THE UNIVERSITY OF CHICAGO

# I'LL TELL YOU TOMORROW: COMMITTING TO FUTURE COMMITMENTS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY

ANDREW BONGJUNE CHOI

CHICAGO, ILLINOIS

JUNE 2023

To my parents, Yongjae Choi and Saeweon Kim

# TABLE OF CONTENTS

# LIST OF FIGURES

## ACKNOWLEDGMENTS

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, Alex Frankel, Takuma Habu, Jinwoo Kim, Roger Myerson, Derek Neal, Minseon Park, Doron Ravid, Christoph Schlom, Nancy Stokey, Tom Winberry, Kai Hao Yang, and various seminar participants for advice and suggestions.

## ABSTRACT

This paper studies a principal who commits to inform an agent about what the principal will eventually do. The principal wishes to promote the agent if and only if the state is good, and he gradually receives private information about the state. The agent always wishes to be promoted, but faces a decreasing outside option and would rather leave if she expects not to be promoted. The principal optimally induces the agent to stay by committing to commit, that is, by committing today to tell the agent tomorrow about her chances of promotion the day after. When the agent has a high initial outside option, with some probability, the principal promotes the agent regardless of his information—even if he realizes early that the state will turn out to be bad. The principal may ask the agent to stay until he fully observes the state, only to deny her promotion; this does not necessarily mean that the principal leads the agent on. We apply our results to worker retention, relationship-specific investment, and forward guidance.

# CHAPTER 1

# I'LL TELL YOU TOMORROW: COMMITTING TO FUTURE COMMITMENTS

## 1.1 Introduction

A firm employs a worker and will, after two years, decide whether or not to promote her. The firm wants to promote the worker if and only if she is productive. At the end of each year, the firm privately receives information about the productivity of the worker. Ideally, the firm would like to wait two years, receiving information after year one and again after year two, and then promote the worker if and only if they believe her to be productive based on the information received. The problem is that the worker's outside option decreases every year because she has competing offers that disappear every year. Although the worker's most preferred outcome is to stay with the firm for two years and be promoted, she prefers any of her outside offers to staying with the firm and failing promotion. In fact, the worker's outside option today is sufficiently attractive that if she expects to be promoted only if the firm prefers to do so after two years, then she would rather leave today. How can the firm retain the worker until they gather more information?

The firm could commit today to promote the worker with a positive probability even if the firm prefers not to promote her after receiving all the information. This could induce the worker to stay in anticipation of being promoted. However, the firm has another important lever; they could commit today to tell the worker after year one about what the firm will do after year two. We call this a commitment about commitments, or a *higher-order commitment*. To commit means to place restrictions on one's future actions and communicate such restrictions to another party. When the firm commits to commit, they commit today that after year one, they will 1) place restrictions on their future promotion decision and 2) communicate these restrictions to the worker.[1]

Committing to commit can convince the worker to stay today in anticipation of future information. To illustrate, suppose the firm makes the following promise: "after year one, if I believe you are unlikely to be productive, I will no longer consider promoting you. If I believe you are likely to be productive, I will continue to consider promoting you, and will promote you after year two if and only if I still believe that you are likely to be productive. Moreover, I will tell you after year one whether or not you are still being considered for promotion." If the worker trusts the firm to follow through on this promise, and her outside option does not decay too much by next year, she would stay with the firm for a year in anticipation of the update that the firm will provide. Next year, if the firm tells her that she will not be promoted, she can then leave; otherwise, she learns that the firm is relatively optimistic about her productivity, and will stay. In fact, we will show that this commitment is optimal for the firm under the assumptions that the worker's outside option today is not too attractive, and the firm cannot lie to the worker about their belief, e.g. by continuing to consider the worker for promotion despite believing that the worker is unlikely to be productive.

Unfortunately, the firm will lie about their belief if they can. Formally, the mechanism described by the above promise is not *incentive compatible* for the firm. The firms knows that the worker will stay after year one if and only if they continue to consider her for promotion. Moreover, if the worker stays after year one, the firm can make the promotion decision based on the additional information that they receive after year two. Therefore, after year one, no matter how pessimistic the firm is about the worker's productivity, as long as there is the smallest possibility that they might believe her to be productive after year two, they should always pretend to believe that the worker is likely to be productive and continue to

1. Formally, a *higher-order commitment* refers to a principal's commitment to a dynamic mechanism, where the principal himself is a player in the mechanism and receives private information after committing to the mechanism, and the principal's report of his private information affects the mechanism's choice of allocation as well as the mechanism's communication to an agent.

consider her for promotion—the firm should never give up a free option. Anticipating this, the worker does not expect to learn anything even if she stays for a year and so will leave today. Because the firm's belief cannot be contracted on, the promise described above is useless.

How can the firm make the correct promotion decision while retaining the worker until the firm receives enough information? More generally, how should a decision-maker who gradually receives private, noncontractible information incentivize another person to pay the opportunity cost of waiting for the decision? This question is relevant to a wide range of social interactions. A regulator choosing how much to subsidize a firm would benefit from having more time to privately assess the cost and benefit of the subsidy, but the firm may wish to shut down early if they do not expect to be sufficiently subsidized. Alice might wish to delay responding to Bob's dinner invitation because she may or may not be hungry, but if she equivocates, Bob could ask Carol instead.

This paper sheds light on these problems by introducing a parsimonious three-period principal-agent model. The principal represents the firm in our leading example, and the agent represents the worker. A binary state, which can be good or bad, represents the productivity of the worker at the firm. In periods 0 and 1, the agent decides whether to stay and continue interacting with the principal or to leave and take her outside option, which decreases every period. If the agent stays until period 2, the principal chooses whether to promote the agent. The agent always values being promoted, 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 his belief about the state over time. In period 1, he observes a signal that is correlated with the state; in period 2, he observes the state.

A mechanism takes as input the principal's report of his updated belief about the state in each period. In period 1, the mechanism outputs a recommendation to the agent, providing information about the likelihood of being promoted. In period 2, the mechanism decides whether the agent will be promoted. Our problem is to find the mechanism that maximizes the principal's ex ante payoff while ensuring that the agent stays in period 0, the agent obeys the recommendation in period 1, the principal reports truthfully in periods 1 and 2.

We first consider a benchmark case in which the principal's information is contractible and characterize the contractible-optimal mechanism that maximizes the principal's ex ante payoff subject to the agent's participation and obedience constraints, but without being subject to the principal's incentive constraints. In period 1, if the posterior probability that the state is good conditional on the principal's signal is below a threshold, the contractible-optimal mechanism informs the agent that she will fail promotion regardless of the realized state, and she leaves. If the conditional probability is above the threshold, the mechanism asks the agent to stay, promising to promote her in the good state and sometimes also promising to promote her 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. The chance of leaving early in period 1 and the prospect of being promoted in the bad state incentivize the agent to stay in period 0.

We then characterize the optimal mechanism, which maximizes the principal's ex ante payoff subject to the principal's incentive constraints in periods 1 and 2, as well as the agent's participation and obedience constraints. 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 never promotes the agent if the principal reports that his 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 the optimal mechanism, the two parameters of the contractible-optimal mechanism are bundled into one.

The optimal mechanism always places a positive weight on at least one threshold mechanism, meaning that the optimal mechanism always involves a commitment to commit. That is, the principal commits in period 0 to the restrictions on his promotion decision that may be placed in period 1, and also commits to communicate the period-1 restriction to the agent, in part or in full. 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. 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 her 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 can implement threshold mechanisms by committing to conduct a midterm review and a final review. A mixture of threshold mechanisms and the mechanism that always promotes the agent can be implemented by reviewing the worker only a fraction of the time and otherwise promoting them by default. Although the mechanism may ask the agent to stay in period 1 and then deny promotion in period 2, the agent is not necessarily led on—even if the agent knew everything that the principal knew when she was recommended to stay, she may still have chosen to stay.

Our model does not have transfers. The probability of promoting the agent in the bad state plays a role in our analysis similar to that of transfers in a standard single-object monopolist screening model. This observation allows us to appeal to the envelope theorem and an extreme point characterization due to Winkler [1988] to characterize the optimal mechanism.

The final section of the paper presents additional results and extensions. The dual of our problem is the problem of an agent who has commitment power and delegates the promotion decision to a principal who learns about the state and has a participation constraint. In some environments, being able to contract on marginal distributions of outcomes may allow the contractible-optimal mechanism to be implemented. The principal may be worse off if they learn more about the state in period 1.

**Related Literature** This paper 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]. 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.[2] In these models, the principal 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 seller

2. In particular, rules versus discretion studies how a principal who expects to receive private information in the future optimally restricts his 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.

at some point during the game. However, because the buyer's outside option stays constant throughout her interaction with the seller, it is without loss to consider mechanisms in which the buyer always stays until the end of the game. Thus there is no reason for the seller 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 she 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 his own future decision. One might think of our principal as solving an information design problem where the state of the world is also chosen by the principal.

Methodologically, we appeal to the revelation principle for multistage games introduced by Myerson [1986], which states that it is without loss for the mechanism designer to restrict attention to direct mechanisms that satisfy incentive compatibility and obedience. To solve for the optimal mechanism, we apply Proposition 2.1. of Winkler [1988], which characterizes the extreme points of a subset of a linear space defined by a finite number of linear constraints. In our context, the proposition characterizes the extreme points of the feasible set of direct mechanisms. As Kleiner et al. [2021] notes, to be able to apply Winkler's result, it is crucial that the number of linear constraints is finite.[3] Although our problem initially involves infinitely many constraints, we use the envelope theorem to reduce these to a finite number

3. Kleiner et al. [2021] characterizes the set of extreme points of monotone functions that majorize or are majorized by a given function and applies these results to various settings, including mechanism design or information design problems. Because there are uncountably many majorization constraints, Winkler's characterization does not hold in their setting.

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].

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]. This paper considers how a firm can optimally retain its worker via the prospect of promotion. Finally, our results are 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.

## 1.2 Model

A principal (he) and an agent (she) 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 she chooses not to participate, the game ends, the principal receives a payoff of 0, and the agent receives a payoff of $c_0 > 0$. If the agent participates, the game proceeds to period 1.

**Period 1 (Interim)** First, the principal privately updates his belief that the state is good to $\mu \in [0, 1]$. We assume that $\mu$ is drawn according to a distribution $F \in \Delta[0, 1]$ that has a density $f > 0$ and satisfies $\mathbb{E}_F[\mu] = \mu_0$. For technical reasons, we require $f$ to be continuous at $\mu = 1$. The distribution $F$ is commonly known to both the principal and the agent in period 0, but only the principal observes the realized $\mu$.[4]

The principal then sends to the agent a message, which may contain information about the principal's updated belief. We later describe the exact mode of communication. Having observed the principal's message, the agent chooses whether to stay or leave. Let $A = \{\text{stay}, \text{leave}\}$ denote the action set of the agent in period 1. If the agent leaves, the game ends, the principal receives 0, and the agent receives $c_1 \in (0, c_0)$. If the agent stays, the game proceeds to period 2.

**Period 2 (Ex post)** First, the principal privately observes the state $\theta$. Then, the principal decides whether to promote the agent. If the agent is promoted, the agent receives $b > c_0$ and the principal receives $\theta$. 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. $c_0 > b\mu_0$. 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 $\theta = 1$.

**Mechanism** Appealing to the revelation principal [Myerson, 1986], we restrict attention to direct mechanisms, which we denote by $\sigma = (\sigma_1, \sigma_2)$. In period 1, the principal reports his 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

4. The distribution $F$ can be generated by a signal $\pi : \{-1, 1\} \to \Delta[0, 1]$ such that, for all $s \in [0, 1]$,

$$f(s) = \mu_0 \pi(s|1) + (1 - \mu_0)\pi(s| - 1)$$

and

$$(1 - s)\mu_0 \pi(s|1) = s(1 - \mu_0)\pi(s| - 1).$$

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 period-2 report is $\hat{\theta}$, and the agent stayed in period 1.[5]

**Timeline** Given a mechanism $\sigma$, the timeline of the game is as follows.

$t = 0$: Principal chooses mechanism $\sigma$.

