the recommendation to stay in period 1 even if he knows the principal's interim belief μ and knows that the convex combination of threshold mechanisms is being played.*

The thresholds are greater than $c_1/2b$ if, for example, condition (4) holds and $c_0 \geq \bar{c}$ (Proposition 3 (*ii*)).

7 Alternative Interpretations

Although the leading interpretation of our model throughout this paper is that of worker retention, our model can be also used to understand relationship-specific investment or forward guidance in policy-making.

7.1 Relationship-Specific Investment

Let us continue to interpret the principal as a firm and the agent as a worker. However, suppose the worker does not have outside options. Instead, in period 0, the worker chooses an amount of human capital investment, $e \geq 0$, that is specific to the firm. The cost of e units of this firm-specific human capital is $1/2e^2$. In period 1, the worker chooses whether to incur a cost of ke to maintain the investment. If the worker maintains the investment in period 1 and is promoted in period 2, he receives a benefit of $be + d$, where $d \geq 0$. The worker's benefit equals 0 regardless of his choice of e if he does not maintain the investment or if he is not promoted in period 2. Neither the worker's choice of e in period 0, nor his choice of whether to maintain the investment in period 1, is observed by the firm.

The firm receives a payoff of 1 from promoting the worker in period 2 if the state is good, and the worker invested at least $\bar{e}$ in period 0 and maintained this investment in period 1. If the state is bad, the worker's investment was less than $\bar{e}$, or the worker did not maintain the investment, the firm's payoff from promotion is -1. The firm obtains 0 from not promoting the worker in period 2. In period 1, the firm privately observes a signal about the state and forms a belief $\mu \sim F$. In period 2, the firm observes the state.

The firm's ideal contract would have the worker invest $\bar{e}$ and maintain it, and then promote the worker if and only if the state is good. On the other hand, the worker does not wish to invest unless he believes the firm is likely to promote him. Because the worker's choice of e is never observed by the firm, it cannot be contracted on; for example, the firm cannot commit to promote the worker only if the worker invested $\bar{e}$. The only way for the firm to incentivize the worker to invest is by committing to promote him with a high probability so that the worker is more likely to benefit from his investment, and by committing to let the worker know in advance, in period 1, if he is unlikely to be promoted, so that he may avoid paying the cost of maintaining his investment.

If the worker chooses to invest a strictly positive amount $e > 0$, he will obey the firm's recommendation to maintain the investment in period 1, as otherwise, he should not have invested to begin with. If the worker is promoted in period 2 with ex ante probability x and is asked with ex ante probability y to maintain the investment in period 1, then the worker's expected payoff from investing e in period 0 is

$$xbe - \frac{1}{2}e^2 - key + xd.$$

Thus the worker's optimal choice of investment level in period 0 is $xb - ky$. The optimal mechanism $\sigma = (\sigma_1, \sigma_2)$ maximizes the firm's expected payoff subject to the firm's incentive compatibility constraints and the constraint that the worker invests at least $\bar{e}$, i.e.

$$\bar{e} \leq b \int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu, 1) + (1-\mu)\sigma_2(\mu, -1))\, \mathrm{d}F(\mu) - k \int_0^1 \sigma_1(\mu)\, \mathrm{d}F(\mu)\,.$$

This is equivalent to our model if we let $c_1 = k$ and $c_0 = \bar{e} + k$.

7.2 Forward Guidance

Suppose the agent is a company that may exert positive externalities in the future but will require a government subsidy to be profitable. For example, the company could be making investments to develop a source of renewable energy that may or may not end up being valuable. The company is willing to incur the investment cost only if it expects the principal, who is a regulator, to subsidize its final product.

In period 2, the regulator decides the level of subsidy, $x \in [0, 1]$, to be provided to the company. The regulator's payoff is θx, where $\theta \in \{-1, 1\}$ is an uncertain state of the world and represents the net marginal benefit of subsidizing the company – the marginal value of the positive externality less the financial cost of a subsidy. The regulator forms a belief $\mu \sim F$ about the state in period 1 and observes the state in period 2. In each of period 0 and 1, the company can either irreversibly shut down or continue to invest in the product. If the company invests in both periods and receives a subsidy of x in period 2, the company's payoff is bx, where $b > 0$. If the company shuts down in period 0 (1), it receives a scrap value of c_0 (c_1). Before the company chooses whether to invest in period 0, the regulator can commit to a mechanism which communicates to the company in period 1 and chooses the subsidy level in period 2 as functions of the regulator's reports.

Our analysis shows how the regulator should provide forward guidance about her future policy. The optimal forward guidance tells the company not only about the subsidy level in period 2, but also about how the regulator will further commit herself in period 1. By promising to reduce uncertainty for the company in period 1, the regulator can induce the company to invest in period 0. Intuitively, when it invests in period 0, the company purchases a real option which allows it to either shut down or invest once more in period 1. By committing to commit, the regulator increases the value of the real option to the company.

8 Conclusion

This paper studies a principal who must make a decision in the future, gradually receives private information about her payoffs from the decision, and faces an agent who wants know what the principal will do. This problem is not uncommon – workers ask firms about promotion prospects, firms ask regulators about future policy, and friends ask one another to reply to dinner invitations – and yet have received little attention from the literature. We introduce a parsimonious model that captures this problem and characterize the optimal mechanism. To convince the agent to wait for her decision, the principal commits today to commit tomorrow. When it is difficult to convince the agent, the principal sometimes ignores her information and decides in the agent's favor.

A Omitted Proofs

A.1 Proof of Proposition 1

Clearly, μ_E must be unique in (i) since the principal's ex ante payoff is strictly decreasing in μ_E. Likewise, uniqueness holds for q_E in (ii) and μ_E in (iii).

The Lagrangian of the problem can be written as

$$\int_0^1 \sigma_1(\mu)(\mu - (1-\mu)q_E)\mathrm{d}F(\mu)$$
$$+\lambda\left(b\int_0^1 \sigma_1(\mu)(\mu + (1-\mu)q_E)\mathrm{d}F(\mu) + c_1\int_0^1 (1-\sigma_1(\mu))\mathrm{d}F(\mu) - c_0\right),$$

where $\lambda \geq 0$ is the multiplier for the constraint A-IR. Since we assumed that $c_0 > b\mu_0$, A-IR must bind. By the Lagrangian sufficiency theorem, if (σ_1, q_E) maximizes the Lagrangian given some $\lambda > 0$, then (σ_1, q_E) is the contractible-optimal mechanism for the value of c_0 such that A-IR holds with equality under (σ_1, q_E).

Rearrange the Lagrangian as

$$\int_0^1 \Big(\sigma_1(\mu)(\mu - (1-\mu)q + \lambda b(\mu + (1-\mu)q) - \lambda c_1) + \lambda c_1\Big)\mathrm{d}F(\mu) - \lambda c_0.$$

The integrand is affine in $\sigma_1(\mu)$, and the coefficient of $\sigma_1(\mu)$ is increasing in μ. Therefore, for fixed values of q_E and λ, μ_E maximizes the Lagrangian if and only if $\sigma_1(\mu) = \mathbb{1}\{\mu \geq \mu_E\}$, where $\mu_E := \frac{q_E + \lambda(bq_E - c_1)}{1+q_E+\lambda b(1-q_E)}$. We may thus rewrite the Lagrangian as a function of q_E and μ_E as follows:

$$\int_{\mu_E}^1 (\mu - (1-\mu)q_E)\,\mathrm{d}F(\mu) + \lambda\left(b\int_{\mu_E}^1 (\mu + (1-\mu)q_E)\,\mathrm{d}F(\mu) + c_1 F(\mu_E) - c_0\right).$$

The derivative of this Lagrangian with respect to q_E is

$$(\lambda b - 1)\int_{\mu_E}^1 (1-\mu)\,\mathrm{d}F(\mu)\,.$$

First, if $\lambda < 1/b$, then the Lagrangian is maximized by $q_E = 0$ and $\mu_E = \frac{\lambda c_1}{1+\lambda b}$. This (q_E, μ_E) is contractible-optimal when the constraint A-IR holds with equality, i.e.

$$b\int_{\mu_E}^1 \mu\,\mathrm{d}F(\mu) + c_1 F(\mu_E) = c_0.$$

As we increase λ continuously from 0 to $1/b$, μ_E increases continuously from 0 to $c_1/2b$, and the LHS of the above equality increases continuously from $b\mu_0$ to $\hat{c}$. This proves case (i) of the proposition.

