# Cowles Foundation

**for Research in Economics**
**at Yale University**

Cowles Foundation Discussion Paper No. 1823


OPTIMALLY EMPTY PROMISES AND ENDOGENOUS SUPERVISION


David A. Miller and Kareen Rozen



October 2011

# OPTIMALLY EMPTY PROMISES AND ENDOGENOUS SUPERVISION*

David A. Miller† UCSD

Kareen Rozen‡ Yale

September 30, 2011

**Abstract**

We study optimal contracting in a team setting with peer monitoring and moral hazard. This environment reflects stylized characteristics of production environments with complex tasks: agents have many opportunities to shirk, task-level monitoring is needed to provide useful incentives, and it is difficult to write performance-based clauses into explicit contracts. Incentives are provided informally, using wasteful punishments like guilt and shame, or slowed promotion. These features give rise to optimal contracts with "empty promises" and endogenous supervision structures. Agents make promises that they don't necessarily intend to keep, leading to the optimal concentration of supervisory responsibility in the hands of one or two agents.

**Keywords:** Partnership, teams, moral hazard, monitoring, supervision, costly punishments.
**JEL Codes:** C72, D03, D86.

*This work subsumes our earlier paper, "Monitoring with collective memory: Forgiveness for optimally empty promises." We are grateful to Andy Postlewaite, Larry Samuelson, Todd Sarver, Andy Skrzypacz, Joel Sobel, Juuso Välimäki, Joel Watson, and Bob Wilson for helpful comments and suggestions. We also thank seminar participants at Columbia, London Business School, LSE, Northwestern, Penn State, Princeton, Queen Mary, Stanford GSB, Toronto, Yale, Zurich, the 2010 NSF/NBER/CEME Conference on General Equilibrium and Mathematical Economics, and the 2011 North American Winter Meeting of the Econometric Society. Max Dvorkin, Jung-Myun Moon, and Aniela Pietrasz provided excellent research assistance.

†Address: Dept. of Economics, 9500 Gilman Dr. #0508, La Jolla, CA 92093–0508. E-mail: d9miller@ucsd.edu. Home page: http://dss.ucsd.edu/~d9miller. Miller thanks Yale University and the Cowles Foundation for Research in Economics for hospitality and financial support.

‡Address: Dept. of Economics and the Cowles Foundation for Research in Economics, 30 Hillhouse Ave., New Haven, CT 06511. E-mail: kareen.rozen@yale.edu. Home page: http://www.econ.yale.edu/~kr287/. Rozen thanks the NSF for financial support through grant SES-0919955.

# 1 Introduction

Carol is Bob's supervisor. Bob is capable of performing up to four complex tasks that benefit both of them, but each task requires costly effort to complete. When the time comes for Bob to perform the tasks, he privately learns which of the tasks are feasible, and then privately decides which feasible tasks he will complete and which ones he will shirk. At the end, Carol will monitor Bob's tasks. Upon finding that a task fails her inspection, however, she cannot tell whether the task was infeasible or whether Bob intentionally shirked. The only instruments for motivating Bob are punishments that do not transfer utility to Carol; for instance, he can be reprimanded, demoted, or fired.[1]

Should Bob always exert effort on all his feasible tasks? Not if task feasibility is moderately uncertain. Then it is optimal for Bob to make "empty promises," and for Carol to be "forgiving" when she decides how much to punish him. Empty promises arise when Bob is assigned four tasks but completes, say, at most two of them, regardless of how many are feasible. To implement Bob's empty promises, Carol should not punish him unless three or more of his tasks fail her inspection. Of course, if Bob is forgiven his first two failures, he will have no incentive to complete more than two of his tasks. But the likelihood that three or four of his tasks will be feasible is not high, so the cost of making two empty promises is small; and forgiving those empty promises saves Bob from punishment in the more likely event that two or fewer of his tasks are feasible. Empty promises thus serve as a buffer against task infeasibility.