Agent chooses whether to participate.

$t = 1$: Principal updates belief to $\mu$ and reports a belief $\hat{\mu}$ to $\sigma$.

$\sigma$ recommends "stay" with probability $\sigma_1(\hat{\mu})$ and recommends "leave" with probability $1 - \sigma_1(\hat{\mu})$.

Agent stays or leaves.

$t = 2$: Principal observes state $\theta$ and reports $\hat{\theta}$ to $\sigma$.

$\sigma$ promotes agent with probability $\sigma_2(\hat{\mu}, \tilde{a}, \hat{\theta})$.

**Optimal Mechanism** The principal's expected payoff from a mechanism $\sigma$ is

$$\int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu, \text{stay}, 1) - (1 - \mu)\sigma_2(\mu, \text{stay}, -1))\, \mathrm{d}F(\mu)\,. \tag{1.1}$$

Our goal is to find a mechanism $\sigma$ 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 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 $\mu$ and $\theta$ and trivially satisfy the

5. Formally, we a mechanism a pair of functions $\sigma_1 : [0, 1] \to [0, 1]$ and $\sigma_2 : [0, 1] \times A \times \{-1, 1\} \to [0, 1]$, each of which are Borel measurable.

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

$$c_1 \leq b \frac{\int_0^1 \sigma_1(\mu)(\mu\sigma_2(\mu,1) + (1-\mu)\sigma_2(\mu,-1))\mathrm{d}F(\mu)}{\int_0^1 \sigma_1(\mu)\mathrm{d}F(\mu)}. \tag{1.2}$$

The agent's ex ante individual rationality constraint is

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

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 she is going to disobey when asked to stay, then she will always receive a payoff of $c_1$ from participating in the mechanism. However, by not participating in the first place, she receives $c_0 > c_1$. We can thus ignore the agent's obedience constraints.

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

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

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 his 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:

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

A mechanism $\sigma$ 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 he will prefer to obtain a payoff of 0 by not inducing the agent to participate in the first place. However, since it is straightforward to check whether the principal's ex ante payoff is positive, we restrict attention to mechanisms 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 1.8 for alternative interpretations. The state $\theta \in \{-1,1\}$ represents the worker's productivity at the firm. We assume that the firm, but not the worker, updates information about the worker's productivity over time. This would be the case if the worker's productivity depends on the demand for the firm's goods, which only the firm observes. Even if productivity is determined 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 herself is unaware of. We make the simplifying assumption that 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 her marginal product until she is promoted.[6] 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.

We assume that utility is not transferable between the firm and the worker, and the firm retains the worker through the prospect of promotion. For example, in the United States, the majority of civilian white-collar federal employees 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 belongs to, but cannot change the salary for each grade or introduce arbitrary bonus schemes. Similarly, an executive in a large company may have the discretion to promote a worker, but the salary package may be determined at the corporate level.

## 1.3 Example

We illustrate via an example how the principal benefits from providing interim information to the agent about the principal's future promotion decision. Let $b = 5$, $c_0 = 2.8$, $c_1 = 2$, and $F = U[0, 1]$. Let us restrict attention to mechanisms such that $\sigma_2(\mu, 1) = 1$ for all $\mu$. Without any constraints, the mechanism that maximizes the principal's payoff would be given by $\sigma_1(\mu) = 1$ and $\sigma_2(\mu, -1) = 0$ (Figure 1.1 (a)). That is, the agent always stays in period 1 and is promoted if and only if $\theta = 1$. Unfortunately, this gives the agent an ex ante payoff of 2.5, which is lower than her ex ante outside option.

One way to incentivize the agent to participate is to promote her in the bad state with some probability. For instance, choosing $\sigma_1(\mu) = 1$ and $\sigma_2(\mu, -1) = 0.12$ would satisfy the

6. 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 herself to an alternative project that precludes her from working for the firm.

agent's participation constraint with equality and give the principal an ex ante payoff of 0.44 (Figure 1.1 (b)). However, the principal does better under the mechanism given by $\sigma_1(\mu) = \mathbb{1}\{\mu \geq 0.2\}$ and $\sigma_2(\mu, -1) = 0$ (Figure 1.1 (c)). That is, the agent is asked to leave if the principal's interim belief is below 0.2; and if the agent is recommended to stay and obeys, she is promoted in period 2 if and only if $\theta = 1$. This mechanism also satisfies the agent's participation constraint with equality, but gives the principal a higher ex ante payoff of 0.48. Intuitively, it is efficient to ask the worker to leave in advance if she is unlikely to be valuable to the principal.

The problem with the last mechanism is that the principal's incentive compatibility constraints are violated.[7] To see this, consider the principal's choice of report in period 1 when his belief realization is $\mu < 0.2$. If he reports truthfully, the agent leaves, so his interim expected payoff is 0. However, if he deviates by reporting some $\mu' \geq 0.2$, the agent stays. By then reporting the state truthfully in period 2, the principal can induce the mechanism to promote the agent in the good state. This deviation gives the principal an interim expected payoff of $\mu$ and is thus profitable.

To incentivize the principal to truthfully report a low belief that leads the agent to leave, the mechanism must make it costly for the principal to report a high belief that leads the agent to stay. Consider the mechanism given by $\sigma_1(\mu) = \mathbb{1}\{\mu \geq \mu^* \approx 0.0725\}$ and $\sigma_2(\mu, -1) = q \approx 0.0782$ (Figure 1.1 (d)). This mechanism satisfies the agent's participation constraint with equality and gives the principal an ex ante payoff that is approximately equal to 0.4637. Under this mechanism, it is costly for the principal to induce the agent to stay in period 1 because the agent must then be promoted with probability $q$ in the bad state. The interim expected cost of promotion in the bad state is higher when the bad state is more likely, that is, when the principal's interim belief is lower. In fact, the principal is willing to incur this cost precisely when $\mu \geq \mu^*$; in other words, this mechanism is incentive

7. We later show that this mechanism is optimal if we ignore the principal's incentive compatibility constraints.

compatible for the principal. Our results in section 1.5 imply that this mechanism is indeed optimal among all feasible mechanisms.

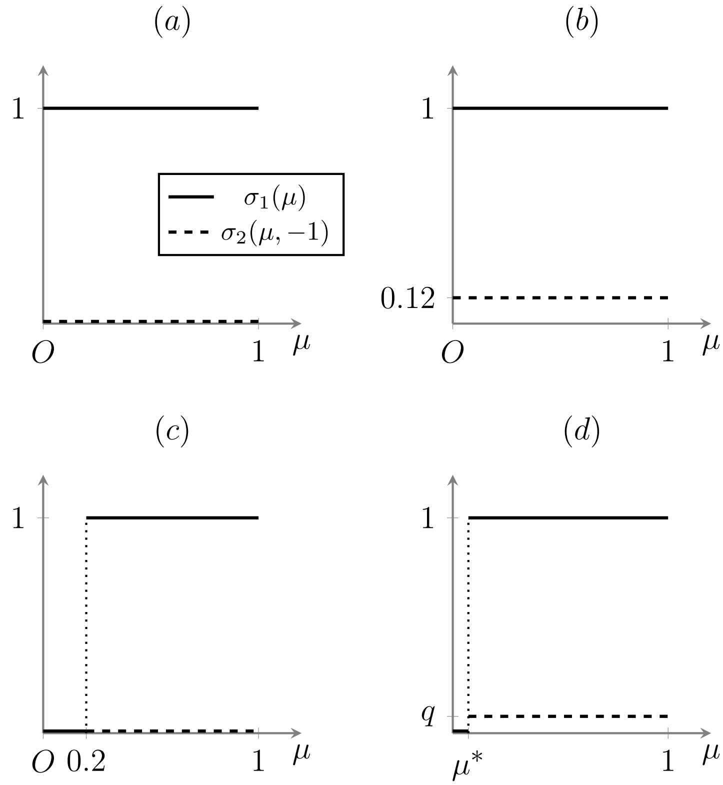


Figure 1.1: Example

## 1.4 Contractible Signals Benchmark

### *1.4.1 Characterization*

Let us first consider as a benchmark 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 his signals were contractible, so that the mechanism can 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 signal comes from a formal evaluation of the worker, the result of which can be publicly verified when the promotion decision is made.[8]

A contractible-optimal mechanism $\sigma$ solves

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

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

If $\sigma_2(\mu,1) < 1$, we can increase $\sigma_2(\mu,1)$ to increase the objective while relaxing A-IR. Intuitively, 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

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

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

The following lemma simplifies the problem.

**Lemma 1** (Memoryless Promotion)**.** *There exists a contractible-optimal mechanism $\sigma = (\sigma_1, \sigma_2)$ such that $\sigma_2(\mu,-1)$ is constant in $\mu$.*

*Proof.* Suppose $\sigma$ is a contractible-optimal mechanism. By the intermediate value theorem, there exists $q_E \in [0,1]$ such that

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

8. The firm would not benefit from announcing the exam results before the worker chooses whether to stay or leave in period 1, since this gives the worker additional information about her productivity and therefore about her chance of promotion.

Let $\sigma_2'(\mu, -1) \coloneqq q_E$ for all $\mu \in [0,1]$. Clearly, the mechanisms $(\sigma_1, \sigma_2)$ and $(\sigma_1, \sigma_2')$ induce the same probability of promotion under each state and the same probability of the agent staying in period 1. $\square$

The proof of Lemma 1 shows that if there is a contractible-optimal mechanism such that $\sigma_2(\mu, -1)$ varies with the period-1 belief report, we can make it constant without affecting the outcome induced by the mechanism. Note that this argument fails when signals are not contractible, as making $\sigma_2(\mu, -1)$ constant may lead to a violation of the principal's incentive compatibilty to report truthfully in period 1.

In a general mechanism, the probability of promotion in period 2 in a given state can depend non-trivially on the principal's period-1 belief report, even after conditioning on the period-1 recommendation to the agent. Indeed, we will see in Section **??** that this is sometimes the case in the optimal mechanism. 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 $\sigma_2$ 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 probability that she will face in each state if she decides to stay.[9]

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

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

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

9. 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.

Since both integrands are increasing in $\mu$, we have the following lemma.

**Lemma 2.** *There exists a contractible-optimal mechanism such that* $\sigma_1(\mu) = \mathbb{1}\{\mu \geq \mu_E\}$ *for some* $\mu_E \in [0,1]$.

*Proof.* See Appendix 1.11.1. □

We may therefore restrict attention to $(q_E, \mu_E) \in [0,1]^2$ that solves

$$\max_{q_E,\mu_E} \int_{\mu_E}^{1} (\mu - (1-\mu)q_E)\mathrm{d}F(\mu)$$
$$\text{s.t.} \quad c_0 \leq b\int_{\mu_E}^{1} (\mu + (1-\mu)q_E)\mathrm{d}F(x) + c_1 F(\mu_E).$$

Define

$$\hat{c}_0 \coloneqq b\int_{c_1/2b}^{1} \mu\mathrm{d}F(\mu) + c_1 F(c_1/2b)$$
$$\check{c}_0 \coloneqq b\,(1 - F(c_1/2b)) + c_1 F(c_1/2b).$$

The value $\hat{c}_0$ is the ex ante payoff of the agent if she 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 she stayed and the state is good. The value $\check{c}_0$ is the ex ante payoff of the agent if she stays in period 1 if and only if the principal's belief is above $c_1/2b$ and is always promoted conditional on staying. Clearly, we have $b\mu_0 < \hat{c}_0 < \check{c}_0 < b$. 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* $c_0 \in (b\mu_0, \hat{c}_0)$. *Then, there exists a unique mechanism* $(q_E, \mu_E)$ *with* $q_E = 0$ *and* $\mu_E \in (0, c_1/2b)$ *that is contractible-optimal.* $\mu_E$ *is strictly and continuously increasing in* $c_0$.