Next, if $\lambda = 1/b$, then the Lagrangian is maximized by $\mu_E = c_1/2b$ and any $q_E \in [0,1]$. This (q_E, μ_E) is contractible-optimal when A-IR holds with equality, i.e.

$$b\int_{c_1/2b}^{1} (\mu + (1-\mu)q_E)\,\mathrm{d}F(\mu) + c_1 F(c_1/2b) = c_0.$$

As we increase q_E continuously from 0 to 1, the LHS of the above equality increases continuously from $\hat{c}$ to $\check{c}$. This proves case (ii).

Finally, if $\lambda > 1/b$, then the Lagrangian is maximized by $q_E = 1$ and $\mu_E = \max\left\{0, \frac{1-\lambda(b-c_1)}{2}\right\}$. This (q_E, μ_E) is contractible-optimal when A-IR holds with equality, i.e.

$$b(1 - F(\mu_E)) + c_1 G(\mu_E) = c_0.$$

As we increase λ continuously $1/b$ to $1/(b-c_1)$, μ_E decreases continuously from $c_1/2b$ to 0, and the LHS of the above equality increases continuously from $\check{c}$ to b. This proves case (iii).

A.2 Proof of Theorem 1

Define $\phi := \sigma^+ + \sigma^-$ and rewrite P-IC$_1$ as

$$\mu\phi(\mu) - \sigma^-(\mu) \geq \mu\phi(\mu') - \sigma^-(\mu') \quad \forall \mu, \mu' \in [0,1].$$

By standard envelope theorem arguments, this is equivalent to

$$\begin{aligned} &\sigma^-(\mu) = \sigma^-(0) + \mu\phi(\mu) - \int_0^{\mu} \phi(x)\,\mathrm{d}x \quad \forall \mu \in [0,1] \\ &\phi(\mu) \text{ is non-decreasing.} \end{aligned}$$

The problem $(\mathcal{P})$ is thus equivalent to the problem of choosing $\sigma^-(0), \sigma^-(1) \in [0,1]$ and $\phi : [0,1] \to [0, \sigma^-(1) + 1]$ to solve

$$\max \int_0^1 (\mu\phi(\mu) - \sigma^-(\mu))\,\mathrm{d}F(\mu)$$

$$\text{s.t.} \quad c_0 \leq b\int_0^1 (\mu\phi(\mu) + (1-2\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu) + c_1 \int_0^1 (1 - \phi(\mu) + \sigma^-(\mu))\,\mathrm{d}F(\mu) \tag{A-IR}$$

$$\phi(\mu) \text{ is non-decreasing} \tag{P-IC$_1^b$}$$

$$\phi(\mu) \geq 2\sigma^-(\mu) \quad \text{for } \mu = 0, 1 \tag{P-IC$_2$}$$

$$\sigma^-(\mu) = \sigma^-(0) + \mu\phi(\mu) - \int_0^{\mu} \phi(x)\,\mathrm{d}x \quad \forall \mu \in [0,1]. \tag{P-IC$_1^a$}$$

Fix $\sigma^-(0), \sigma^-(1) \in [0,1]$ such that $\sigma^-(0) \le \sigma^-(1)$.[26] Fix $\phi(1) = \sigma^{-1}(1)+1$, which implies $\phi(1) \ge 2\sigma^-(\mu)$, and is without loss by Lemma 3. Consider the constrained problem $(\mathcal{P}')$ of choosing $\phi : [0,1] \to [2\sigma^-(0), \sigma^-(1)+1]$:

$$\max_{\phi} \int_0^1 (\mu\phi(\mu) - \sigma^-(\mu))\,\mathrm{d}F(\mu)$$

$$\text{s.t.} \quad c_0 \le b\int_0^1 (\mu\phi(\mu) + (1-2\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu) + c_1\int_0^1 (1-\phi(\mu)+\sigma^-(\mu))\,\mathrm{d}F(\mu) \tag{A-IR}$$

$$\phi(\mu) \text{ is non-decreasing} \tag{P-IC$_1^b$}$$

$$\phi(1) = \sigma^-(1)+1$$

$$\sigma^-(\mu) = \sigma^-(0) + \mu\phi(\mu) - \int_0^\mu \phi(x)\,\mathrm{d}x \quad \forall \mu \in [0,1]. \tag{P-IC$_1^a$}$$

We wish to show that there exists ϕ which solves $(\mathcal{P}')$ and takes the form described in the statement of the theorem. Let $L^1([0,1],[2\sigma^-(0),\sigma^-(1)+1])$ denote the normed linear space of Lebesgue integrable functions from $[0,1]$ to $[2\sigma^-(0),\sigma^-(1)+1]$. Let $\mathcal{M} \subset L^1([0,1],[2\sigma^-(0),\sigma^-(1)+1])$ denote the convex set of non-decreasing functions in $L^1([0,1],[2\sigma^-(0),\sigma^-(1)+1])$. Let $\mathcal{F} \subset \mathcal{M}$ be the subset of functions in $\mathcal{M}$ that satisfy two linear constraints, A-IR and

$$\begin{aligned} &\sigma^-(1) = \sigma^-(0) + \phi(1) - \int_0^1 \phi(x)\,\mathrm{d}x \\ \Leftrightarrow &\int_0^1 \phi(x)\,\mathrm{d}x = \sigma^-(0)+1. \end{aligned} \tag{6}$$

We may view $\mathcal{F}$ as the feasible set of $(\mathcal{P}')$.[27]

By Helly's selection principle[28], a sequence of functions contained in $\mathcal{M}$ has a subsequence that converges pointwise to an element of $\mathcal{M}$. By the dominated convergence theorem, this subsequence converges in the L^1 norm. Therefore, $\mathcal{M}$ is sequentially compact and thus compact. It is well known that the set $E := \{e : [0,1] \to \{2\sigma^-(0), \sigma^-(1)+1\} \mid e \text{ is non-decreasing}\}$ is the set of extreme points of $\mathcal{M}$. Since $\mathcal{F}$ is the subset of a compact set $\mathcal{M}$ that is the preimage of a linear mapping from $\mathcal{M}$ into a convex set in $\mathbb{R}^2$, Proposition 2.1. in Winkler (1988) allows us to conclude that any extreme point of $\mathcal{F}$, if it exists, is a convex combination of at most three elements of E.

Since $\mathcal{F} \subset \mathcal{M}$ is the continuous preimage of a closed set in $\mathbb{R}^2$, $\mathcal{F}$ is also compact.

[26] P-IC$_1^b$ and P-IC$_1^a$ imply that σ^- is non-decreasing.

[27] A solution in $\mathcal{F}$ will be an equivalence class of functions that are almost-everywhere equivalent. Once we obtain a solution in $\mathcal{F}$, we may select a ϕ that is non-decreasing everywhere and satisfies $\phi(1) = \sigma^-(1)+1$. This ϕ will be a solution to $(\mathcal{P}')$.

[28] See, for example, Kolmogorov and Fomin (1975).