In the above arrangement, Bob is making empty promises even though Carol is monitoring all of his tasks. Would Carol's efforts be put to better use in performing some tasks, which Bob would monitor? If Bob and Carol are each capable of performing or monitoring up to four tasks, in total, then there are other arrangements under which four promises are made and all of them are monitored. For example, they could each make two promises and monitor the other's two. Or Bob could make one promise and Carol could make three, with each monitoring all the promises of the other. Are all these arrangements welfare-equivalent? No—if empty promises are optimal when one person purely supervises the other, then this arrangement dominates the mutual monitoring arrangements. So in the optimal contract an endogenous supervisor emerges. On larger teams, supervisory responsibility is optimally concentrated even if tasks can be monitored with probability

[1] We assume task completion is not formally contractible, so that Carol cannot be legally bound to pay Bob a bonus for completing tasks. If Carol and Bob are not very patient, then even in a relational contracts setting (e.g., Baker, Gibbons, and Murphy, 1994; Che and Yoo, 2001; Levin, 2003; MacLeod, 2003; Radner, 1985) there is limited scope for Carol to commit to discretionary bonuses or for Bob to commit to voluntary fines. Furthermore, even the limited scope for such commitment may be needed to support other aspects of their relationship, such as investment in human capital.

less than one.

Our task-based approach with peer monitoring and wasteful punishments fits some stylized characteristics of an important class of partnership and team environments: those in which production is complex and requires accumulated job-specific human capital. According to Lazear and Shaw (2007), from 1987 to 1996 "the percent of large firms with workers in self-managed work teams rose from 27 percent to 78 percent," and moreover "the firms that use teams the most are those that have complex problems to solve." Similarly, Boning, Ichniowski, and Shaw (2007) find that steel minimills with more complex production processes are more likely to organize workers into "problem-solving teams." In such environments, agents face many opportunities to shirk, rather than one or several. For instance, an agent may face numerous tasks in a single workday, and yet output, being complex, may be measurable only weekly or monthly, as well as noisy or hard to quantify. Task-level monitoring is needed to provide useful incentives, and monitoring the performance of complex tasks is difficult for anyone who is not intimately familiar with related tasks. One example is the shopfloor of a Japanese manufacturing firm, on which Aoki (1988, p. 15) writes,

> The experience-based knowledge shared by a team of workers on the shopfloor may be tacit and not readily transferable in the form of formal language, but is quite useful in identifying local emergencies, such as product defects and machine malfunctions, on the spot and solving them autonomously. Those workers nurtured in a wide range of [job-specific] skills may be able to understand, as individuals or as a collective, why defective products have increased, and may be able to devise and implement measures to cope with the situation and thus prevent the problem from recurring. This can be done without much, if any "outside" help…

Finally, since the output of a specialist is hard to describe, it is hard to write performance-based incentives into a formal contract. So incentives must be provided informally, using wasteful instruments like guilt and shame (Barron and Gjerde, 1997; Carpenter, Bowles, Gintis, and Hwang, 2009; Kandel and Lazear, 1992), or slowed promotion.When extreme punishments arise, typically the worst available is separation, with its attendant search and dislocation costs. Aoki (1988, p. 58) writes "unless the employee possesses special skills that might be needed elsewhere, the value of those skills, accumulated in the context of teamwork and internal personal networking, would by and large be lost;" and, moreover, "midcareer separation may signal negative attributes."

If, in our example, Bob and Carol are on a shopfloor team, Bob's complex tasks may be to try to fix small malfunctions in the production equipment in order to avoid introducing defects into the product. A task is feasible if Bob is capable of fixing the particular malfunction on his own, but Bob can also shirk the task by ignoring the malfunction. When Carol monitors Bob's tasks, she can observe whether a product is defective, but she cannot discern whether the defects arose from a malfunction that Bob could have fixed. To motivate Bob, if there are too many defects then Carol can reprimand him in front of the other workers, and add demerits to his personnel file that will slow his path of promotion through the firm or ultimately lead the firm to fire him.

This paper fits into the literature on partnership and teams with moral hazard, but emphasizes a new perspective on teamwork. Inspired by complex production environments, where output is hard to quantify and agents have built up specialized, job-specific human capital, we bring together three features: peer monitoring, high-dimensional effort, and wasteful punishments. These features provide the foundation for studying two important tradeoffs—between production and monitoring, and between punishment and forgiveness—as well as the endogenous allocation of monitoring responsibility. Much of the literature on teams addresses contracts that depend on stochastic team output, and focuses on the problem of free-riding,[2] or allows for exogenously specified individual-level monitoring.[3] In contrast, our approach endogenizes individual-level monitoring by putting the agents in charge of monitoring each other. We assume that assigning an agent to monitor her peers crowds out her own productivity.[4] This allows us to study the tradeoff between productive and supervisory activity at both the individual level and the team level, and to study the optimal assignment of agents into productive and supervisory roles.