(*ii*) *Suppose* $c_0 \in [\hat{c}_0, \check{c}_0]$. *Then, there exists a unique mechanism* $(q_E, \mu_E)$ *with* $q_E > 0$ *and* $\mu_E = c_1/2b$ *that is contractible-optimal.* $q_E$ *is strictly and continuously increasing in* $c_0$.

(*iii*) *If* $c_0 \in (\check{c}_0, b)$, *then there exists a unique mechanism* $(q_E, \mu_E)$ *with* $q_E = 1$ *and* $\mu_E \in (0, c_1/2b)$ *that is contractible-optimal.* $\mu_E$ *is strictly and continuously decreasing in* $c_0$.

*Proof.* See 1.11.2. □

Figure 1.2 depicts the contractible-optimal mechanism for different values of the agent's initial outside option, $c_0$. Note that we have drawn $\sigma_2(\mu, -1)$ only for the values of $\mu$ such that $\sigma_1(\mu) > 0$, as $\sigma_2(\mu, -1)$ is irrelevant if $\sigma_1(\mu) = 0$.

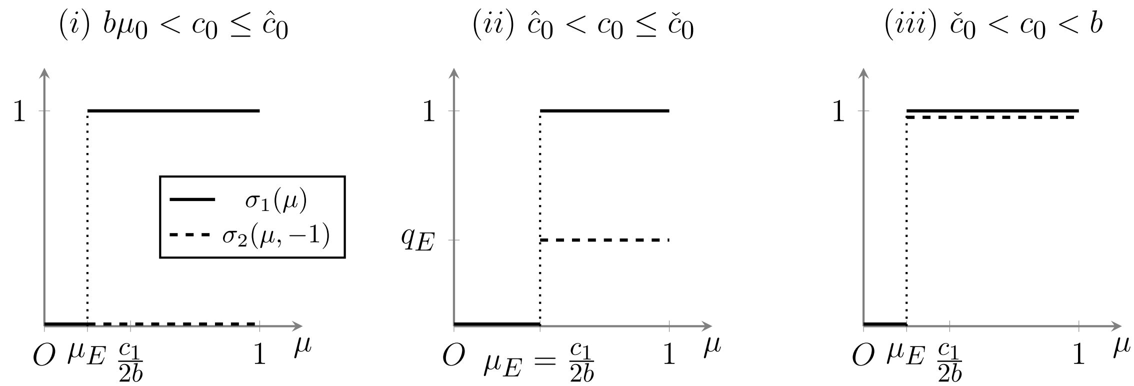


Figure 1.2: Contractible-Optimal Mechanism

Proposition 1 describes how the contractible-optimal mechanism changes as we increase $c_0$, starting from $b\mu_0$. First, the threshold belief $\mu_E$ increases up to $c_1/2b$ (case (*i*)), then $q_E$ increases from 0 to 1 (case (*ii*)), and finally, $\mu_E$ decreases back to 0 (case (*iii*)). In case (*i*), raising $\mu_E$ 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 to leave. A higher $\mu_E$ hurts the principal because he would rather have the agent stay, given that he promotes her in period 2 if and only if the state is good. In case (*iii*), lowering $\mu_E$ benefits the agent because she will be promoted whenever she obeys the recommendation to stay. A lower $\mu_E$ 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 her for sure. Since we assume $c_0 > c_1$, if $c_1$ is sufficiently high, it may be that we are never in cases (*i*) or (*ii*).

A naive principal may have considered a mechanism that does not communicate to the agent at the interim stage (or, equivalently, always asks her to stay) and makes the promotion decision based on the state $\theta$. Proposition 1 says that the principal can do better by using the interim information.[10] 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 she 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.

Finally, we note that when the agent's ex ante outside option is sufficiently high ($c_0 > \check{c}_0$), whether or not she will be promoted in period 2 is fully determined in period 1, and this is fully communicated to the agent in period 1. That is, in period 0, the principal says to the agent, "I will tell you in period 1 whether or not you will be promoted in period 2."

### *1.4.2 Violation of Principal's Incentive Compatibility*

The contractible-optimal mechanism characterized by Proposition 1 is never incentive compatible for the principal except in a knife-edge case.[11] When $b\mu_0 < c_0 \leq \hat{c}_0$ (case (*i*)), even if the principal's belief is below the threshold $\mu_E$, he can deviate upward and misreport that his belief is above $\mu_E$, inducing the agent to stay. This is profitable for the principal because he can then promote truthfully in period 2 and promote the agent if and only if $\theta = 1$.

10. 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.

11. The only case in which the contractible-optimal mechanism is incentive compatible is when $\mu_E = c_1/2b$ and $q_E = \mu_E/(1 - \mu_E)$. The value of $c_0$ such that this is true will be defined as $\tilde{c}_0$ when we describe the optimal mechanism in Proposition 3.

On the other hand, when $\check{c}_0 \leq c_0 < b$ (case (*iii*)), the principal with an interim belief in $(\mu_E, 1/2)$ will have incentives to deviate downward and claim that his belief is below the threshold $\mu_E$, since he prefers not to promote the agent.

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

## 1.5 Optimal Mechanism

### *1.5.1 Simplifying the Problem*

Recall from section 1.2 that an optimal mechanism solves

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

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

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

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

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

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

$$\sigma_1(\mu) \ge \sigma^+(\mu) \quad \forall \mu \in [0,1]. \tag{Feasibility}$$

The feasibility constraint ensures that $(\sigma_1, \sigma^+, \sigma^-)$ corresponds to a mechanism $(\sigma_1, \sigma_2)$ with $\sigma_2 \le 1$.[12]

**Lemma 3.** *It is without loss of optimality to set $\sigma_1(\mu) = \sigma^+(\mu)$.*

*Proof.* Whenever $\sigma_1(\mu) > \sigma^+(\mu)$, reducing $\sigma_1(\mu)$ to $\sigma^+(\mu)$ relaxes the A-IR constraint, does not violate any other constraints, and does not affect the objective function. □

Lemma 3 implies that $\sigma_2(\mu, 1) = 1$. That is, 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 $\mu$. 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.[13]

By Lemma 3, $\sigma^+(\mu)$, which is the probability of promotion when the period-1 belief is $\mu$ and the state is 1, coincides with the probability of recommending "stay" in period 1 when the period-1 belief is $\mu$. Thus our goal is to find $\sigma^+(\mu)$ and $\sigma^-(\mu)$ that solve the following

12. 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 $\mu$.

13. One might be tempted by an alternative proof of Lemma 3 where, rather than decreasing $\sigma_1(\mu)$, we increase $\sigma^+(\mu)$ to $\sigma_1(\mu)$, i.e. increase $\sigma_2(\mu, 1)$ to 1. We took this approach in finding the contractible-optimal mechanism. However, this may violate P-IC$_1$.

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 4.** *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 $\sigma$ 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 since P-IC$_1$ implies $\sigma^+(1) \ge \sigma^+(\mu)$ for any $\mu \in [0,1]$. It is straightforward to check that $\tilde{\sigma}$ satisfies all the constraints of $\mathcal{P}$ and gives the principal a strictly higher ex ante payoff. Thus $\sigma$ 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 $\lim_{\mu \to 1} \sigma^+(\mu) = \sigma^+(1)$. □

### *1.5.2 Characterization*

To characterize optimal mechanisms, we start by considering two simple mechanisms that are incentive compatible for the principal. The first is a mechanism that promotes the agent regardless of the interim belief $\mu$ or the state $\theta$. This mechanism is clearly incentive compatible because it does not depend on the principal's reports.

**Definition 1.** Let $\sigma_1 : [0,1] \to [0,1]$ and $\sigma^- : [0,1] \to [0,1]$. The pair $(\sigma^+, \sigma^-)$ is an *always-promote mechanism* if $\sigma^+(\mu) = \sigma^-(\mu) = 1$ for all $\mu \in [0,1]$.

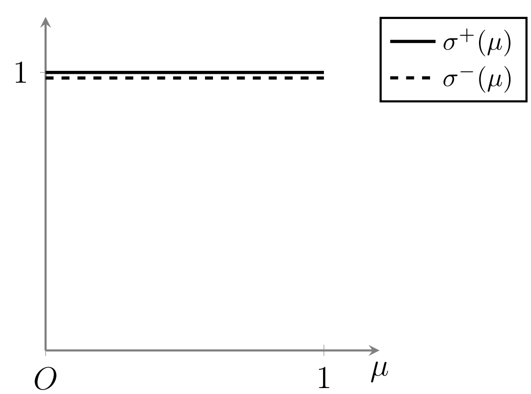


Figure 1.3: Always-Promote Mechanism

One may recall that, in Figure 1.2, 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 reason. First, we have shown that $\sigma_1(\mu) = \sigma^+(\mu)$

for all $\mu$. Moreover, 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 contractible-optimal mechanism, this can induce the agent to participate because the agent prefers to receive $c_1$ for sure rather than receive $b$ with a low probability. However, to satisfy the principal's interim incentive compatibility constraints, the mechanism must, upon keeping the agent in period 1, sometimes promote her even in the bad state.

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

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

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

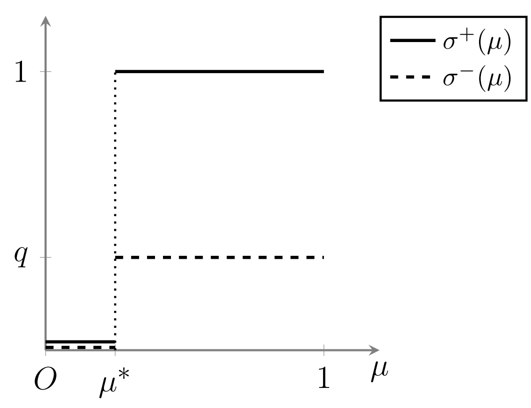


Figure 1.4: A Threshold Mechanism

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 $(\sigma^+, \sigma^-)$ satisfying the conditions of

Definition 2 as a "threshold pre-mechanism". Later in this section, we will construct an optimal mechanism mechanism by taking a convex combination (an operation we define shortly) over multiple threshold mechanisms. The resulting convex combination must be a mechanism, but the individual threshold mechanisms need not be.

A threshold mechanism is parametrized by the threshold $\mu^*$. 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 his belief is above $\mu^*$ is increasing in his true belief. Therefore, the mechanism is incentive compatible for the principal if and only if the principal is indifferent between his two options when his belief is equal to the threshold $\mu^*$. At $\mu^*$, 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 $\mu^*$, 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 $\mu_E$ and $q_E$, here $\mu^*$ and $q$ are *bundled*. To incentivize the principal to truthfully reveal whether his 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 $(\sigma_i^+, \sigma_i^-)$, $i = 1, \ldots, I$, we may define a new mechanism $(\sigma^+, \sigma^-)$ 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$_1$ is preserved under convex combination. Figure 1.5 illustrates a convex combination of a threshold mechanism and the always-promote mechanism.

Theorem 1 characterizes optimal mechanisms.

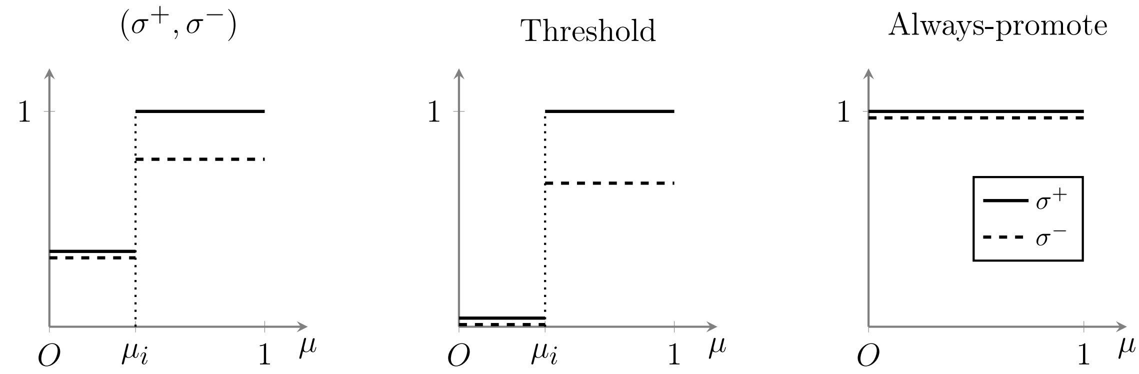


Figure 1.5: $(\sigma^+, \sigma^-)$ is a convex combination of threshold and always-promote.