Moreover, the objective function is affine in ϕ. Thus by the Bauer Maximum Principle, given each choice of $\sigma^-(0)$ and $\sigma^-(1)$, there exists an extreme point of $\mathcal{F}$ that maximizes the objective. That is, given $\sigma^-(0)$ and $\sigma^-(1)$, there exists an optimal solution ϕ to $(\mathcal{P}')$ that is a convex combination of at most three elements of E.

Now, let us unwrap (ϕ, σ^-) back into (σ^+, σ^-). Let e_1, e_2, e_3 be the three elements of E whose convex combination gives ϕ. Let μ_i denote the point at which e_i is discontinuous. By P-IC$_1^a$, σ^- is also non-decreasing and is a convex combination of at most three functions e_1^-, e_2^-, e_3^- contained in $E^- := \{e^- : [0,1] \to \{\sigma^-(0), \sigma^-(1)\} \mid e^- \text{ is non-decreasing}\}$. We index each e_i^- so that it shares the same discontinuity as e_i. Likewise, σ^+ is a convex combination of at most three functions e_1^+, e_2^+, e_3^+ contained in $E^+ := \{e^+ : [0,1] \to \{\sigma^-(0), 1\} \mid e^+ \text{ is non-decreasing}\}$, and each e_i^+ shares the same discontinuity as e_i and e_i^-.

Unless $\sigma^-(0) = 1$, in which case the optimal mechanism is simply the always-promote mechanism, none of e_i^+ can be equal to the constant function $e(x) = \sigma^-(0)$.[29] This means that each (e_i^+, e_i^-) is the convex combination of a threshold mechanism with threshold μ_i and the always-promote mechanism, where the weight on the latter is $\sigma^-(0)$. Therefore, the constrained-optimal mechanism, which is a convex combination of (e_i^+, e_i^-) for $i = 1, 2, 3$, is a convex combination of three threshold mechanisms and the always-promote mechanism.

We have thus shown that, for any $\sigma^-(0)$ and $\sigma^-(1)$, there exists a convex combination of three threshold mechanisms and the always-promote mechanism that solves the constrained problem $(\mathcal{P}')$. It remains to prove that there exists a solution of this form to the unconstrained problem $(\mathcal{P})$. We defer this to Lemma 6 in Appendix B.3.

References

Akbarpour, M. and S. Li (2020). Credible auctions: A trilemma. *Econometrica 88*(2), 425–467.

Athey, S., A. Atkeson, and P. J. Kehoe (2005). The optimal degree of discretion in monetary policy. *Econometrica 73*(5), 1431–1475.

Ball, I. (2022). Dynamic information provision: Rewarding the past and guiding the future. *Available at SSRN 3103127*.

Baron, D. P. and D. Besanko (1984). Regulation and information in a continuing relationship. *Information Economics and policy 1*(3), 267–302.

[29]Otherwise, σ^+ would be bounded away from 1, contradicting Lemma 3.

Barro, R. J. and D. B. Gordon (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of monetary economics 12*(1), 101–121.

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Bergemann, D. and S. Morris (2013). Robust predictions in games with incomplete information. *Econometrica 81*(4), 1251–1308.

Bernanke, B. S. (1983). Irreversibility, uncertainty, and cyclical investment. *The quarterly journal of economics 98*(1), 85–106.

Bizzotto, J., J. Rüdiger, and A. Vigier (2021). Dynamic persuasion with outside information. *American Economic Journal: Microeconomics 13*(1), 179–94.

Courty, P. and L. Hao (2000). Sequential screening. *The Review of Economic Studies 67*(4), 697–717.

Dasgupta, S. (2023). Optimal test design for knowledge-based screening. *Available at SSRN 4403119*.

Dixit, R. K., A. K. Dixit, and R. S. Pindyck (1994). *Investment under uncertainty*. Princeton university press.

Ely, J. C. (2017). Beeps. *American Economic Review 107*(1), 31–53.

Ely, J. C., G. Georgiadis, S. Khorasani, and L. Rayo (2023). Optimal feedback in contests. *The Review of Economic Studies*, 2370–2394.

Ely, J. C. and M. Szydlowski (2020). Moving the goalposts. *Journal of Political Economy 128*(2), 468–506.

Gulen, H. and M. Ion (2016). Policy uncertainty and corporate investment. *The Review of Financial Studies 29*(3), 523–564.

Halac, M. and P. Yared (2014). Fiscal rules and discretion under persistent shocks. *Econometrica 82*(5), 1557–1614.

Halac, M. and P. Yared (2022). Instrument-based versus target-based rules. *The Review of Economic Studies 89*(1), 312–345.

Hashimoto, M. (1981). Firm-specific human capital as a shared investment. *The American Economic Review 71*(3), 475–482.

Hashimoto, M. and B. T. Yu (1980). Specific capital, employment contracts, and wage rigidity. *The Bell Journal of Economics*, 536–549.

Jackson, M. O. and H. F. Sonnenschein (2007). Overcoming incentive constraints by linking decisions 1. *Econometrica 75*(1), 241–257.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Kleiner, A., B. Moldovanu, and P. Strack (2021). Extreme points and majorization: Economic applications. *Econometrica 89*(4), 1557–1593.

Kolmogorov, A. N. and S. V. Fomin (1975). *Introductory real analysis*. Courier Corporation.

Krähmer, D. and R. Strausz (2015). Optimal sales contracts with withdrawal rights. *The Review of Economic Studies 82*(2), 762–790.

Kydland, F. E. and E. C. Prescott (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy 85*(3), 473–491.

Lin, X. and C. Liu (2022). Credible persuasion. *arXiv preprint arXiv:2205.03495*.

Lizzeri, A., M. Meyer, and N. Persico (2005). Interim evaluations in dynamic tournaments: The effects of midterm exams. *Working Paper*.

Mattock, M. G., B. J. Asch, J. Hosek, and M. Boito (2019). *The Relative Cost-Effectiveness of Retaining Versus Accessing Air Force Pilots*. Santa Monica, CA: RAND Corporation.

Mortensen, D. T. (1978). Specific capital and labor turnover. *The Bell Journal of Economics*, 572–586.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic theory 18*(2), 301–317.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research 6*(1), 58–73.

Myerson, R. B. (1983). Mechanism design by an informed principal. *Econometrica: Journal of the Econometric Society*, 1767–1797.

Myerson, R. B. (1986). Multistage games with communication. *Econometrica: Journal of the Econometric Society*, 323–358.

Oi, W. Y. (1962). Labor as a quasi-fixed factor. *Journal of political economy 70*(6), 538–555.

Orlov, D., A. Skrzypacz, and P. Zryumov (2020). Persuading the principal to wait. *Journal of Political Economy 128*(7), 2542–2578.

Renault, J., E. Solan, and N. Vieille (2017). Optimal dynamic information provision. *Games and Economic Behavior 104*, 329–349.

Riley, J. and R. Zeckhauser (1983). Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics 98*(2), 267–289.

Rodrik, D. (1991). Policy uncertainty and private investment in developing countries. *Journal of Development Economics 36*(2), 229–242.

Smolin, A. (2021). Dynamic evaluation design. *American Economic Journal: Microeconomics 13*(4), 300–331.

Winkler, G. (1988). Extreme points of moment sets. *Mathematics of Operations Research 13*(4), 581–587.

Yildirim, H. (2005). Contests with multiple rounds. *Games and Economic Behavior 51*(1), 213–227.

B Online Appendix

B.1 Necessary Properties of the Optimal Mechanism

In this section, we show that any optimal mechanism must feature committing to commit, in the sense that σ^+ must vary with μ. We also show that when the agent's ex ante outside option is sufficiently high, any optimal mechanism must sometimes ignore information, in the sense that it must ask the agent to stay when $\mu = 0$ and then, conditional on staying, promote him with probability 1 in period 2.