Whereas the prior literature generally studies agents who exert effort along one dimension or several complementary dimensions,[5] in our model each task constitutes an independent dimension of effort.[6] This assumption imposes a natural structure on the stochastic relationship among ef-

[2]For example: Battaglini (2006); d'Aspremont and Gérard-Varet (1998); Legros and Matsushima (1991); Legros and Matthews (1993). Free-riding, of course, also arises in public goods problems (e.g., Palfrey and Rosenthal, 1984).

[3]For example: Carpenter, Bowles, Gintis, and Hwang (2009); Che and Yoo (2001); Holmström (1982); Holmström and Milgrom (1991); Kvaløy and Olsen (2006); Laux (2001); Matsushima, Miyazaki, and Yagi (2010); McAfee and McMillan (1991); Miller (1997); Mirrlees (1976).

[4]Li and Zhang (2001) formalize the conjecture of Alchian and Demsetz (1972) that if monitoring is costly then it should be the responsibility of a residual claimant. Rahman (2010) and Rahman and Obara (2010) show that the monitor need not be the residual claimant when a mediator can make correlated recommendations. Like Li and Zhang (2001), we show that monitoring responsibilities are optimally concentrated in the hands of one agent, but like Rahman (2010) and Rahman and Obara (2010) we do not need to give the monitoring agent residual claims. There is also a small literature on costly monitoring in principal-agent relationships (e.g., Border and Sobel, 1987; Mookherhjee and Png, 1989; Snyder, 1999; Williamson, 1987).

[5]For example: Alchian and Demsetz (1972); Aoki (1994); Barron and Gjerde (1997); Battaglini (2006); Carpenter, Bowles, Gintis, and Hwang (2009); Che and Yoo (2001); Holmström (1982); Kandel and Lazear (1992); Kvaløy and Olsen (2006); Li and Zhang (2001); McAfee and McMillan (1991); Mirrlees (1976).

[6]Matsushima, Miyazaki, and Yagi (2010) also study a model where agents have access to private information about the feasibility of arbitrarily many independent tasks, but assume that monitoring is exogenous and utility is transferable. d'Aspremont and Gérard-Varet (1998); Holmström and Milgrom (1991); Laux (2001); Legros and Matsushima (1991); Legros and Matthews (1993); Miller (1997) allow for arbitrarily many dimensions of effort, but assume the agents have no private information.

fort, output, and monitoring, and enables us to make more specific predictions about promises and supervision than would be possible with a single dimension of continuous effort.

Finally, a majority of the literature assumes that all incentives are provided through monetary payments, such that only the imbalance must be burned (or given to a residual claimant).[7] Instead, we rule out formal monetary payments, and focus on providing incentives through informal punishments that are socially wasteful.[8] Such punishments are a natural instrument in an environment in which the agents cannot commit to inspection-contingent payments. In principle, formal monetary payments are one polar case, while wasteful punishments are another. In practice, if a team of three or more agents *can* commit to inspection contingent payments, then it is easy for them to implement the first best (see footnote 12). Since many realistic scenarios share some features of both polar cases, the natural next step is to study the implications of wasteful punishments.

Our results build on the intuition behind Bob's empty promises and Carol's monitoring responsibilities. First, Section 3.1 considers the simple case of two identical agents with bounded capacity, one of whom is exogenously assigned to supervise the other. We show that it is optimal for the worker—Bob, in this scenario—to use a cutoff strategy of performing as many feasible tasks as he can, up to some cutoff $p^*$. The schedule of punishments that implements this strategy is "kinked-linear": it punishes him with a constant per-task penalty whenever he completes fewer than $p^*$ tasks, and does not punish him at all when he completes at least $p^*$ tasks. The cutoff $p^*$ is increasing in the probability, $\lambda$, that any given task is feasible. Empty promises optimally arise for an intermediate range of $\lambda$, but empty promises arise for any $\lambda \in (0, 1)$ if either the agents' capacities are sufficiently large or the ratio of private costs to team benefits is sufficiently close to unity. These results illustrate the basic structure of optimal contracts when there are enough supervisory resources to monitor every task that an agent promises to complete.