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

*Sketch of proof.* See Appendix 1.11.3 for a formal proof; here, we provide a sketch. Recall from Section 1.5.1 that our problem $(\mathcal{P})$ is

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

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

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

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

Because P-IC$_1$ implies that $\sigma^+(\mu) - \sigma^-(\mu)$ is non-decreasing when $\mu \leq 1/2$ and non-increasing when $\mu \geq 1/2$, P-IC$_2$ is equivalent to $\sigma^+(0) \geq \sigma^-(0)$ and $\sigma^+(1) \geq \sigma^-(1)$. 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

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

This is reminiscent of the incentive compatibility constraints in a standard monopolist screening problem with a single good if we interpret $\mu$ as value, $\phi$ as allocation, and $\sigma^-$ as transfer. Although the principal cannot pay money to the agent, $\sigma^-$ serves as a way of transferring utility from the principal to the agent. By appealing to the envelope theorem, we may conclude 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 $\phi$ pins down $\sigma^-(\mu)$ up to a constant, and our problem can be reduced to that of choosing $\phi$ and $\sigma^-(0) \in [0,1]$.

If we fix $\sigma^-(0) \in [0,1]$, 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 $\phi(1) \geq 2\sigma^-(1)$. By the Bauer maximum principle, there exists $\phi$ 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 function $\phi$ that 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. As we show in the formal proof, this implies that we may restrict $\sigma^+$ and $\sigma^-$ to be non-decreasing step functions with at most three shared discontinuities. We can then prove that each mechanism $(\sigma^+, \sigma^-)$ is a convex combination of at most three threshold mechanisms and the always-promote mechanism, where the weight on the always-promote mechanism is $\sigma^-(0)$. □

Corollary 1, which follows from Theorem 1, explicitly describes the outcome of the convex combination. Figure 1.6 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* $\{p_0, p_1, p_2, 1\}$, *where* $0 \leq p_0 \leq p_1 \leq p_2 \leq 1$.

2. $\sigma^-(\mu)$ *is a non-decreasing step function taking values in* $\{p_0, q_1, q_2, 1\}$, *where* $p_0 \leq q_1 \leq q_2 \leq 1$

3. $\sigma^+(\mu) \geq \sigma^-(\mu)$ *for all* $\mu \in [0, 1]$.

4. $\sigma^+$ *and* $\sigma^-$ *share the same points of discontinuity.*

5. $\sigma^+(1) = 1$.

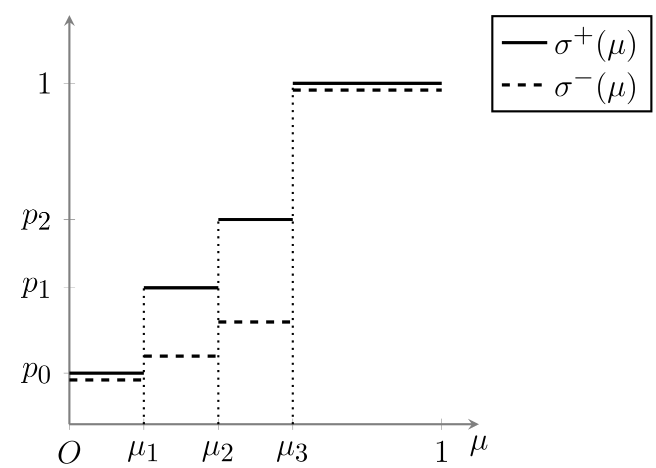


Figure 1.6: Optimal Mechanism

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 their belief belongs to. For example, a firm may evaluate worker 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.[14]

In light of Theorem 1, let us identify each of the three threshold mechanisms with the pair $(q_i, \mu_i)$ for $i = 1, 2, 3$. Let $p_0$ be the weight placed on the always- mechanism. We now describe how the optimal mechanism depends as the agent's ex ante outside option.

**Proposition 2** (General Comparative Statics). *There exists $\bar{c}_0 \in [\tilde{c}_0, b)$ such that the following statements are true.*

*(i) If $c_0 \in (b\mu_0, \bar{c}_0]$, 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 $c_0, c_0' \in (b\mu_0, \bar{c}_0]$ 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 $c_0 \in (\bar{c}_0, b]$, we can find an optimal mechanism that is a convex combination of the always-promote mechanism and three threshold mechanism 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 1.11.5. □

14. On the other hand, it is not without loss for the principal to *receive* a coarse signal in period 1. As we show in section 1.6, such a coarsening of the information structure relaxes the principal's incentive constraints in period 1 and may increase the principal's ex ante payoff under the optimal mechanism.

When the agent's ex ante outside option is low ($c_0 \leq \bar{c}_0$), 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 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 $\mu^*$ 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 most likely failing promotion and receiving 0, while the principal does not lose from preemptively letting the agent leave because he was unlikely to promote her even if she stayed. As a result, for low values of $c_0$, threshold mechanisms dominate the always-promote mechanism.

Once $c_0$ is above $\bar{c}_0$, the thresholds remain constant at $\bar{\mu}_i$, and the weight $p_0$ 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 her most of the time, 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 $\mu_i$-s fixed.

### *1.5.3 Optimal Mechanism: Single Threshold Case*

We 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. Define

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

$T(\mu^*, \lambda)$ is the sum of the principal's and the agent's payoffs from a threshold mechanism, where the agent's payoff is weighted by $\lambda \geq 0$.

Consider the following condition:

$$T(\mu^*, \lambda) \text{ is strictly concave in } \mu^* \text{ for any } \lambda \geq \lambda_0. \tag{1.3}$$

Intuitively, condition (1.3) 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 (1.3) 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 (1.3) always holds if the belief distribution $F$ is Uniform.[15]

15. For an alternative sufficient condition for (1.3) that does not require $f$ to be monotone, see Appendix 1.11.7.

**Theorem 2.** *Suppose condition* (1.3) *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 1.11.6. □

Define

$$\tilde{c}_0 := b \int_{c_1/2b}^{1} \left( \mu + (1-\mu)\frac{c_1}{2b - c_1} \right) \mathrm{d}F(\mu) + c_1 F(c_1/2b)$$

to be the ex ante payoff to the agent under the threshold mechanism with threshold $c_1/2b$. It is straightforward to verify that $\tilde{c}_0 \in (\hat{c}_0, \check{c}_0)$. The following proposition is an analogue of Proposition 2 for when condition (1.3) holds.

**Proposition 3** (Comparative Statics)**.** *Suppose condition* (1.3) *holds. Then, there exist* $\bar{c}_0 \in [\tilde{c}_0, \check{c}_0)$ *and* $\bar{\mu} \in (c_1/2b, 1/2]$ *such that:*

*(i) If* $c_0 \in (b\mu_0, \bar{c}_0]$*, there exists a unique mechanism that is a threshold mechanism and is optimal. The threshold satisfies* $\mu^* \leq \bar{\mu}$*, and* $\mu^*$ *is strictly increasing in* $c_0$*.*

*(ii) If* $c_0 \in (\bar{c}_0, b)$*, 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 1.11.8. □

The two cases of Proposition 3 are depicted in Figure 1.7. When the agent's ex ante outside option is low ($c_0 \leq \bar{c}_0$), the optimal mechanism is a threshold mechanism, and $\mu^*$ increases as $c_0$ increases. Once $c_0$ is above $\bar{c}_0$, $\mu^*$ stays fixed at $\bar{\mu}$, and the weight $p$ on the always-promote mechanism increases.

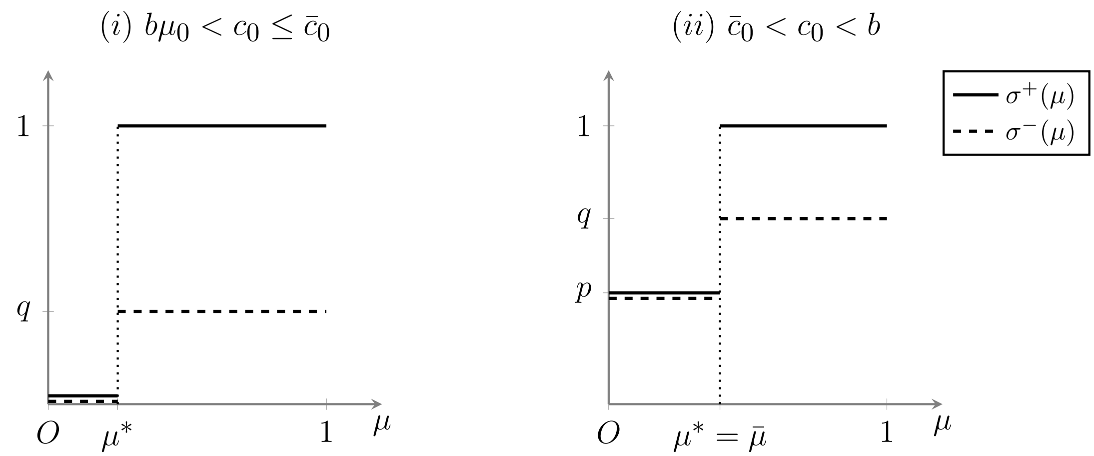


Figure 1.7: Optimal Mechanism When (1.3) Holds

## 1.6 Properties of the Optimal Mechanism

### *1.6.1 Interim Screening: Committing to Commit*

Proposition 2 tells us that, 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 $\mu^* > 0$.[16] The important feature of a threshold mechanism that it screens the agent based on the principal's interim beliefs, asking the agent to leave if the principal's belief is below the threshold. As was the case for the contractible-optimal mechanism, such interim screening incentivizes the agent to participate in the mechanism by reducing her 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 interim screening 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 his decision in period 2; in other words, the principal should commit to commit.

16. Even in case $(i)$, not all thresholds can equal 0, as otherwise the the agent's ex ante payoff from the mechanism would equal $b\mu_0 < c_0$.

It is crucial that the mechanism in period 1 not only restricts the period-2 promotion decision decision, but *communicates* this restriction to the agent. In our environment, the value of interim screening 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, he would have no reason to make any decision in period 1 with only partial information about the state.[17]

The following observation illustrates the importance of interim screening: 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 $\sigma^-(\mu) = 0$.[18] This means that the agent's benefit from being able to make a better interim decision more than outweighs her loss from a lower ex ante probability of promotion.

**Necessity of Interim Screening** Although we do not claim that the optimal mechanism characterized by Proposition 2 is the unique optimal mechanism, the following results imply that interim screening is a necessary component of any optimal mechanism.

**Lemma 5.** *If $\sigma$ is an optimal mechanism, it cannot be that $\sigma^+(0) > \sigma^-(0) > 0$.*

*Proof.* See Appendix 1.11.9. □

**Proposition 4.** *If $\sigma$ is an optimal mechanism, $\sigma^+$ cannot be constant in $\mu$.*

*Proof.* Suppose to the contrary that $\sigma^+$ is constant. By P-IC$_1$, $\sigma^-$ must be constant as well. By Lemma 4 and Lemma 5, the only two cases in which $\sigma^+$ and $\sigma^-$ may be constant is if

17. There may be other environments in which the principal benefits from committing to restrict his 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. distribution over decisions conditional on each state) that he could not implement without commitment.