Lemma 4. *If σ is an optimal mechanism, it cannot be that $\sigma^+(0) > \sigma^-(0) > 0$.*

Proof. Suppose to the contrary that $\sigma = (\sigma^+, \sigma^-)$ is optimal and satisfies $\sigma^+(0) > \sigma^-(0) > 0$. For $\epsilon > 0$, consider a perturbation $\sigma_\epsilon = (\sigma_\epsilon^+, \sigma_\epsilon^-)$, where $\sigma_\epsilon^+, \sigma_\epsilon^- : [0,1] \to \mathbb{R}$, defined by

$$\sigma_\epsilon^+(\mu) = \begin{cases} \sigma^-(0) - \sigma^+(0) & \text{if } \mu < \epsilon \\ 0 & \text{if } \mu \geq \epsilon \end{cases}$$

$$\sigma_\epsilon^-(\mu) = \begin{cases} \frac{\epsilon}{1-\epsilon}(\sigma^-(0) - \sigma^+(0)) & \text{if } \mu < \epsilon \\ 0 & \text{if } \mu \geq \epsilon. \end{cases}$$

Evaluating the principal's payoff at σ_ϵ gives

$$\int_0^\epsilon (\sigma^-(0) - \sigma^+(0))(\mu - \frac{\epsilon}{1-\epsilon}(1-\mu))\, \mathrm{d}F(\mu) \approx O(\epsilon).$$

Evaluating the agent's payoff at σ_ϵ gives

$$\int_0^\epsilon (\sigma^-(0) - \sigma^+(0)) \left((b\mu - c_1) + \frac{b\epsilon}{1-\epsilon}(1-\mu) \right) \mathrm{d}F(\mu),$$

which, as $\epsilon \to 0$, converges to $c_1(\sigma^+(0) - \sigma^-(0)) > 0$.

Let $\sigma_{\text{ideal}} = (\sigma_{\text{ideal}}^+, \sigma_{\text{ideal}}^-)$ given by $\sigma_{\text{ideal}}^+ \equiv 1$ and $\sigma_{\text{ideal}}^- \equiv 0$ denote the principal's ideal mechanism that always recommends the agent to stay and promotes if and only if the state is good. For $\alpha \in [0,1]$, consider the mechanism

$$\sigma' \coloneqq \alpha(\sigma + \sigma_\epsilon) + (1-\alpha)\sigma_{\text{ideal}} = (\alpha(\sigma^+ + \sigma_\epsilon^+) + (1-\alpha),\, \alpha(\sigma^- + \sigma_\epsilon^-)).$$

Since σ^- is non-decreasing, and $\sigma^+ - \sigma^-$ is non-decreasing when $\mu \leq 1/2$, it must be that $\sigma^-(\mu) \geq \sigma^-(0)$ and $\sigma^+(\mu) - \sigma^-(\mu) \geq \sigma^+(0) - \sigma^-(0)$ for $\mu \in [0, 1/2]$. For ϵ small enough, we have $\frac{\epsilon}{1-\epsilon}(\sigma^-(0) - \sigma^+(0)) > -\sigma^-(0)$, so that σ' is a mechanism. σ' is incentive compatible by construction. When ϵ is small enough, $\sigma + \sigma_\epsilon$ represents an arbitrarily efficient transfer of utils from the principal to the agent, relative to σ. On the other hand, taking a convex combination with σ_{ideal} represents a transfer of utils from the agent to the principal that is linear in α. Therefore, there must exist ϵ and α such that σ' satifies A-IR and gives a strictly higher payoff to the principal. □

Lemma 4 allows us to prove the following result.

Proposition 5 (Necessity of Committing to Commit)**.** *If σ is an optimal mechanism, σ^+ cannot be constant in μ.*

Proof. Suppose to the contrary that σ^+ is constant. By P-IC$_1$, σ^- must be constant as well. By Lemma 3, it must be that $\sigma^+ = 1$. By Lemma 4, $\sigma^- = 0$ or $\sigma^- = 1$. The former is ruled out by our assumption that $c_0 > b\mu_0$. The latter is ruled out by our assumption that $c_0 < b$. □

Lemma 4 can also be used to show that ignoring information is necessary when the agent's ex ante outside option is high.

Proposition 6 (Necessity of Ignoring Information)**.** *There exists $c \in (b\mu_0, b)$ such that, if $c_0 \geq c$, then any optimal mechanism must have $\sigma^+(0) = \sigma^-(0) > 0$.*

Proof. By Lemma 4, it is enough to show that when c_0 is sufficiently high, A-IR cannot be satisfied by a mechanism that has $\sigma^-(0) = 0$. If $\sigma^-(0) = 0$, by P-IC$_1$, it must be that $\sigma^-(\frac{1}{4}) \leq \frac{1}{3}$. Since σ^- is increasing, it follows that $\sigma^-(\mu) \leq \frac{1}{3}$ for $\mu \in [0, \frac{1}{4}]$. Thus the agent's ex ante expected payoff under any mechanism with $\sigma^-(0) = 0$ cannot exceed $c := b - \int_0^{\frac{1}{4}} \frac{2}{3}(1-\mu)\,\mathrm{d}F(\mu) < b$. □

Note that the necessary condition $\sigma^+(0) = \sigma^-(0) > 0$ cannot be cast aside as measure-zero properties. This is because $\sigma^-(0) > 0$ serves as a lower bound on $\sigma^-(\mu)$ for all $\mu \in [0, 1]$. Moreover, interim incentive compatibility implies that σ^- must be continuous at $\mu = 0$, and it is without loss to require that σ^+ is continuous at $\mu = 0$.

B.2 Additional Results

B.2.1 Agent-Optimal Mechanism: Tell Me Tomorrow

Suppose the principal has an ex ante outside option of terminating her relationship with the agent. What mechanism maximizes the agent's ex ante payoff subject to the participation and incentive compatibility constraints of the principal? This can be viewed as a model of optimal delegation, where the agent has commitment power and delegates the promotion decision to the principal, who receives private, noncontractible information. Although the agent always wants to be promoted, he must meet the principal's participation constraint and thus chooses to delegate the decision to the principal by committing to a mechanism that makes promotion decisions as a function of the firm's reports. Our novelty relative to most of the literature on delegation is that the principal and the agent disagree not only about *what* the promotion decision should be, but also about the *speed* with which the uncertainty about the decision should be resolved. Unlike the agent, the principal does not incur a cost from waiting to receive more information. The agent-optimal delegation mechanism must therefore induce the principal not only to decide in the agent's favor, but to swiftly restrict her future decision.

Recall that, to solve our original problem of finding the principal-optimal mechanism subject to the agent's participation constraint, we maximize the Lagrangian, which is a weighted sum of the principal's and agent's payoffs. Therefore, if a mechanism σ is a principal-optimal mechanism given that the agent's outside option is c_0, and the principal's ex ante payoff from this mechanism is x, then σ maximizes the agent's ex ante payoff subject to the constraint that the principal's ex ante payoff must be at least x, and the agent's payoff from σ equals c_0. In the optimal mechanism, the agent asks the principal not only to sometimes promote him against her wishes, but also to inform the agent in period 1 about his chance of promotion in period 2.

B.2.2 Commitment to Marginal Distributions

One way to justify the use of a stochastic mechanism is to assume that there is a continuum of agents. For example, if a firm employs a large cohort of workers, the firm may commit to pass 80% of the workers in the midterm review and promote at least 50% of those who passed the midterm. Note that this is only possible if the principal can make different decisions for different agents; this may not be the case, for instance, for a regulator who is legally required to equally subsidize all companies in an industry.