Next, Section 3.2 addresses the question of who should supervise whom, while maintaining the restrictions that every promised task must be monitored and that monitoring one task reduces an agents' capacity to perform tasks by one. Even though Bob and Carol are identical, it is indeed optimal for one of them to specialize in performing tasks while the other specializes in monitoring. Intuitively, concentrating supervisory responsibility in Carol's hands eliminates the possibility that

[7]For example: Alchian and Demsetz (1972); Battaglini (2006); d'Aspremont and Gérard-Varet (1998); Holmström (1982); Holmström and Milgrom (1991); Laux (2001); Legros and Matsushima (1991); Legros and Matthews (1993); Matsushima, Miyazaki, and Yagi (2010); McAfee and McMillan (1991); Miller (1997); Mirrlees (1976); Rahman and Obara (2010). In addition, there is a literature that views bonuses and penalties as financially equivalent, depending only on the reference point, but then applies reference-dependent preferences to distinguish between their incentive effects (e.g., Aron and Olivella, 1994; Frederickson and Waller, 2005).

[8]Wasteful punishments are also studied by Barron and Gjerde (1997); Carpenter, Bowles, Gintis, and Hwang (2009); Che and Yoo (2001); Kandel and Lazear (1992).

Bob could have more feasible tasks than he is willing to complete even while Carol has fewer feasible tasks than she would be willing to complete. For example, compare an asymmetric contract in which Carol supervises and Bob completes at most two of his four assigned tasks to a symmetric contract in which Carol and Bob are each assigned two tasks and complete at most one of them. In both contracts, four tasks are assigned in total, and at most two of them will be completed. But in the symmetric contract there is the possibility that Bob has two feasible tasks while Carol has none, in which case only one task is completed. In contrast, when all the tasks are assigned to Bob and two of them are feasible, he completes both of them. More generally, when there are many identical players it is optimal for at least one of them to devote herself monitoring. Having a monitoring specialist—a supervisor—is strictly optimal whenever there are empty promises.

Section 4 then shows how to economize on monitoring. Since the agents have unbounded liability, with large penalties even a small chance of being monitored can enforce behavior. So in a team of $N$ agents, it is simple but effective to appoint one agent to supervise; she simply randomizes over which of the other $N - 1$ agents to monitor. But it is possible to further increase productivity by appointing two "partial supervisors," each of whom spends less than half her capacity on monitoring. In this arrangement, more resources are devoted to task completion. There is a tradeoff: since no agent will ever have all his tasks monitored, the punishment schedule must be coarser, and in the most general environment it may not even be optimal for agents to use cutoff strategies. If $\lambda$ is sufficiently high, however, appointing two partial supervisors improves over the single-supervisor arrangement.

For our main results, we assume that monitoring is costless (aside from its opportunity cost, of course), and therefore take for granted that agents are willing to monitor each other. But our results hold up even if monitoring is costly, as shown in Section 5.1, because agents can discipline each other for failing to monitor. Suppose that Bob's first task turns out not to be feasible, but Carol does not flag it as uncompleted. Then Bob can reveal that he did not complete it. Bob suffers no consequence for this revelation, but Carol is punished. By calibrating Carol's punishment to her cost of monitoring, they can assure that Carol monitors properly along the equilibrium path. We also show that our results are robust to the possibility of exchanging messages after tasks are performed but before they are monitored (Section 5.2), and to imperfections in monitoring (Section 5.3).

Finally, Section 6 shows that optimally empty promises and endogenous supervision extend to teams within firms. In a team within a firm, we assume that the agents do not receive any inherent benefits from accomplishing tasks. The firm offers the agents a contract that specifies fixed wages, task and monitoring responsibilities, bonuses based on the aggregate output of the team, and a schedule of wasteful punishments as a function of monitoring outcomes. Because individual rationality must bind, the firm faces the same optimization problem as the partnerships we study, except that now the team-based bonus is a choice variable rather than a parameter. We show that for appropriate parameters the firm's optimal contract involves empty promises and endogenous supervision, just as in a partnership.