18. This is the case, for example, if $b = 10$, $c_0 = 8.975$, $c_1 = 8$, and $f(\mu) = (-1/10)(\mu - 1/2) + 1$. We can easily check 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 $\int_0^1 \mu f(\mu)\mathrm{d}\mu \approx 0.4917$.

$\sigma^+ = 1$ and $\sigma^- = 0$ or if $\sigma^+ = \sigma^- = 1$. The former is ruled out by our assumption that $c_0 > b\mu_0$. The latter is ruled out by our assumption that $c_0 < b$. □

By Proposition 4, in any optimal mechanism, it cannot be that $\sigma^+(\mu) = 1$ for all $\mu \in [0, 1]$. This means that in any optimal mechanism, the agent must sometimes be asked to leave in period 1, i.e. there must be interim screening.

### *1.6.2 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 her in period 2 even when the principal already knows in period 1 that the state is bad, i.e. $\mu = 0$.[19]

On the other hand, the contractible-optimal mechanism always asks the agent to leave if the principal's interim belief is sufficiently low, $\mu \in [0, \mu_E)$. That is, the contractible-optimal mechanism never ignores the principal's information. 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.

### *1.6.3 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 being

19. Note that the probability that the agent stays in period 1 is constant at $p_0$ 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 $\mu$ and sometimes leaves when the belief is $\mu'$. It would be more efficient to leave more often $\mu$ and stay more often at $\mu'$, but such Pareto improvements are not achievable because of the principal's incentive constraints.

$\theta$. In the optimal mechanism, does $\sigma_2(\mu, \theta)$ depend on $\mu$? One might have answered in the negative, since $\mu$ is payoff-irrelevant given $\theta$, but our characterization of the optimal mechanism tells otherwise. Although we know from Lemma 3 that $\sigma(\mu, 1)$ is constant at 1, $\sigma(\mu, -1)$ can depend non-trivially on $\mu$. For instance, if condition (1.3) holds and $c_0$ is large (statement *(ii)* of Proposition 3), we have $\sigma_2(\mu, -1) = 1$ for $\mu < \bar{\mu}$ and $\sigma_2(\mu, -1) = p < 1$ for $\mu \geq \bar{\mu}$. In other words, when making the promotion decision in period 2, the optimal mechanism must *remember* what the principal reported in period 1. 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).

## 1.7 Implementation

How would a firm deciding whether to promote a worker actually implement the optimal mechanism? When $\sigma_2(\mu, -1)$ is decreasing in $\mu$, if the worker turns out to be unproductive in period 2, she is *less* likely to be promoted if the firm had a *higher* belief about her value in the past. One might worry that this feature makes implementation difficult, but the firm can implement the optimal mechanism in a natural way that mirrors how the mechanism is constructed. Theorem Theorem 1 implies that the optimal 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 $\mu$ under any convex combination of threshold mechanisms. To see this, suppose we place weight $k_i \in (0,1)$ on a threshold mechanism with threshold $\mu_i$, for $i = 1,2,3$. Let $\mu_1 \le \mu_2 \le \mu_3$. 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 $\mu$ on $\mu \in [0,1]$.[20]

Therefore, we may interpret the convex combination of threshold mechanisms as a promotion scheme that consists of a midterm review and a final review. During the midterm review, which takes place in period 1, the firm forms a belief about whether the worker is productive. The probability that the worker passes the midterm review is a non-decreasing step function of the firm's belief.[21] The worker is asked to stay with the firm if she passes the midterm; otherwise, she is no longer considered for promotion and is asked to leave. If the worker obeys the recommendation to stay, she is reviewed again in period 2. During this final review, the firm observes whether the worker is productive. If the worker is productive, she is promoted with certainty. Even if she is not productive, she 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. Thus, to implement the optimal mechanism, the firm commits to review the worker only some of the time. With probability $1-p$, a review takes

20. 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.

21. Such randomization may be implemented, for example, 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.

place, and the worker is promoted if she passes both a midterm and a final review. With probability $p$, the worker is promoted by default.

Can the worker know whether she is subject to a review or not? 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 she will be reviewed in the future; otherwise, the worker will leave if she knows that she 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 she was subject to passed the midterm review, her conditional probability of being promoted decreases, and this may lead her 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 she has been reviewed. For example, if the review consists of an interview, the firm may need to nominally interview the worker even if she will be promoted by default.

However, the threshold mechanism may be sufficiently attractive that the worker's interim obedience holds even if she learns that she has been reviewed. 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

$$\begin{aligned}
&\frac{b}{1-F(\mu^*)}\int_{\mu^*}^{1}\left(\mu+(1-\mu)\frac{\mu^*}{1-\mu^*}\right)\mathrm{d}F(\mu) \\
>&\frac{b}{1-F(\mu^*)}\int_{\mu^*}^{1}\left(\mu^*+(1-\mu^*)\frac{\mu^*}{1-\mu^*}\right)\mathrm{d}F(\mu) \\
=&\,2b\mu^* \\
>&\,c_1.
\end{aligned} \tag{1.4}$$

(1.4) shows that if a worker knows that she is playing a threshold mechanism with $\mu^* > c_1/2b$, and she has been recommended to stay, she 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 recommendation to stay even if she learns that she is playing the convex combination of the threshold mechanisms.

In fact, (1.4) implies an even stronger result. Suppose the worker is asked to stay in period 1 and know that she is subject to reviews. Furthermore, suppose the worker becomes aware that she barely passed the midterm review. That is, the firm's belief about her was $\bar{\mu}$, which is the lowest possible belief under which she is asked to stay. Even then, the worker will be willing to stay because her expected payoff of staying. Our interpretation is that 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 conveyed their pessimism to the worker, she would still choose to stay when asked to do so.

**Proposition 5** (No Leading On)**.** *Suppose $\sigma$ 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 she knows the principal's interim belief $\mu$ and knows that the convex combination of threshold mechanisms is being played.*

The thresholds are greater than $c_1/2b$ if, for example, condition (1.3) holds and $c_0 \geq \bar{c}_0$ (Proposition 3 (*ii*)).

## 1.8 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.

### 1.8.1 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, she receives a benefit of $be + d$, where $d \geq 0$. The worker receives 0 regardless of her choice of $e$ if she does not maintain the investment or if she is not promoted in period 2. Neither the worker's choice of $e$ in period 0, nor her 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 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 she believes the firm is likely to promote her. 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 her with a high probability so that the worker is more likely to benefit from her investment, and by committing to let the worker know in advance, in period 1, if she is unlikely to be promoted, so that she may avoid paying the cost of maintaining her investment.

If the worker chooses to invest a strictly positive amount $e > 0$, she will obey the firm's recommendation to maintain the investment in period 1, as otherwise, she 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 $c_0 = \bar{e} + k$.

Note that even if we allow the firm to make monetary transfers to the worker, the firm cannot incentivize the worker to invest by paying her. Because the worker's level of investment is noncontractible, transfers would not affect the worker's marginal benefit of investing.

### *1.8.2 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 they expect 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 $\theta x$, 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), they receive 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 how the regulator should provide *forward guidance* about their 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 provide additional information in period 1. By committing to reduce uncertainty for the company in period 1, the regulator can induce the company to invest in period 0. Intuitively, when they invest in period 0, the company purchases a real option which allows them to either shut down or invest once more in period 1. By committing to provide information in period 1 about the period-2 subsidy, the regulator increases the value of the real option to the company and induces them to purchase it.

## 1.9 Additional Results and Extensions

### *1.9.1 Agent-Optimal Mechanism: Tell Me Tomorrow*

Suppose the principal has an ex ante outside option of terminating his 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 firm, who receives private, noncontractible information. Although the agent always wants to be promoted, she 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* in 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 their future decision.

The problem of finding the agent-optimal mechanism subject to the principal's participation constraint is the dual of our original problem of finding the principal-optimal mechanism subject to the agent's participation constraint.[22] Since the problem is linear in the mechanism $\sigma = (\sigma^+, \sigma^-)$, we may invoke strong duality. If a mechanism $\sigma$ 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 $\sigma$ 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 $\sigma$ equals $c_0$. Thus the agent asks the principal not only to sometimes promote her against his wishes, but also to inform the agent in period 1 about her chance of promotion in period 2.

### *1.9.2 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 to equally subsidize all companies in an industry.

22. Both problems are also subject to the principal's incentive compatibility constraints.

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.[23] What mechanisms can the principal implement? The answer to this question depends crucially on the distribution of the state $\theta$. First, it may be that there is a continuum of agents that are only *ex ante* homogeneous, and both $\mu$ and $\theta$ are drawn independently and identically for each agent. This would be the case, for example, if $\theta$ represents the innate ability of each worker. On the other hand, it may be that the agents are *ex post* homogeneous, so that a single $\mu$ is drawn in period 1 and a single $\theta$ is drawn in period 2. For example, $\theta$ may represent the demand for the firm's goods and thus be 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 $c_0 \in [\hat{c}_0, \check{c}_0]$ 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 $m_1 \coloneqq 1 - F(c_1/2b)$, and the measure of agents who are promoted is

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

Suppose the principal commits to recommend "stay" to $m_1$ agents and promote $m_2$ agents, and suppose he 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 he is the most optimistic

23. 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.

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 his commitment to the the marginal distributions.

Next, suppose the agents are ex post homogeneous. Since only one $\mu$ 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 (1.3) holds and $c_0 \in (\bar{c}_0, b)$. 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.[24] 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 5 describes the condition under which the agents obey the recommendation to stay even if they become aware of such information.

### *1.9.3 Comparative Statics on Belief Distribution*

The principal is better off when he is ex ante more optimistic about the state.

**Proposition 6.** *Let $F, F'$ be two belief distributions with supp$(F) =$ supp$(F') = [0, 1]$ and $\mathbb{E}_F[\mu] = \mathbb{E}_{F'}[\mu] = \mu_0$. If $F$ first-order stochastic dominates $F'$, then the principal obtains a higher ex ante payoff under $F$ than $F'$.*

*Proof.* Given a mechanism $\sigma$, the partial derivative of the principal's interim payoff with

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

respect to interim belief $\mu$ at $\mu \neq \mu^*$ is

$$\begin{aligned} &\sigma^+(\mu) + \sigma^-(\mu) + \mu(\sigma^+)'(\mu) - (1-\mu)(\sigma^-)'(\mu) \\ =&\sigma^+(\mu) + \sigma^-(\mu) \\ \geq& 0. \end{aligned}$$

Also, the principal's interim payoff is continuous at $\mu^*$. Therefore, the principal's interim payoff is increasing in $\mu$. Likewise, the agent's interim payoff is increasing in $\mu$. Therefore, if $\sigma'$ is an optimal mechanism under $F'$, the principal with belief distribution $F$ can choose $\sigma'$ and do no worse than under $F'$. □

Does the principal always prefer a more informative interim belief distribution $F$? It is clear that no $F$ can be strictly worse for the firm than the uninformative distribution $\underline{F} = \delta_{\mu_0}$, since the firm 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 $\bar{F}(\mu)$.

However, the principal sometimes prefers to receive a *less* informative signal in period 1. 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 1.8, and the principal's expected ex ante payoff is 0.156.