Formally, suppose that there is a unit mass of agents, and that the principal can deviate from the mechanism as long as the marginal distribution of outcomes – the measure of agents who stay in period 1 and the measure of agents in period 2 – is a distribution that can arise from implementing the actual mechanism.[30] What mechanisms can the principal implement? The answer to this question depends crucially on the distribution of the state θ. First, it may be that there is a continuum of agents that are only *ex ante* homogeneous, and both μ and θ are drawn independently and identically for each agent. This would be the case, for example, if θ represents the innate ability of

[30] This is an application of quota mechanisms pioneered by Jackson and Sonnenschein (2007). See Lin and Liu (2022) for a recent application to Bayesian persuasion.

each worker. On the other hand, it may be that the agents are *ex post* homogeneous, so that a single μ is drawn in period 1 and a single θ is drawn in period 2. This may be because θ represents the demand for the firm's goods and thus shared by all workers at the firm.

When the agents are ex ante homogeneous but ex post heterogeneous, being able to commit to marginal distributions allows the contractible-optimal mechanism to be implemented even when the principal's signals are non-contractible. To illustrate, suppose $c_0 \in [\hat{c}, \check{c}]$ and consider the contractible-optimal mechanism $(q_E, \mu_E) = (q_E, c_1/2b)$ (Proposition 1. (ii)). If the principal implements this mechanism, the measure of agents who stay in period 1 is $m_1 := 1 - F(c_1/2b)$, and the measure of agents who are promoted is

$$m_2 := \int_{c_1/2b}^{1} (\mu + (1-\mu)q_E)\,\mathrm{d}F.$$

Suppose the principal commits to recommend "stay" to m_1 agents and promote m_2 agents, and suppose she is allowed to deviate to any direct mechanism as long as these two moment conditions are satisfied. In period 2, the principal will promote all agents with $\theta = 1$ and additionally promote agents with $\theta = -1$ until m_2 agents have been promoted. Knowing this, in period 1, the principal will recommend "stay" to the agents that she is the most optimistic about. This is precisely what the contractible-optimal mechanism specifies. Intuitively, since the contractible-optimal mechanism already allows the principal to keep the agents who are the most likely to be productive and then promote the most productive agents, the principal cannot profitably deviate while honoring her commitment to the the marginal distributions.

Next, suppose the agents are ex post homogeneous. Since only one μ is realized in period 1, period 1 incentive compatibility must be satisfied, and the contractible-optimal mechanism cannot generally be implemented. However, the optimal mechanism can be implemented. For example, suppose condition (4) holds and $c_0 \in (\bar{c}, b)$. To implement the optimal mechanism given by statement (ii) of Proposition 3, says that the principal commits to either keep p agents in period 1 and promote all of them in period 2, or keep all agents in period 1 and promote q of them in period 2.

Finally, one may wish to microfound the principal's ability to commit to marginal distributions by requiring that each agent observes the measure of agents who stay in period 1.[31] This means that agents receive additional information about both the principal's belief about the state and the promotion probabilities conditional on the state. Proposition 4 describes the condition under which the agents obey the recommendation

[31] This is related to the notion of credibility studied by Akbarpour and Li (2020).

to stay even if they become aware of such information.

B.2.3 Principal's Value for Interim Information

Does the principal always prefer a more informative interim belief distribution F? It is clear that no F can be strictly worse for the principal than the uninformative distribution $\underline{F} = \delta_{\mu_0}$, since the principal with an informative F could simply commit not to communicate in the interim period and obtain the same payoff as under $\underline{F}$. Also, no distribution F can be strictly better than the fully informative distribution $\bar{F}$. However, it is possible for the principal to prefer an interim signal that is *less* informative in the sense of Blackwell. To illustrate, let $F = U[0,1]$, $b = 15$, $c_0 = 13$, and $c_1 = 7$. The optimal mechanism is depicted in the left-hand panel of Figure 7, and the principal's expected ex ante payoff is 0.156.

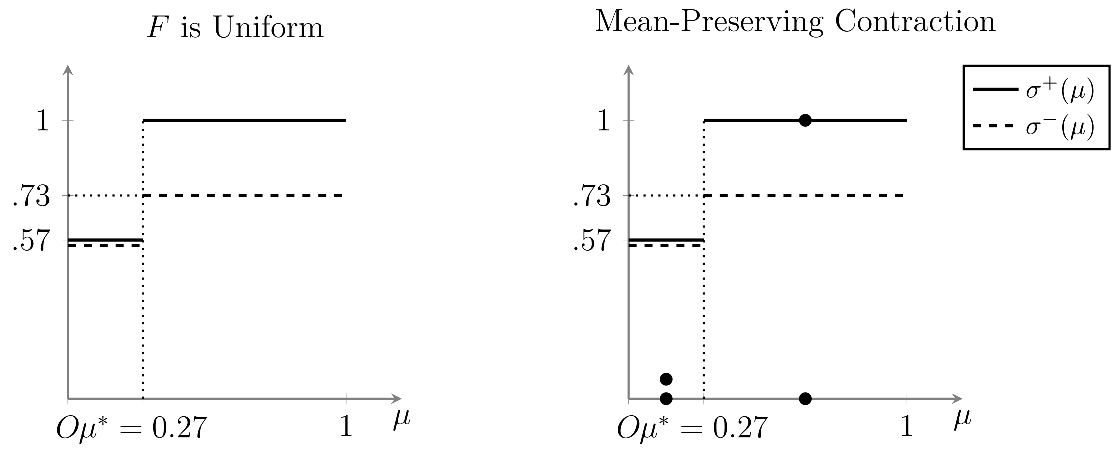

Figure 7: Effect of Mean Preserving Contraction ($b = 15$, $c_0 = 13$, $c_1 = 7$)

Consider the following mean-preserving contraction of the Uniform distribution:

$$\mu = \begin{cases} \mu^*/2 & \text{with proability } \mu^* \\ (1+\mu^*)/2 & \text{with proability } 1-\mu^*. \end{cases}$$

Although this less informative signal does not have a density, it is straightforward to show that under the new distribution, the principal can obtain a payoff of 0.183 by choosing

$$\sigma^+(\mu) = \sigma^-(\mu) = \begin{cases} 0.07 & \text{if } \mu = 0.135 \\ 1 & \text{if } \mu = 0.635. \end{cases}$$

Intuitively, receiving less information relaxes the principal's period-1 incentive compatibility constraints. Because the principal cannot promise not to act upon her period-1 belief, learning less in period 1 can help her commit to a mechanism that otherwise would not have been incentive compatible. Thus a firm may benefit from degrading the quality of information that it acquires about its employees, even when information can be acquired for free.

B.2.4 Deterministic Mechanisms

The optimal mechanism may involve randomization of both action recommendations and promotion decisions. If one were to restrict attention to deterministic mechanisms, there would only remain two direct mechanisms that are incentive compatible and may satisfy the agent's ex ante participation constraint: the always-promote mechanism and the threshold mechanism with $\mu^* = 1/2$. These are depicted in Figure 8. The principal's optimal deterministic mechanism would be the threshold mechanism with $\mu^* = 1/2$ if $c_0 \in (b\mu_0, c_1F(1/2) + b(1 - F(1/2)]$, and would be the always-promote mechanism if $c_0 \in (c_1F(1/2) + b(1 - F(1/2), b)$.