# 2 Model and preliminaries

Consider a team of $N \geq 2$ risk-neutral agents, each of whom may perform or monitor up to $M$ tasks. There is a countably infinite set of tasks $\mathcal{X}$, each of which is an identical single-agent job. Any given task $x \in \mathcal{X}$ is feasible with independent probability $\lambda \in (0, 1)$. If a task is infeasible, then it cannot be completed. If a task is feasible, the agent performing it can choose whether to shirk or exert effort cost $c > 0$ to complete it. Shirking is costless, but yields no benefit to the team. If the agent exerts effort to complete the task, each member of the team (including him) receives a benefit $\frac{b}{N}$, where $b > c > \frac{b}{N}$. Hence each task is socially beneficial, but no agent is willing complete it without further incentives. To simplify exposition, we assume that monitoring requires zero effort cost (Section 5.1 shows this can be relaxed without affecting our results).

The timing of the game is as follows:

- At $\tau = 1$, each agent publicly promises to perform a set of tasks. Each task can be promised by at most one agent.[9] We call a promised task a "promise" for short.
- At $\tau = 2$, each agent privately observes the feasibility of each task he promised, and, for each feasible task, privately decides whether to shirk or exert effort.
- At $\tau = 3$, agents monitor each other. An agent who made $p$ promises at $t = 1$ can monitor up to $M - p$ of the other agents' promises. Each task can be monitored by at most one agent, but the agents can employ an arbitrary correlation device to coordinate their monitoring activities. Conditional on being monitored, with probability 1 a completed task will pass inspection, and an uncompleted task will fail inspection.[10] The monitoring agent, however, cannot distinguish whether the task was infeasible or intentionally shirked.
- At $\tau = 4$, the agents reveal the results of their inspections.[11]
- At $\tau = 5$, each agent can impose unbounded punishments on other agents, at no cost to himself.

[9] The spirit of this requirement is that the agents discuss and then agree on who will promise which tasks. But formally it is simpler to assume that the agents make their promises sequentially, in any arbitrary order.

[10] Perfect monitoring simplifies the exposition; extension to imperfect monitoring is discussed later.

[11] Whether inspection results are verifiable does not matter in most of our analysis. In Section 5.1, we show that our analysis is robust to costly monitoring with unverifiable inspections.

We consider a setting in which it is not possible to commit to transfers that are contingent on inspection outcomes; instead, any penalty imposed on an agent is pure waste.[12] We study perfect Bayesian equilibria of this game. Since the punishments at $\tau = 5$ are unbounded and costless for each agent to impose, the agents can discourage any observable deviations from the equilibrium path—in particular, deviations at time $\tau = 1$ are immediately observable. Similarly, since monitoring is costless, it is without loss of generality to restrict attention to equilibria in which players do not deviate from the equilibrium monitoring scheme at time $\tau = 3$ and reveal their inspection results truthfully at time $\tau = 4$. Accordingly, we limit attention to behavior along the equilibrium path. In such equilibria, the main concern is to discourage unobservable deviations at time $\tau = 2$. We call the specification of equilibrium-path behavior a *contract*, for reasons that we address in Remark 1, below.

In what follows, for any countable set $Z$, let $\mathcal{PS}(Z)$ denote the power set of $Z$, and let $\Delta(Z)$ denote the set of probability distributions over $Z$.

**Definition 1.** *A* contract *specifies equilibrium behavior, for each agent $i = 1, 2, \ldots, N$, of the following form:*

1. *A promise scheme $P \in \times_{i=1}^{N} \mathcal{PS}(\mathcal{X})$, with each distinct $P_i$ and $P_j$ disjoint in $\mathcal{X}$, specifying which tasks should be promised by which agents at $\tau = 1$;*
2. *A task completion strategy $s_i : \mathcal{PS}(P_i) \to \Delta(\mathcal{PS}(P_i))$ for each agent i, with every realization a subset of the argument, specifying the subset of her promises to complete among those that are feasible at $\tau = 2$;*
3. *A monitoring scheme $r \in \Delta \times_{i=1}^{N} \mathcal{PS}\big(\bigcup_{j \neq i} P_j\big)$, with every realization a vector of disjoint subsets of $\mathcal{X}$, specifying a joint distribution over which agents should monitor which tasks at $\tau = 3$;*
4. *A punishment scheme $v : \times_{i=1}^{N} \big(\mathcal{PS}(P_i) \times \mathcal{PS}(P_i)\big) \to \mathbb{R}_{-}^{N}$, specifying the net punishment imposed on each player $i = 1, \ldots, N$ at $\tau = 5$ as a function of which tasks were shown to pass and fail inspection at $\tau = 4$.*