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

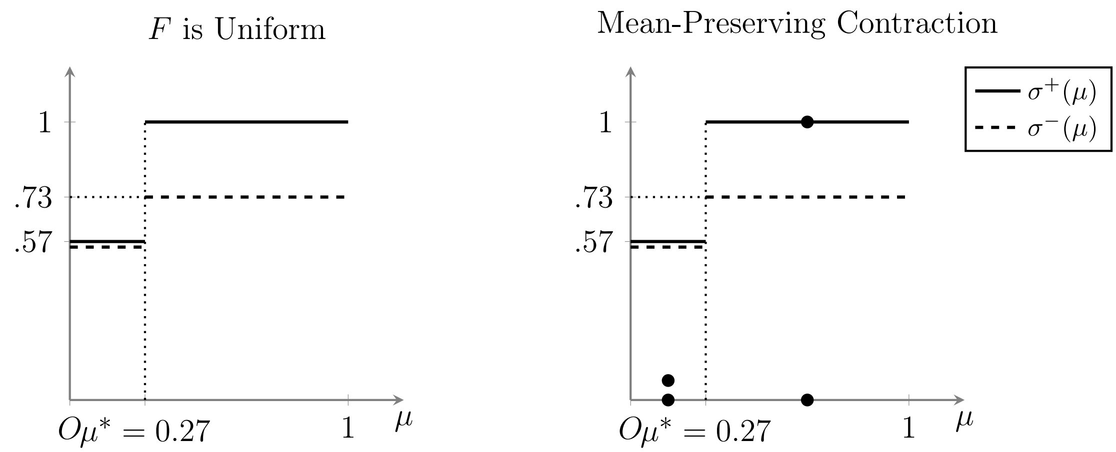


Figure 1.8: Effect of Mean Preserving Contraction ($b = 15$, $c_0 = 13$, $c_1 = 7$)

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 his period-1 belief, learning less in period 1 can help him 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.

### *1.9.4 Deterministic Mechanisms*

The optimal mechanism generally involves 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 1.9. 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 $c_0 \in (c_1F(1/2)+b(1-F(1/2), b)$. Thus the restriction to deterministic mechanisms interacts with 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 $c_0 = c_1F(1/2)+b(1-F(1/2)$.[25][26]

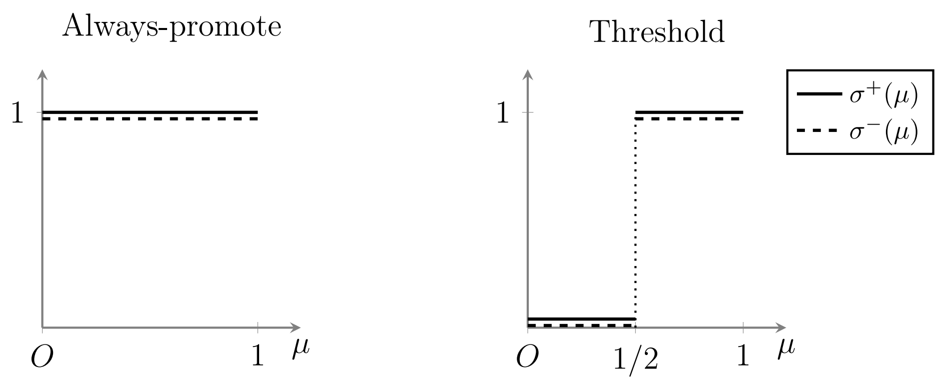


Figure 1.9: Two Deterministic Mechanisms

In contrast, when the principal's signals are contractible, Proposition 1 shows that there exists a deterministic contractible-optimal mechanism as long as $c_0 \in (b\mu_0, \hat{c}_0] \cup c_0 \in [\check{c}_0, b)$. In addition, if signals are contractible, it is without loss for interim recommendations to be deterministic regardless of the value of $c_0$.

## 1.10 Conclusion

This paper studies the problem of a principal who must make a decision in the future, gradually receives private information about his payoffs from the decision, and faces an agent who wants know what the principal will do. This problem is not uncommon—workers

25. 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.

26. 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.

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 his decision, the principal commits today to commit tomorrow. Because his private information cannot be contracted on, the principal sometimes ignores his information and decides in the agent's favor.

## 1.11 Omitted Proofs

### 1.11.1 *Proof of Lemma 2*

Fix $q \in [0,1]$ and consider the problem

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

Letting $\sigma_1 \in L^1[0,1]$, the problem has a solution $\sigma_1^*$. Since both the objective and the constraint are affine in $\sigma_1$, there must exist a Lagrangian multiplier[27] $\lambda \ge 0$ such that choosing $\sigma_1 = \sigma_1^*$ maximizes

$$\begin{aligned} &\int_0^1 \sigma_1(\mu)(\mu - (1-\mu)q)\mathrm{d}F(\mu) \\ &+ \lambda\left(b\int_0^1 \sigma_1(\mu)(\mu + (1-\mu)q)\mathrm{d}F(\mu) + c_1 \int_0^1 (1-\sigma_1(\mu))\mathrm{d}F(\mu) - c_0\right). \end{aligned}$$

Therefore, it must be that $\sigma_1(\mu) = 1$ if $\mu \ge \mu_E := \frac{q+\lambda(bq-c_1)}{1+q+\lambda b(1-q)}$ and $\sigma_1(\mu) = 0$ if otherwise.

### 1.11.2 *Proof of Proposition 1*

The problem of maximizing the Lagrangian is

$$\max_{\substack{q\in[0,1] \\ \mu\in[0,1]}} \int_{\mu_E}^1 (\mu - (1-\mu)q)\,\mathrm{d}F(\mu) + \lambda\left(b\int_{\mu_E}^1 (\mu + (1-\mu)q)\,\mathrm{d}F(\mu) + c_1 F(\mu_E) - c_0\right).$$

If we find $q_E$ and $\mu_E$ that maximizes the Lagrangian and satisfies the A-IR constraint, then $(q_E, \mu_E)$ is an efficient mechanism. Conversely, the statement and the proof of Lemma 2

27. See, for example, Luenberger [1997].

implies that given an efficient mechanism $(q_E, \mu_E)$, there exists a multiplier $\lambda \geq 0$ such that $q_E$ and $\mu_E$ maximize the Lagrangian.

If the A-IR constraint does not bind, $\lambda = \mu_E = q_E = 0$ solves the Lagrangian. Suppose that the constraint binds, so that $\lambda > 0$.

The first derivatives of the Lagrangian with respect to $q_E$ and $\mu_E$ are, respectively,

$$(\lambda b - 1) \int_{\mu_E}^{1} (1-\mu)\, \mathrm{d}F(\mu) \tag{1.5}$$

$$(\mu_E(\lambda b q_E - \lambda b - q_E - 1) + q_E - \lambda b q_E + \lambda c_1) f(\mu_E). \tag{1.6}$$

First, suppose $\lambda > 1/b$. Then, setting $q_E = 1$ and

$$\mu_E = \max\{0, \frac{1 + \lambda(c_1 - b)}{2}\}$$

maximizes the Lagrangian. $(q_E, \mu_E)$ solves the problem when the constraint holds with equality, i.e.

$$b(1 - F(\mu_E)) + c_1 G(\mu_E) = c_0. \tag{1.7}$$

This corresponds to case $(iii)$.

Next, if $\lambda < 1/b$, setting $q_E = 0$ and

$$\mu_E = \frac{\lambda c_1}{1 + \lambda b}$$

maximizes the Lagrangian. $(q_E, \mu_E)$ solves the problem when the constraint holds with equality, i.e.

$$b \int_{\mu_E}^{1} \mu\, \mathrm{d}F(\mu) + c_1 F(\mu_E) - c_0 = 0. \tag{1.8}$$

This corresponds to case $(i)$.

Finally, suppose $\lambda = 1/b$. Then,

$$\mu_E = \frac{c_1}{2b}$$

and any value $q_E \in [0,1]$ maximizes the Lagrangian. $(q_E, \mu_E)$ solves the problem when the constraint holds with equality, i.e.

$$b \int_{c_1/2b}^{1} (\mu + (1-\mu)q) \, \mathrm{d}F(x) + c_1 F(c_1/2b) - c_0 = 0. \tag{1.9}$$

This corresponds to case $(ii)$.

### *1.11.3 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

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

$\phi(\mu)$ is non-decreasing.

The problem ($\mathcal{P}$) is to choose $\phi : [0,1] \to [0, \sigma^-(1)+1]$ and $\sigma^- : [0,1] \to [0,1]$ to solve

$$
\begin{aligned}
&\max \int_0^1 (\mu\phi(\mu) - \sigma^-(\mu))\,\mathrm{d}F(\mu) \\
\text{s.t.}\quad & c_0 \le b\int_0^1 (\mu\phi(\mu) + (1-2\mu)\sigma^-(\mu))\,\mathrm{d}F(\mu) + c_1\int_0^1 (1-\phi(\mu)+\sigma^-(\mu))\,\mathrm{d}F(\mu) && \text{(A-IR)} \\
& \phi(\mu) \text{ is non-decreasing} && (\text{P-IC}_1^b) \\
& \phi(\mu) \ge 2\sigma^-(\mu) \quad \text{for } \mu = 0,1, && (\text{P-IC}_2) \\
& \sigma^-(\mu) = \sigma^-(0) + \mu\phi(\mu) - \int_0^\mu \phi(x)\,\mathrm{d}x \quad \forall \mu \in [0,1]. && (\text{P-IC}_1^a)
\end{aligned}
$$

Since $\phi$ is non-decreasing, without loss of generality, we may restrict attention to $\phi$ (and thus $\sigma^-$) that is right-continuous on $\mu \in [0,1]$.

Fix $\sigma^-(0), \sigma^-(1) \in [0,1]$ such that $\sigma^-(0) \le \sigma^-(1)$. 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 the feasible set $\mathcal{F} \subset \mathcal{M}$ be the subset of functions in $\mathcal{M}$ that satisfy A-IR and[28]

$$
\begin{aligned}
&\sigma^-(1) = \sigma^-(0) + \phi(1) - \int_0^1 \phi(x)\,\mathrm{d}x \\
\Leftrightarrow &\int_0^1 \phi(x)\,\mathrm{d}x = \sigma^-(0) + 1. \qquad (1.10)
\end{aligned}
$$

Thus $\mathcal{F}$ is the subset of $\mathcal{M}$ that satisfies two linear constraints, A-IR and (1.10).

By Helly's selection principle[29], 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.

28. All other constraints have been built into the definition of $\mathcal{M}$.

29. See, for example, Kolmogorov and Fomin [1975].

It is well known that the set $E \coloneqq \{e : [0,1] \to \{2\sigma^-(0), \sigma^-(1)+1\} \mid e \text{ is non-decreasing}\}$ 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. Moreover, the objective function is affine in $\phi$. 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 a constrained-optimal mechanism such that $\phi$ is a convex combination of at most three elements of $E$.

Now, let us unwrap $(\phi, \sigma^-)$ back into $(\sigma^+, \sigma^-)$. Keeping $\sigma^-(0)$ and $\sigma^-(1)$ fixed, consider a constrained-optimal mechanism. Let $e_1, e_2, e_3$ be the three elements of $E$ that constitute the $\phi$ of this constrained-optimal mechanism. Let $\mu_i$ denote the point at which $e_i$ is discontinuous.

By P-IC$_1^a$, $\sigma^-$ is also non-decreasing and is a convex combination of at most three functions $e_1^-, e_2^-, e_3^-$ contained in $E^- \coloneqq \{e^- : [0,1] \to \{\sigma^-(0), \sigma^-(1)\} \mid e^- \text{ is non-decreasing}\}$. We index each $e_i^-$ so that it shares the same discontinuity as $e_i$. Likewise, $\sigma^+$ is a convex combination of at most three functions $e_1^+, e_2^+, e_3^+$ contained in $E^+ \coloneqq \{e^+ : [0,1] \to \{\sigma^-(0), 1\} \mid e^+ \text{ is non-decreasing}\}$, 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 $e(x) = \sigma^-(0)$.[30] This means that each $(e_i^+, e_i^-)$ is the convex combination of a threshold mechanism with threshold $\mu_i$ 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

30. In an optimal mechanism, the function $\sigma^+$ can never be written as a convex combination that places a positive weight on $e(x) = \sigma^-(0) < 1$, as this would imply that $\sigma^+$ is bounded away from 1, contradicting Lemma 4.

a convex combination of three threshold mechanisms and the always-promote mechanism. Of course, any mechanism that is a convex combination of three threshold mechanisms and the always-promote mechanism is a convex combination of three elements of $E^+ \times E^-$ for some value of $\sigma^-(0)$ and $\sigma^-(1)$.