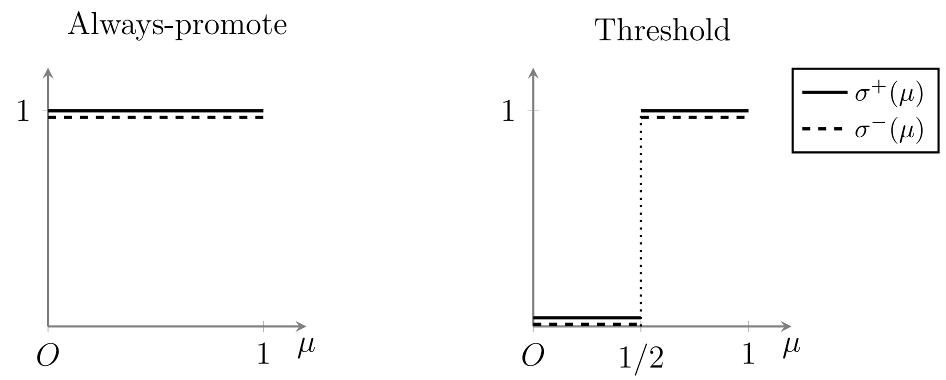

Figure 8: Two Deterministic Mechanisms

In contrast, when the principal's signals are contractible, Proposition 1 shows that there exists a deterministic contractible-optimal mechanism as long as $c_0 \in (b\mu_0, \hat{c}] \cup c_0 \in [\check{c}, b)$. In addition, if signals are contractible, it is without loss for interim recommendations to be deterministic regardless of the value of c_0. Hence, the restriction to deterministic mechanisms interacts with the noncontractibility of signals. It may be costless to use deterministic mechanisms when signals are contractible, but under noncontractibility, the restriction is binding except possibly when the agent's ex ante outside option happens to be $c_0 = c_1F(1/2) + b(1 - F(1/2)$.[32][33]

[32]Even if $c_0 = c_1F(1/2) + b(1 - F(1/2)$, there is no guarantee that the threshold mechanism with $\mu^* = 1/2$ is actually optimal.

[33]Restricting attention to deterministic mechanisms would also be without loss if $c_0 \leq b\mu_0$ or $c_0 = b$, which are corner cases that we have assumed away.

B.3 Lagrangian Approach

This subsection describes the Lagrangian approach to solving problem ($\mathcal{P}$). By Theorem 1, ($\mathcal{P}$) can be reduced to the problem of choosing a 7-tuple $\mathbb{x} = (\mu_1, \mu_2, \mu_3, p, k_1, k_2, k_3)$, which represents the three thresholds, $\mu_1, \mu_2, \mu_3 \in [0, 1)$, the weight p on the always-promote mechanism, and the weight k_i on each of the threshold mechanisms satisfying $p + \sum_{i=1}^{3} k_i = 1$, in order to maximize the principal's expected payoff subject to the individual rationality constraint and $\sigma^+(1) \geq \sigma^-(1)$.[34] Define

$$\mathbb{X}_1 = \{\mathbb{x} \in [0,1)^3 \times [0,1]^4 \mid p + \sum_{i=1}^{3} k_i = 1 \text{ and } \sum_{i=1}^{3} k_i \frac{\mu_i}{1-\mu_i} \leq 1\}$$
$$\mathbb{X}_2 = \{\mathbb{x} \in [0,1)^3 \times [0,1]^4 \mid p + \sum_{i=1}^{3} k_i = 1\}.$$

For multipliers $\lambda, \eta \geq 0$, define

$$\begin{aligned} t(\mu_i, \lambda, \eta) \coloneqq & \int_{\mu_i}^{1} \left(\mu - (1-\mu) \frac{\mu_i}{1-\mu_i} \right) \mathrm{d}F(\mu) \\ & + \lambda \left(b \int_{\mu_i}^{1} \left(\mu + (1-\mu) \frac{\mu_i}{1-\mu_i}) \right) \mathrm{d}F(\mu) + c_1 F(\mu_i) - c_0 \right) \\ & + \eta \left(1 - \frac{\mu_i}{1-\mu_i} \right) \\ a(\lambda) \coloneqq & \int_0^1 (\mu - (1-\mu))\, \mathrm{d}F(\mu) + \lambda (b - c_0)\,. \end{aligned}$$

The expression $t(\mu_i \lambda, \eta)$ represents a weighted sum of the principal's and agent's payoffs (and the term corresponding to P-IC$_2$ at $\mu_i = 1$) induced by a threshold mechanism with a threshold at μ_i. $a(\lambda)$ is the weighted sum of payoffs induced by the always-promote mechanism. The Lagrangians are

$$\mathcal{L}_1 = \mathcal{L}(\mathbb{x}, \lambda, 0) = \sum_{i=1}^{3} k_i t(\mu_i, \lambda, 0) + pa(\lambda)$$
$$\mathcal{L}_2 = \mathcal{L}(\mathbb{x}, \lambda, \eta) = \sum_{i=1}^{3} k_i t(\mu_i, \lambda, \eta) + pa(\lambda)\,.$$

A vector $\mathbb{x}_1 \in \mathbb{X}_1$ solves ($\mathcal{P}$) if and only if there exists $\lambda \geq 0$ such that $\mathbb{x}_1$ solves

[34] A linear combination of threshold mechanisms and the always-promote mechanism by construction satisfies P-IC$_1$ and $\sigma^+(0) \geq \sigma^-(0)$.

$\max_{\mathbb{x}\in\mathbb{X}_1} \mathcal{L}_1$, the A-IR constraint

$$b\int_{\mu_i}^{1}\left(\mu+(1-\mu)\frac{\mu_i}{1-\mu_i})\right)\mathrm{d}F(\mu_i)+c_1F(\mu_i)-c_0\geq 0$$

holds, and the complementary slackness condition is satisfied – either $\lambda = 0$, or the A-IR condition holds with equality. Although $t(\mu_i, \lambda, 0)$ is not guaranteed to be concave in $\mathbb{x}$ for arbitrary belief distributions F, the "only if" direction holds because the problem $(\mathcal{P})$ can be formulated to be linear in $(\phi, \sigma^-(0))$, where $\phi := \sigma^+ + \sigma^-$ as in Section A.2.[35] Similarly, a vector $\mathbb{x}_2 \in \mathbb{X}_2$ solves $(\mathcal{P})$ if and only if there exist $\lambda, \eta \geq 0$ such that $\mathbb{x}_2$ solves $\max_{\mathbb{x}\in\mathbb{X}_2} \mathcal{L}_2$, A-IR and $\sum_{i=1}^{3} k_i \frac{\mu_i}{1-\mu_i} \leq 1$ hold, and the complementary slackness conditions are satisfied – either $\lambda = 0$, or A-IR holds with equality; and either $\eta = 0$, or $\sum_{i=1}^{3} k_i \frac{\mu_i}{1-\mu_i} = 1$. We will make use of both Lagrangians to prove our results.

We next show that we may restrict each μ_i to lie inside a compact interval.

Lemma 5. *There exists $\epsilon > 0$, possibly dependent on b, c_1, and F but independent of c_0, such that it is without loss to solve $(\mathcal{P})$ across solutions $\mathbb{x}$ that have $\mu_i \in [0, 1-\epsilon]$ for each $i = 1, 2, 3$.*

Proof. The partial derivative of $t(\mu_i, \lambda, \eta)$ with respect to μ_i is

$$\frac{\lambda b-1}{(1-\mu_i)^2}\int_{\mu_i}^{1}(1-\mu)\,\mathrm{d}F\,(\mu)-\lambda f(\mu_i)(2b\mu_i-c_1)-\frac{\eta}{(1-\mu_i)^2}, \tag{7}$$

which can be rearranged to

$$\frac{-1}{(1-\mu_i)^2}\int_{\mu_i}^{1}(1-\mu)\,\mathrm{d}F\,(\mu)+\lambda\left(\frac{b}{(1-\mu_i)^2}\int_{\mu_i}^{1}(1-\mu)\,\mathrm{d}F\,(\mu)-f(\mu_i)(2b\mu_i-c_1)\right)-\frac{\eta}{(1-\mu_i)^2}. \tag{8}$$

[35] $(\mathcal{P})$ can be written as follows: choose $\sigma^-(0) \in [0,1]$ and a Borel-measurable and non-decreasing $\phi : [0,1] \to [0,2]$ to solve

$$\max \int_0^1 (\mu\phi(\mu)-\sigma^-(\mu))\,\mathrm{d}F(\mu)$$

$$\text{s.t.}\quad c_0 \leq b\int_0^1(\mu\phi(\mu)+(1-2\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu)+c_1\int_0^1(1-\phi(\mu)+\sigma^-(\mu))\,\mathrm{d}F(\mu) \tag{A-IR}$$

$$\phi(\mu)\geq 2\sigma^-(\mu)\quad \text{for } \mu=0,1, \tag{P-IC$_2$}$$

$$\sigma^-(\mu)=\sigma^-(0)+\mu\phi(\mu)-\int_0^{\mu}\phi(x)\,\mathrm{d}x\quad \forall\mu\in[0,1] \tag{P-IC$_1^a$}$$

$$\phi(1)\leq\sigma^-(1)+1.$$

Clearly, this problem is linear in $(\sigma^-(0), \phi)$.