**Remark 1.** *We refer to truthful equilibrium path behavior as a "contract" to emphasize that this game environment can also be interpreted as a contractual setting. Suppose some external principal offers the agents a contract in which the principal formally commits to pay each agent $b/N$ for each task completed by the team, and informally recommends a promise scheme, task completion strategies, a monitoring scheme, and a punishment scheme. Then it should be a perfect Bayesian*

[12]For at least three players, solving this problem would be trivial if agents could commit to outcome-contingent transfers: *without affecting team welfare*, whenever one of his tasks fails inspection, an agent could be forced to make payments to a third agent (not the monitor of that task) that are sufficiently high to ensure task completion.

*equilibrium for the agents to be obedient to the recommendations and report their inspection outcomes truthfully, as well as for the principal to implement the recommended punishment scheme. We investigate this principal-agent interpretation formally in Section 6.*

A contract must respect each agent's bounded capacity. A contract is *feasible* if $|P_i| + \max_{R \in \text{supp } r} |R_i| \leq M$ for all $i$; i.e., no player is asked to perform and monitor more than $M$ tasks in total. A contract is *incentive compatible* if no agent has an incentive to deviate from his task completion strategy, given the promise, monitoring, and punishment schemes, and assuming that all agents reveal their inspection results truthfully. A contract is *optimal* if it maximizes the team's aggregate utility within some class of incentive compatible contracts.

**Remark 2.** *A more general space of contracts would allow the agents to employ a randomized promise scheme. However, for our purposes it is without loss of generality to restrict attention to deterministic promise schemes. For any optimal contract with a random promise scheme, there would be an equally good deterministic promise scheme in the support of the randomization.*

*If agents had the ability to opt out of the game before time $\tau = 1$, then for any contract yielding positive social welfare the agents would be willing to accept the contract "behind the veil of ignorance," i.e., before their "roles" (as workers or supervisors) are randomly assigned. Alternatively, if they could make noncontingent transfers, then they could spread the wealth so as to make everyone willing to accept the contract, no matter how asymmetric are the roles. Accordingly, we do not impose individual rationality constraints on the contract.*

Before formalizing the incentive compatibility constraints, we show that the relevant space of contracts can be simplified without loss of generality.

**Lemma 1.** *There exists an optimal contract satisfying the following, for each agent i:*

1. *The number of tasks agent $i$ completes is a deterministic function of the set $A \subseteq P_i$ of his tasks that are feasible (so with some abuse of notation let $|s_i(A)|$ be this number);*
2. *Agent $i$'s penalty depends only on his tasks that failed inspection, so without loss of generality $v_i : \mathcal{PS}(P_i) \to \mathbb{R}_-$;*
3. *"Upward" incentive compatibility constraints for task completion are slack—when $A \subseteq P_i$ tasks are feasible, agent $i$ strictly prefers to complete $s_i(A)$ over any completing any feasible set $A'$ for which $|s_i(A)| < |A'|$;*
4. *$s_i(s_i(A)) = s_i(A)$; in addition, $A \subseteq A'$ implies $|s_i(A)| \leq |s_i(A')|$.*

Given a promise scheme $P$, a task completion strategy profile $s$, and a monitoring scheme $r$, let $\rho_i(f; s_i(A))$ be the probability that $f \subseteq P_i$ is the set of player $i$'s tasks that fails inspection when player $i$ completes the set of tasks $s_i(A)$, and let $p_i = |P_i|$. The optimal contract $(P, s, r, v)$ maximizes

$$\sum_{i=1}^{N} \sum_{A \subseteq P_i} \lambda^{|A|}(1-\lambda)^{p_i - |A|} \Big( |s_i(A)|(b-c) + \sum_{f \subseteq P_i} v_i(f) \rho_i(f; s_i(A)) \Big) \tag{1}$$

subject to feasibility and downward incentive compatibility (IC)

$$\sum_{f \subseteq P_i} v_i(F) \rho_i(f; s_i(A)) + |s_i(A)| \Big( \frac{b}{N} - c \Big) \geq \sum_{f \subseteq P_i} v_i(F) \rho_i(f; A') + |A'| \Big( \frac{b}{N} - c \Big) \tag{2}$$

for each downward deviation $A' \subset s_i(A)$, for each set of feasible tasks $A \subseteq P_i$, and for each agent $i$.