We have so far shown that, given each $\sigma^-(0)$ and $\sigma^-(1)$, there exists a convex combination of three threshold mechanisms and the always-promote mechanism that solves the principal's constrained problem. It remains to prove that there exists a solution of this form to the unconstrained problem. For this, it is enough to show the existence of a mechanism which is a convex combination of three threshold mechanisms and the always-promote mechanism, and which solves the principal's unconstrained problem across all mechanisms that are a convex combination of three threshold mechanisms and the always-promote mechanism. We defer this to Lemma 8 in subsection 1.11.4.

### *1.11.4 Lagrangian Approach*

This subsection describes the Lagrangian approach to solving the problem $(\mathcal{P})$.

By Theorem 1, $(\mathcal{P})$ can be reduced to the choice of a 7-tuple $\mathbb{x} = (\mu_1, \mu_2, \mu_3, k_1, k_2, k_3, k_4)$, representing the three thresholds, $\mu_1, \mu_2, \mu_3 \in [0, 1)$, and a weight $k_i$ for each of the four mechanisms such that $\sum_{i=1}^4 k_i = 1$, to maximize the principal's expected payoff subject to the individual rationality constraint and $\sigma^+(1) \geq \sigma^-(1)$.[31] Define

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

$$\mathbb{X}_2 = \{\mathbb{x} \in [0,1)^3 \times [0,1]^4 \mid \sum_{i=1}^{3} k_i = 1\}.$$

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

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

$$\begin{aligned} t(\mu_i, \lambda, \eta) &:= \int_{\mu_i}^{1} \left(\mu - (1-\mu)\frac{\mu_i}{1-\mu_i}\right) \mathrm{d}F(x) \\ &\quad + \lambda \left(b \int_{\mu_i}^{1} \left(\mu + (1-\mu)\frac{\mu_i}{1-\mu_i})\right) \mathrm{d}F(x) + c_1 F(\mu_i) - c_0\right) \\ &\quad + \eta \left(1 - \frac{\mu_i}{1-\mu_i}\right) \\ a(\lambda) &:= \int_{0}^{1} (\mu - (1-\mu)) \, \mathrm{d}F(x) + \lambda(b - c_0) \, . \end{aligned}$$

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 $\mu_i$. $a(\lambda)$ is the weighted sum of payoffs induced by the always-promote mechanism.

The Lagrangians are

$$\begin{aligned} \mathcal{L}_1 = \mathcal{L}(\mathbb{x}, \lambda, 0) &= \sum_{i=1}^{3} k_i t(\mu_i, \lambda, 0) + k_4 a(\lambda) \\ \mathcal{L}_2 = \mathcal{L}(\mathbb{x}, \lambda, \eta) &= \sum_{i=1}^{3} k_i t(\mu_i, \lambda, \eta) + k_4 a(\lambda) \, . \end{aligned}$$

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

$$b \int_{\mu_i}^{1} \left(\mu + (1-\mu)\frac{\mu_i}{1-\mu_i})\right) \mathrm{d}F(\mu_i) + c_1 F(\mu_i) - c_0 \geq 0$$

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 of the statement holds here because the problem $(\mathcal{P})$ is linear in $(\sigma^+, \sigma^-)$.[32]

Similarly, a vector $\mathbb{x}_2 \in \mathbb{X}_2$ solves the $(\mathcal{P})$ if and only if there exist $\lambda, \eta \geq 0$ such that $\mathbb{x}_2$ solves $\max_{\mathbb{X}_2} \mathcal{L}_2$, 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 the both Lagrangians to prove our results. The following lemma provides a justification for this.

**Lemma 6.** *Let $\lambda \geq 0$. The following are equivalent.*

*(i) $\mathbb{x}$ maximizes $\mathcal{L}_1 = \mathcal{L}(\mathbb{x}, \lambda, 0)$ across $\mathbb{X}_1$, satisfies A-IR, and satisfies complementary slackness—either $\lambda = 0$, or A-IR binds.*

*(ii) There exists $\eta \geq 0$ such that $\mathbb{x}$ maximizes $\mathcal{L}_2 = \mathcal{L}(\mathbb{x}, \lambda, \eta)$ across $\mathbb{X}_2$, satisfies A-IR and $\sum_{i=1}^3 k_i \frac{\mu_i}{1-\mu_i} \leq 1$, and satisfies complementary slackness—either $\lambda = 0$, or A-IR holds with equality; and either $\eta = 0$, or $\sum_{i=1}^3 k_i \frac{\mu_i}{1-\mu_i} = 1$.*

*Proof.* Interpret $\mathcal{L}_1 = \sum_{i=1}^3 k_i t(\mu_i, \lambda, 0) + k_4 a(\lambda)$ as the objective function and $\sum_{i=1}^3 \frac{\mu_i}{1-\mu_i} \leq 1$ as the single constraint. $(ii)$ implies $(i)$ by the Lagrangian sufficiency theorem. Conversely, $(i)$ implies $(ii)$ by the Lagrangian necessity theorem, which holds because the problem of maximizing $\mathcal{L}_1$ across $\mathbb{X}_1$ can be recast as the linear problem of choosing a mechanism $(\sigma^+, \sigma^-)$. □

The Lagrangian necessity theorem tells us that if $\mathbb{x}$ solves the $(\mathcal{P})$, there exists $\lambda_1$ such that $(i)$ is true, and that there exists $\lambda_2$ such that $(ii)$ is true. Lemma 6 shows that we may take the two $\lambda$-s to be equal. We next show that we can choose each $\mu_i$ from a compact interval.

**Lemma 7.** *There exists $\epsilon > 0$, dependent on $b$, $c_1$, and $F$ but independent of $c_0$, such that any solution $\mathbb{x}$ to the problem $(\mathcal{P})$ must satisfy $\mu_i \in [0, 1-\epsilon]$ for each $i = 1, 2, 3$.*

32. See, for example, Luenberger [1997].

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

$$\frac{\lambda b - 1}{(1-\mu_i)^2} \int_{\mu_i}^{1} (1-\mu)\,\mathrm{d}F(\mu) - \lambda f(\mu_i)(2b\mu_i - c_1) - \frac{\eta}{(1-\mu_i)^2}. \tag{1.11}$$

This can be rewritten as

$$\frac{-1}{(1-\mu_i)^2} \int_{\mu_i}^{1} (1-\mu)\,\mathrm{d}F(\mu) + \lambda \left( \frac{b}{(1-\mu_i)^2} \int_{\mu_i}^{1} (1-\mu)\,\mathrm{d}F(\mu) - f(\mu_i)(2b\mu_i - c_1) \right) - \frac{\eta}{(1-\mu_i)^2}. \tag{1.12}$$

L'Hopital's Rule and the continuity of $f$ at $\mu = 1$ imply that

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

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

Both are strictly negative since $f(1) > 0$. Therefore, there exists $\epsilon > 0$, independent of $\lambda$ and $\eta$, such that for any $\mu_i \in (1-\epsilon, 1)$, we have (1.11) $< 0$. Thus it can never be optimal to choose $\mu_i > 1 - \epsilon$. □

In light of Lemma 7, we now restrict attention to

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

$$\bar{\mathbb{X}}_2 = \{\mathbb{x} \in [0, 1-\epsilon]^3 \times [0,1]^4 \mid \sum_{i=1}^{3} k_i = 1\}.$$

Lemma 7 allows us to complete the proof of Theorem 1 by proving the following lemma.

**Lemma 8.** *The restriction of the problem* $(\mathcal{P})$ *to* $\bar{\mathbb{X}}_2$ *has a solution.*

*Proof.* $\bar{\mathbb{X}}_2$ is compact. All three terms in $t(\mu_i, \lambda, \eta)$ are continuous in $\mu_i$. Thus the objective function, which is a convex combination of the first term of $t(\mu_i, \lambda, \eta)$ and the first term of

$a(\lambda)$, is continuous in $x \in \bar{\mathbb{X}}_2$. Similarly, the two constraints, A-IR and P-IC$_2$, are continuous in $x \in \bar{\mathbb{X}}_2$. Thus there exists $x \in \bar{\mathbb{X}}_2$ that solves the restriction of the problem $(\mathcal{P})$ to $\bar{\mathbb{X}}_2$. □

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 9.** *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* $k_4 = 0$.

(*ii*) *For* $\lambda = \bar{\lambda}$, *there exists* $\mathbb{x}_1 \in \bar{\mathbb{X}}_1$ *with* $k_4 > 0$ *that maximizes* $\mathcal{L}_1$.

*Proof.* Consider the problem of maximizing the first Lagrangian $\mathcal{L}_1$ across $\mathbb{x}_1 \in \bar{\mathbb{X}}_1$. Any solution $\mathbb{x}_1$ has $k_i > 0$ only if the corresponding $t(\mu_i, \lambda, 0)$ or $a(\lambda)$ has the largest value among the four terms of the Lagrangian. Moreover, if $k_4 < 1$, then $\mathbb{x}_1$ must solve

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

It is easy to see that $V(0) > a(0)$. We now argue that when $\lambda$ is large enough, we have $V(\lambda) < a(\lambda)$. Consider the difference between the ex ante payoffs to the agent under a mechanism with $k_4 = 1$ and the the maximum ex ante payoff that can be given to the agent under a mechanism with $k_4 = 0$:

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

It is clear that $D$ is non-negative, since the agent cannot do better than obtaining a payoff of $b$ with probability 1. Moreover, we can show that $D$ is bounded away from zero:

$$\begin{aligned} D &\geq bF(\frac{1}{2}) - \max_{\substack{\mathbb{x}\in\bar{\mathbb{X}}_1 \\ \mu_i\leq 1/2}} \sum_{i=1}^{3} k_i \left( b\int_{\mu_i}^{1/2} \left(\mu + (1-\mu)\frac{\mu_i}{1-\mu_i})\right) \mathrm{d}F(\mu) + c_1 F(\mu_i) \right) \\ &= \min_{\mu_1\leq 1/2} (b-c_1)F(\mu_1) + b\int_{\mu_1}^{1/2} (1-\mu)\frac{1-2\mu_1}{1-\mu_1}\mathrm{d}F(\mu) \\ &\geq \min\{b\int_{1/4}^{1/2} \frac{2}{3}(1-\mu)\mathrm{d}F(\mu),\, (b-c_1)F(\frac{1}{4})\} > 0\,. \end{aligned}$$

The first inequality holds because the RHS minimizes the difference between the ex ante payoffs conditional on $\mu \leq 1/2$, and because the difference is positive at every $\mu$. The equality holds because we may take all $\mu_i$'s to be equal.

Also, the difference between the maximum ex ante payoff that can be given to the principal under a mechanism with $k_4 = 0$ and the ex ante payoff to the principal under a mechanism with $k_4 = 1$ 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 $\lambda$ is large enough, we have $V(\lambda) < a(\lambda)$. Intuitively, as we place an increasingly large weight on the agent's payoff, it must eventually be that the always-promote mechanism is used.

By the Maximum Theorem, $V(\lambda)$ is continuous in $\lambda$. Clearly, $a(\lambda)$ is also continuous in $\lambda$. Hence there exists a smallest $\lambda$, strictly greater than 0, such that $V(\lambda) = a(\lambda)$; take $\bar{\lambda}$ to be this number. For $\lambda < \bar{\lambda}$, we have $V(\lambda) > a(\lambda)$, so any optimal $\mathbb{x}$ must have $k_4 = 0$. When $\lambda = \bar{\lambda}$, we can maximize the Lagrangian with any value of $k_4$. □

Finally, we state a basic property of the Lagrangian.