L'Hopital's Rule implies that

$$\lim_{\mu_i \to 1} \frac{-1}{(1-\mu_i)^2} \int_{\mu_i}^{1} (1-\mu)\,\mathrm{d}F(\mu) = -\frac{1}{2} \lim_{\mu_i \to 1} f(\mu_i)$$

$$\lim_{\mu_i \to 1} \left(\frac{b}{(1-\mu_i)^2} \int_{\mu_i}^{1} (1-\mu)\,\mathrm{d}F(\mu) - f(\mu_i)(2b\mu_i - c_1) \right) = \left(-\frac{3}{2} b + c_1 \right) \lim_{\mu_i \to 1} f(\mu_i).$$

Both are strictly negative due to our assumption that $\lim_{\mu \to 1} f(\mu) > 0$. Therefore, there exists $\epsilon > 0$, independent of λ and η, such that for any $\mu_i \in (1-\epsilon, 1)$, we have (7) < 0. Thus it is without loss to impose $\mu_i \leq 1 - \epsilon$. □

In light of Lemma 5, it is without loss to restrict the feasible sets to

$$\bar{\mathbb{X}}_1 = \{\mathbb{x} \in [0, 1-\epsilon]^3 \times [0,1]^4 \mid p + \sum_{i=1}^{3} k_i = 1 \text{ and } \sum_{i=1}^{3} k_i \frac{\mu_i}{1-\mu_i} \leq 1\}$$
$$\bar{\mathbb{X}}_2 = \{\mathbb{x} \in [0, 1-\epsilon]^3 \times [0,1]^4 \mid p + \sum_{i=1}^{3} k_i = 1\}.$$

We may now complete the proof of Theorem 1 by stating the following lemma, which holds since $\bar{\mathbb{X}}_2$ is compact.

Lemma 6 (Completing the Proof of Theorem 1)**.** *Problem* ($\mathcal{P}$) *has a solution.*

The following lemma is similar to Propositions 3 and 2, but is stated in terms of the Lagrangian multipler rather than the ex ante outside option.

Lemma 7. *There exists a unique* $\bar{\lambda} > 0$ *such that*

(i) For $\lambda < \bar{\lambda}$*, any* $\mathbb{x}_1 \in \bar{\mathbb{X}}_1$ *that maximizes* $\mathcal{L}_1$ *must have* $p = 0$.

(ii) For $\lambda = \bar{\lambda}$*, there exists* $\mathbb{x}_1 \in \bar{\mathbb{X}}_1$ *with* $p > 0$ *that maximizes* $\mathcal{L}_1$.

Proof. Consider the problem of maximizing $\mathcal{L}_1$ across $\mathbb{x}_1 \in \bar{\mathbb{X}}_1$. If $p < 1$, then $\mathbb{x}_1$ must solve

$$V(\lambda) \equiv \max_{\mathbb{x}_1 \in \bar{\mathbb{X}}_1} \sum_{i=1}^{3} k_i t(\mu_i, \lambda, 0)\,.$$

Clearly, we have $V(0) > a(0)$. Consider the difference between the ex ante payoffs to the agent under the always-promote mechanism and the the maximum ex ante payoff that can be given to the agent under a mechanism with $p = 0$:

$$D := b - \max_{\mathbb{x} \in \bar{\mathbb{X}}_1} \sum_{i=1}^{3} k_i \left(b \int_{\mu_i}^{1} \left(\mu + (1-\mu) \frac{\mu_i}{1-\mu_i}) \right) \mathrm{d}F(\mu) + c_1 F(\mu_i) \right).$$

D must be strictly positive.[36] Also, the difference between the maximum ex ante payoff that can be given to the principal under a mechanism with $p = 0$ and the ex ante payoff to the principal under a mechanism with $p = 1$ is bounded above by 2, since the principal's payoff cannot be higher than 1 or lower than -1. Therefore, it must be that when λ is large enough, we have $V(\lambda) < a(\lambda)$.[37]

By the Maximum Theorem, $V(\lambda)$ is continuous in λ. Since $a(\lambda)$ is also continuous in λ, there exists a smallest λ, strictly greater than 0, such that $V(\lambda) = a(\lambda)$; let $\bar{\lambda}$ denote this number. For $\lambda < \bar{\lambda}$, we have $V(\lambda) > a(\lambda)$, so any optimal $\mathbb{x}$ must have $p = 0$. When $\lambda = \bar{\lambda}$, we can maximize the Lagrangian with any value of p. □

Finally, we state a basic property of the Lagrangian.

Lemma 8. *Fix b, c_1, and F. Let λ and λ' be two Lagrangian multipliers such that $\lambda > \lambda'$. Suppose $\mathbb{x} = (\mu_1, \mu_2, \mu_3, p, k_1, k_2, k_3)$ maximizes $\mathcal{L}(\mathbb{x}, \lambda, 0)$ across $\mathbb{X}_1$, and $\mathbb{x}' = (\mu_1', \mu_2', \mu_3', p', k_1', k_2', k_3')$ maximizes $\mathcal{L}(\mathbb{x}, \lambda', 0)$ across $\mathbb{X}_1$. Then the agent's ex ante payoff must be weakly higher under $\mathbb{x}$ than under $\mathbb{x}'$.*

B.4 Proof of Proposition 2

Let $\bar{\lambda}$ be as defined in Lemma 7. Let $\bar{c}$ be the smallest value of the agent's ex ante payoff that can be obtained from a mechanism $\mathbb{x} \in \bar{\mathbb{X}}_1$ that maximizes $\mathcal{L}(\mathbb{x}, \bar{\lambda}, 0)$:

$$\bar{c} := \min\{pb + \sum_{i=1}^{3} k_i \left(b \int_{\mu_i}^{1} \left(\mu + (1-\mu)\frac{\mu_i}{1-\mu_i}) \right) \mathrm{d}F(\mu) + c_1 F(\mu_i) \right) \mid \mathbb{x} \in \arg\max_{\mathbb{x} \in \bar{\mathbb{X}}_1} \mathcal{L}(\mathbb{x}, \bar{\lambda}, 0)\},$$

which is well defined because $\arg\max_{\mathbb{x} \in \bar{\mathbb{X}}_1} \mathcal{L}(\mathbb{x}, \lambda, 0)$ is compact. To induce $\bar{c}$, it must be that $p = 0$, so we have $\bar{c} < b$.

By Lemma 8, for any $c_0 < \bar{c}$, the multiplier λ in the first Lagrangian must be weakly smaller than $\bar{\lambda}$. By the definition of $\bar{c}$, it cannot be that $\lambda = \bar{\lambda}$. Thus $\lambda < \bar{\lambda}$, and by Lemma 7, there exists $\mathbb{x} \in \bar{\mathbb{X}}$ with $p = 0$ that solves the problem $(\mathcal{P})$ and induces c_0 as the agent's ex ante payoff.