# 3 Empty promises and endogenous supervision

In this section we show the optimality of empty promises and how they endogenously give rise to optimal supervision structures. Throughout this section, we impose the restriction that every task must be monitored. Formally, a contract is *optimal* if it maximizes the sum of the agents' utility among individually rational and incentive compatible contracts in some class. A task completion strategy $s_i$ has *empty promises* if there is some subset $A$ of the promises of agent $i$ for which $s_i(A) \neq A$. Otherwise (i.e., if $s_i(A) = A$ for all $A$), the strategy is *promise keeping*. Our results identify *cutoff strategies* as an important class of task completion strategies. A task completion strategy $s_i$ is a cutoff strategy if there is a cutoff $p_i^*$ such that $|s_i(A)| = \min\{|A|, p_i^*\}$ for every subset $A$ of agent $i$'s promises. A cutoff strategy has empty promises if $p_i^* < p_i$.

## 3.1 A worker and a supervisor

Before discussing how endogenous supervision may arise in Section 3.2, we first examine the implications of a very simple supervisory structure. Suppose the team consists of two members, and that the promise scheme calls for a "worker" who promises all the tasks and a "supervisor" who monitors all the tasks. Because only one agent is completing tasks, we drop the $i$ subscript and simply use $p$ to denote the number of promises that the worker makes and $s$ to denote his task completion strategy. The following result characterizes the optimal contract conditional on having such a worker-supervisor structure.

**Theorem 1.** *Conditional on a worker-supervisor structure, the optimal contract satisfies:*

1. *The worker makes as many promises as possible ($M$) but uses a cutoff strategy, completing at most $p^*$ feasible tasks. The supervisor monitors all $M$ tasks.*
2. *The cutoff $p^*$ is increasing in $\lambda$.*
3. *There are empty promises ($0 < p^* < p$) when $1 - \sqrt[M]{2 - \frac{c}{b/2}} < \lambda < \sqrt[M]{\frac{c}{b/2} - 1}$.*
4. *Penalties depend only on the number of failed inspections. No penalty is imposed on the worker up to a threshold of $p - p^*$ inspection failures, but each additional inspection failure results in a marginal penalty of $c - b/2$.*

Theorem 1 says that under a worker-supervisor structure, the optimal contract has the worker complete only up to a cutoff $p^*$ of tasks, even though the worker makes $M$ promises and the supervisor monitors each and every one. The cutoff $p^*$, which is increasing in the probability of task feasibility $\lambda$, is strictly positive whenever $1 - \sqrt[M]{2 - \frac{c}{b/2}} < \lambda$, and is strictly smaller than the number of promises made whenever $\lambda < \sqrt[M]{\frac{c}{b/2} - 1}$. Recall that for $N = 2$, $\frac{b}{2} < c < b$ and so $\frac{c}{b/2} \in (1, 2)$. Hence, the interval of $\lambda$'s for which there are empty promises increases with both the capacity and cost-benefit ratio. Indeed:

**Corollary 1.** *As capacity size $M \to \infty$, there are empty promises for any $\lambda \in (0, 1)$. The same is true when the cost-benefit ratio of a task $\frac{c}{b/2} \to 2$.*

We prove Theorem 1 below. To first understand the intuition behind the optimality of empty promises, note that even if the worker intends to keep all his promises, some of his tasks are likely to be infeasible because $\lambda < 1$, so he will incur penalties anyway. Since penalties are costly, it is possible to reduce the cost of punishment by forgiving a few failures. However, the worker is able to move the support of the monitoring distribution: for example, if he makes ten promises, and the threshold for a penalty is three failures, then he will never fulfill more than eight promises, even if all ten are feasible. When $\lambda$ is not too close to one, this tradeoff is resolved in favor of empty promises.