**Lemma 10.** *Fix $b, c_1$, and $F$. Let $\lambda$ and $\lambda'$ be two Lagrangian multipliers such that $\lambda > \lambda'$. Suppose $\mathbb{x} = (\mu_1, \mu_2, \mu_3, k_1, k_2, k_3, k_4)$ maximizes $\mathcal{L}(\mathbb{x}, \lambda, 0)$ across $\mathbb{X}_1$, and $\mathbb{x}' = (\mu_1', \mu_2', \mu_3', k_1', k_2', k_3', k_4')$ maximizes $\mathcal{L}(\mathbb{x}, \lambda', 0)$ across $\mathbb{X}_1$. Then $\mathbb{x}$ must induce a weakly* *higher ex ante payoff to the agent than does* $\mathbb{x}'$*:*

$$\sum_{i=1}^{4} k_i \left( b \int_{\mu_i}^{1} \left( \mu + (1-\mu)\frac{\mu_i}{1-\mu_i}) \right) \mathrm{d}F(\mu) + c_1 F(\mu_i) \right)$$
$$\geq \sum_{i=1}^{4} k'_i \left( b \int_{\mu'_i}^{1} \left( \mu + (1-\mu)\frac{\mu'_i}{1-\mu'_i}) \right) \mathrm{d}F(\mu) + c_1 F(\mu'_i) \right) .$$

### *1.11.5 Proof of Proposition 2*

We prove $(i)$ using Lemma 9. Consider the second Lagrangian $\mathcal{L}_2$, and suppose $\lambda \leq 1/b$. The partial derivative (1.11) is always strictly negative when $\mu_i > c_1/2b$, so if $\mathcal{L}_2$ is maximized at some $\mathbb{x}$, it must be that $\eta = 0$ and that $\mathbb{x}$ has $\mu_i \leq c_1/2b < 1/2$ for all $i$.

For $\lambda = 1/b$ and $\eta = 0$, setting $\mu = c_1/2b$ gives us $t(\mu, 1/b, 0) > a(1/b)$, which implies $V(1/b) > a(1/b)$. However, for $\mu = c_1/2b$, the coefficient of $\lambda$ in $a(\mu, \lambda, \eta)$,

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

is always smaller than $b - c_0$, which is the coefficient of $\lambda$ in $a(\lambda)$. Thus $t(c_1/2b, \lambda, 0) > a(\lambda)$ for any $\lambda \leq 1/b$, which implies that $V(\lambda) > a(\lambda)$ for any $\lambda \leq 1/b$.

We have shown that $k_4 = 0$ when $\lambda \leq 1/b$. Also, when $\lambda = 1/b$, the only threshold mechanism that is optimal is the one with a threshold at $c_1/2b$. The agent's ex ante payoff from this mechanism is $\tilde{c}_0$.

As in Lemma 9, define $\bar{\lambda}$ to be the smallest $\lambda$ such that $V(\lambda) = a(\lambda)$. Define $\bar{c}_0$ to 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 $\mathcal{L}(\mathbb{x}, \bar{\lambda}, 0)$:

$$\bar{c}_0 := \min\{k_4 b + \sum_{i=1}^{3} k_i \left( b \int_{\mu_i}^{1} \left( \mu + (1-\mu)\frac{\mu_i}{1-\mu_i}) \right) \mathrm{d}F(\mu) + c_1 F(\mu_i) \right) \mid \mathbb{x}$$
$$\in \arg\max_{\mathbb{x} \in \bar{\mathbb{X}}_1} \mathcal{L}(\mathbb{x}, \bar{\lambda}, 0)\}.$$

The minimum is well defined because $\arg\max_{\mathbb{x}\in\bar{\mathbb{X}}_1} \mathcal{L}(\mathbb{x}, \lambda, 0)$ is a compact set. It is clear that we must have $k_4 = 0$ to obtain $\bar{c}_0$. It is clear that $\bar{c}_0 < b$, since the agent cannot obtain a payoff of $b$ from a mechanism with $k_4 = 0$. By the definition of $\bar{\lambda}$, we have $\bar{\lambda} > 1/b$. By Lemma 10, the corresponding ex ante outside options, $\bar{c}_0$ and $\tilde{c}_0$, must satisfy $\bar{c}_0 \geq \tilde{c}_0$.

Note that for any $c_0 < \bar{c}_0$, the multiplier $\lambda$ corresponding to $c_0$ must be weakly smaller than $\bar{\lambda}$ because of Lemma 10. By the definition of $\bar{c}_0$, it cannot be that $\lambda = \bar{\lambda}$. Thus $\lambda < \bar{\lambda}$, and by the definition of $\bar{\lambda}$, there exists $\mathbb{x} \in \bar{\mathbb{X}}$ with $k_4 = 0$ that solves the problem $(\mathcal{P})$ and induces $c_0$ as the agent's ex ante payoff.

$(ii)$ holds because the principal cannot be better off when $c_0$ is higher.

To show $(iii)$, suppose we start from $\mathbb{x}$ that induces $\bar{c}_0$ for the agent. Let $\bar{\mu}_i$ denote the thresholds of $\mathbb{x}$. By keeping each threshold fixed at $\bar{\mu}_i$, increasing $k_4$ 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 $[\bar{c}_0, b]$. Finally, Lemma 5 implies that whenever $p_0 > 0$, we must have $\bar{\mu}_i > 0$ for each $i = 1, 2, 3$.

### *1.11.6 Proof of Theorem 2*

We know from Theorem 1 that there always exists an optimal mechanism that is a convex combination of three threshold mechanisms. Recall that we defined

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

As we argued in the proof of Theorem 1, for any optimal mechanism that is a convex combination of three threshold mechanisms, there exist multipliers $\lambda, \eta \geq 0$ such that, if a nonzero weight $k_i$ is placed on a threshold mechanism with threshold $\mu_i$, then $\mu_i$ maximizes $t(\mu_i, \lambda, \eta)$ across $\mu_i \in [0, 1)$. Since $\eta(1 - \frac{\mu_i}{1-\mu_i})$ is concave in $\mu_i$, $t(\mu_i, \lambda, \eta)$ is strictly concave in $\mu_i$ as long as $T(\mu_i, \lambda)$ is strictly concave in $\mu_i$, which we have assumed to be true for $\lambda \geq \lambda_0$.

It remains to show that it is without loss to restrict attention to $\lambda \geq \lambda_0$. Since $t(\mu_i, \lambda_0, \eta)$ is strictly concave in $\mu_i$, one can see from (1.12) that $t(\mu_i, \lambda_0, 0)$ is maximized at $\mu_i = 0$. That is, when $\lambda = \lambda_0$ and $\eta = 0$, the mechanism that maximizes the Lagrangian (and thus solves $(\mathcal{P})$) is the principal's most-preferred mechanism, $\sigma^+(\mu) = 0$ and $\sigma_2(\mu, -1) = 0$. This mechanism gives a payoff of $b\mu_0$ to the agent. By Lemma 10, 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 $\lambda_0$.

Strict concavity of $t(\mu_i, \lambda, \eta)$ implies that there can be at most one $\mu_i$ that maximizes the Lagrangian. Therefore, it must be that the three threshold mechanisms are actually identical.

### *1.11.7 Alternative Sufficient Condition for Regularity*

The following proposition provides an alternative sufficient condition for the principal's problem to be regular.

**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, the principal's problem is regular.*

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

$$\frac{\partial t}{\partial \mu_i} = \frac{\lambda b - 1}{1 - \mu_i} \left( \frac{2}{(1-\mu_i)^2} \int_{\mu_i}^{1} (1-\mu) f(\mu)\, \mathrm{d}\mu - f(\mu_i) \right) - \lambda f'(\mu_i)(2b\mu_i - c_1) - 2\lambda b f(\mu_i) - \frac{2\eta}{(1-\mu_i)^3} . \tag{1.13}$$

Define $\bar{f} \coloneqq \max\{f(\mu) \mid \mu \in [0,1]\}$ and $\bar{f}' \coloneqq \max\{|f'(\mu)| \mid \mu \in [0,1]\}$, 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 (1.13) is strictly negative if

$$\bar{f}' < \frac{2\lambda b \underline{f}}{|\lambda b - 1| + \lambda(2b - c_1)} \quad \forall \mu \in [0,1]. \tag{1.14}$$

The assumption of the proposition implies that (1.14) holds for $\lambda \to \infty$ and $\lambda = \lambda_0$. But the RHS of (1.14) is single-peaked in $\lambda$ (with the peak at $\lambda = 1/b$). Therefore, (1.14) holds for all $\lambda \ge \lambda_0$.

□

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

### *1.11.8 Proof of Proposition 3*

Let $\bar{\mu}$ be the unique value of the optimal threshold $\mu^*$ when the agent's ex ante outside option is $\bar{c}_0$. For $c_0 \in (\bar{c}_0, b)$, it is optimal to place positive weights on both the always-promote mechanism and a the threshold mechanism with threshold $\bar{\mu}$. We showed in 1.11.5 that a positive weight is placed on the always-promote mechanism only if $\lambda > 1/b$. Since a positive weight is placed on the threshold mechanism and $\lambda > 1/b$, by (1.11), 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 $\bar{\mu} > c_1/2b$.

Uniqueness and monotonicity of $\mu^*$ in $(i)$ hold because the principal's payoff is strictly decreasing in $\mu^*$. Likewise, the uniqueness, monotonicity, and continuity of $p$ in $(ii)$ hold 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.

### *1.11.9 Proof of Lemma 5*

Suppose to the contrary that $\sigma = (\sigma^+, \sigma^-)$ is optimal and satisfies $\sigma^+(0) > \sigma^-(0) > 0$. Since $\sigma^-$ 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 $\epsilon > 0$, consider the functions $\sigma_\epsilon^+$ and $\sigma_\epsilon^-$ defined by

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

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

For $\epsilon$ small enough, we have $\frac{\epsilon}{1-\epsilon}(\sigma^-(0) - \sigma^+(0)) > -\sigma^-(0)$, so that the pair of functions $\sigma + \sigma_\epsilon := (\sigma^+ + \sigma_\epsilon^+, \sigma^- + \sigma_\epsilon^-)$ defines an incentive compatible mechanism. However, for $\epsilon$ small enough, $\sigma + \sigma_\epsilon$ must induce a greater value of the Lagrangian $\mathcal{L}_1$ compared to $\sigma$. To see this, evaluate $\mathcal{L}_1$ at $\sigma_\epsilon$ to get:

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

For $\epsilon$ small enough, the integrand converges to $\lambda c_1(\sigma^+(0) - \sigma^-(0))$, which we know to be strictly positive because we showed that $\lambda \geq \lambda_0 > 0$ in the proof of Theorem 2.

## REFERENCES


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

Susan Athey, Andrew Atkeson, and Patrick J Kehoe. The optimal degree of discretion in monetary policy. *Econometrica*, 73(5):1431–1475, 2005.

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

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

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

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

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

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

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

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

Robert K Dixit, Avinash K Dixit, and Robert S Pindyck. *Investment under uncertainty*. Princeton university press, 1994.

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

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

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

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

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

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

Masanori Hashimoto and Ben T Yu. Specific capital, employment contracts, and wage rigidity. *The Bell Journal of Economics*, pages 536–549, 1980.

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

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

Andreas Kleiner, Benny Moldovanu, and Philipp Strack. Extreme points and majorization: Economic applications. *Econometrica*, 89(4):1557–1593, 2021.

Andrei Nikolaevich Kolmogorov and Sergei Vasilevich Fomin. *Introductory real analysis*. Courier Corporation, 1975.

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

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

Xiao Lin and Ce Liu. Credible persuasion. *arXiv preprint arXiv:2205.03495*, 2022.

David G Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1997.

Dale T Mortensen. Specific capital and labor turnover. *The Bell Journal of Economics*, pages 572–586, 1978.

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

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

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

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

Dmitry Orlov, Andrzej Skrzypacz, and Pavel Zryumov. Persuading the principal to wait. *Journal of Political Economy*, 128(7):2542–2578, 2020.

Jérôme Renault, Eilon Solan, and Nicolas Vieille. Optimal dynamic information provision. *Games and Economic Behavior*, 104:329–349, 2017.

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

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

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

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