It remains to show that $\bar{c} \geq \tilde{c}$. Consider the second Lagrangian $\mathcal{L}_2$, and suppose $\lambda \leq 1/b$. The partial derivative (7) is always strictly negative when $\mu_i > c_1/2b$, so if $\mathcal{L}_2$ is maximized at some $\mathbb{x}_2$, it must be that $\mu_i \leq c_1/2b < 1/2$ for all i. This means that if $\lambda \leq 1/b$, we have $\eta = 0$, and the two Lagrangians are equivalent. Since $t(c_1/2b, 1/b, 0) >$

[36]See the proof of Proposition 6.

[37]Intuitively, as we place an increasingly large weight on the agent's payoff, the using always-promote mechanism eventually becomes optimal.

$a(1/b)$, we have $V(1/b) > a(1/b)$. However, the coefficient of λ in $t(c_1/2b, \lambda, 0)$,

$$b \int_{c_1/2b}^{1} \left(\mu + (1-\mu) \frac{c_1/2b}{1 - c_1/2b} \right) \mathrm{d}F(\mu) + c_1 F(c_1/2b) - c_0 \,,$$

is always smaller than $b - c_0$, which is the coefficient of λ in $a(\lambda)$. Thus $t(c_1/2b, \lambda, 0) > a(\lambda)$ for any $\lambda \le 1/b$, which implies that $V(\lambda) > a(\lambda)$ for any $\lambda \le 1/b$. Thus p must be 0 when $\lambda \le 1/b$, and by the definition of $\bar{\lambda}$, we have $\bar{\lambda} > 1/b$. Also, when $\lambda = 1/b$, it can be seen from (7) that all thresholds must equal $c_1/2b$. The agent's ex ante payoff from this mechanism is $\tilde{c}$. By Lemma 8, we have $\bar{c} \ge \tilde{c}$.

(ii) holds because the principal cannot be better off when c_0 is higher. To show (iii), consider $\mathbb{x} \in \arg\max_{\mathbb{x} \in \bar{\mathbb{X}}_1} \mathcal{L}(\mathbb{x}, \bar{\lambda}, 0)$ that induces $\bar{c}$ for the agent. Let $\bar{\mu}_i$ denote the thresholds of $\mathbb{x}$. By keeping each threshold $\bar{\mu}_i$ fixed, increasing p from 0 to 1, and proportionally decreasing k_i for $i = 1, 2, 3$, we can obtain an optimal mechanism that induces any value of ex ante outside option in $[\bar{c}, b]$. By Lemma 4, it must be that $\bar{\mu}_i > 0$ for each i.

B.5 Proof of Theorem 2

From the definitions of t and T, we have

$$t(\mu_i, \lambda, \eta) = T(\mu_i, \lambda) - \lambda c_0 + \eta \left(1 - \frac{\mu_i}{1 - \mu_i} \right).$$

Since $\eta(1 - \frac{\mu_i}{1-\mu_i})$ is concave in μ_i, as long as $T(\mu_i, \lambda)$ is strictly concave in μ_i, it must be that $t(\mu_i, \lambda, \eta)$ is strictly concave in μ_i. Thus as long as $\lambda \ge \lambda_0$, there can be at most one μ_i that maximizes the second Lagrangian $\mathcal{L}_2$, which means the optimal mechanism can place a positive weight on at most one threshold mechanism.

It remains to show that it is without loss to restrict attention to $\lambda \ge \lambda_0$. Since $t(\mu_i, \lambda_0, 0)$ is strictly concave in μ_i, one can see from (8) that $t(\mu_i, \lambda_0, 0)$ is maximized at $\mu_i = 0$. That is, when $\lambda = \lambda_0$ and $\eta = 0$, the solution to $\mathcal{L}_2$ (and thus $\mathcal{L}_1$) is the principal's most-preferred mechanism, $\sigma^+(\mu) = 1$ and $\sigma^-(\mu) = 0$. This mechanism gives a payoff of $b\mu_0$ to the agent. By Lemma 8, for any mechanism that is optimal given an ex ante outside option $c_0 > b\mu_0$, the corresponding multiplier must be at least as large as λ_0.

B.6 Alternative Sufficient Condition for (4)

The following proposition provides an alternative sufficient condition for (4) to hold.

Proposition 7. *Suppose f is continuously differentiable and satisfies*

$$|f'(\mu)| < \min\left\{\frac{2b}{3b-c_1}, \frac{2\lambda_0 b}{1-\lambda_0 b+\lambda_0(2b-c_1)}\right\}\underline{f} \quad \forall \mu \in [0,1].$$

Then, (4) *holds.*

Proof. The second partial derivative of $t(\mu_i, \lambda, \eta)$ with respect to μ_i is

$$\begin{aligned}\frac{\partial t}{\partial \mu_i} = \frac{\lambda b - 1}{1-\mu_i}\left(\frac{2}{(1-\mu_i)^2}\int_{\mu_i}^{1}(1-\mu)f(\mu)\,\mathrm{d}\mu - f(\mu_i)\right) \\ -\lambda f'(\mu_i)(2b\mu_i - c_1) - 2\lambda b f(\mu_i) - \frac{2\eta}{(1-\mu_i)^3}\,.\end{aligned} \tag{9}$$

Define $\bar{f} := \max\{f(\mu) \mid \mu \in [0,1]\}$ and $\bar{f}' := \max\{|f'(\mu)| \mid \mu \in [0,1]\}$, which are well-defined because f is continuously differentiable. We have

$$\left|\frac{\lambda b - 1}{1-\mu_i}\left(\frac{2}{(1-\mu_i)^2}\int_{\mu_i}^{1}(1-\mu)f(\mu)\,\mathrm{d}\mu - f(\mu_i)\right)\right| \le |\lambda b - 1|\frac{\bar{f} - f(\mu_i)}{1-\mu_i} \le |\lambda b - 1|\bar{f}'\,,$$

and

$$|\lambda f'(\mu_i)(2b\mu_i - c_1)| \le \lambda(2b - c_1)\bar{f}'.$$

Therefore, the second partial (9) is strictly negative if

$$\bar{f}' < \frac{2\lambda b \underline{f}}{|\lambda b - 1| + \lambda(2b - c_1)} \quad \forall \mu \in [0,1]. \tag{10}$$

The assumption of the proposition implies that (10) holds for $\lambda \to \infty$ and $\lambda = \lambda_0$. But the RHS of (10) is single-peaked in λ (with the peak at $\lambda = 1/b$). Therefore, (10) holds for all $\lambda \ge \lambda_0$. □

For example, when $b = 2c_1$, $|f'(\mu)| < 0.47$ implies (4).

B.7 Proof of Proposition 3

Let $\bar{\mu}$ be the unique value of the optimal threshold μ^* when the agent's ex ante outside option is $\bar{c}$. By the same argument that was used to prove case (iii) of Proposition 2, for $c_0 \in (\bar{c}, b)$, it is optimal to place positive weights on both the always-promote mechanism and the threshold mechanism with threshold $\bar{\mu}$. Consider the second Lagrangian $\mathcal{L}_2$. As we showed in Appendix B.4, since a positive weight is placed on the always-promote mechanism, it must be that $\lambda > 1/b$. Since a positive weight is placed on the threshold mechanism and $\lambda > 1/b$, by (7), it must be that either the threshold satisfies $\bar{\mu} > c_1/2b$ or $\eta > 0$. If $\eta > 0$, the period-2 incentive compatibility constraint $\sigma^+(1) \geq \sigma^-(1)$ must bind, so it must be that $\bar{\mu} = 1/2 > c_1/2b$. We have thus shown that $\bar{\mu} > c_1/2b$.

In (i), μ^* must be unique and strictly increasing in c_0 because the principal's payoff is strictly decreasing in μ^*. Likewise, in (ii), p must be unique and strictly and continuously increasing in c_0 because the principal's payoff is strictly and continuously decreasing in p, while the agent's payoff is strictly and continuously increasing in p.

All other results follow directly from Theorem 2 and Proposition 2.