*Proof.* Suppose that the task completion strategy $s$ is optimal and that the worker optimally promises the set of tasks $P$. Note that if the worker completes $s(A)$ when the set $A \subset P$ is feasible, then all of $P \setminus s(A)$ will fail inspection. Incentive compatibility of $s$ requires that for all $A' \subset s(A)$,

$$v(P \setminus s(A)) + |s(A)|(\frac{b}{N} - c) \geq v(P \setminus A') + |A'|(\frac{b}{N} - c). \tag{3}$$

Let $p^* = \max_{A \subseteq P} |s(A)|$ be the size of the largest set of tasks completed under $s$ (in light of Lemma 1, $p^* = |s(P)|$). Examination of Eq. 3 reveals that expected penalties are minimized under the *kinked* *linear* scheme $v(P \setminus s(A)) = (\frac{b}{2} - c)\max\{|P \setminus s(A)| - (p - p^*), 0\}$, which imposes no penalty when $p^*$ or more tasks are completed, but a penalty of $(\frac{b}{2} - c)(p^* - s(A))$ whenever $|s(A)| < p^*$. This penalty scheme is kinked-linear in the number of tasks left uncompleted. Consider extending the definition of $v$ above to all subsets of $P$ of size at most $p^*$. Suppose that for some subset $A \subset P$ of size at most $p^*$, the strategy prescribes $s(A) \subset A$. But then $s$ is suboptimal, since completing the extra tasks $a \setminus s(A)$ both decreases $v(P \setminus s(A))$ and has a positive externality on the supervisor. Hence $s$ must be a cutoff strategy, with cutoff $p^*$. Thus far, points (1) and (4) are proven.

Substituting the kinked-linear penalty function into Eq. 1, the team's welfare reduces to

$$p^*\left(\frac{b}{2} - c\right) + \frac{b}{2}\sum_{a=0}^{p}\binom{p}{a}\lambda^a(1-\lambda)^{p-a}\min\{a, p^*\}. \tag{4}$$

Note that Eq. 4 is maximized at $p = M$. By contrast, $p^*$ has a positive effect on the second term but a negative effect in the first term (since $\frac{b}{2} - c < 0$). The second term, which we call the *truncated expectation*, has increasing differences in $p^*$ and $\lambda$, leading to the monotone comparative statics in point (2). Given that $p^*$ is increasing in $\lambda$, there will be empty promises whenever (i) using $p^* = 1$ gives a larger value in Eq. 4 than does $p^* = 0$, to avoid the degenerate case in which the optimal number of promises may as well be zero; and (ii) using $p^* = M - 1$ gives a larger value in Eq. 4 than does $p^* = M$, so that the cutoff is strictly smaller than the number of promises made. The interval in point (3) then follows from simple algebra. □

It is clear from Eq. 4 that the optimal cutoff is the smallest $p^*$ for which

$$\begin{aligned} p^*\left(\frac{b}{2} - c\right) + \frac{b}{2}\sum_{a=0}^{p}\binom{p}{a}\lambda^a(1-\lambda)^{p-a}\min\{a, p^*\} \\ \geq (p^*+1)\left(\frac{b}{2} - c\right) + \frac{b}{2}\sum_{a=0}^{p}\binom{p}{a}\lambda^a(1-\lambda)^{p-a}\min\{a, p^*+1\}. \end{aligned} \tag{5}$$

Let $\mu_{\lambda,M}(p^*+1) = \sum_{a=p^*+1}^{p}\binom{p}{a}\lambda^a(1-\lambda)^{p-a}$ be the binomial probability that the number of feasible tasks is at least $p^*+1$. Rearranging the above, the optimal cutoff is the first $p^*$ for which $\mu_{\lambda,M}(p^*+1) \leq \frac{c}{b/2} - 1$. It follows that $p^* \leq M\lambda$ when $\frac{c}{b/2} \geq \frac{3}{2}$. More generally, using a normal approximation to the binomial, $p^*/M \approx \lambda - O(M^{-3/2})$ for large $M$ (see Eq. 10 in the appendix).

## 3.2 Endogenous supervision

The features that agents should optimally employ cutoff strategies for promise completion, that those cutoffs are increasing in $\lambda$, and that the optimal penalty scheme is forgiving, are not specific to the worker-supervisor structure studied above. Indeed, the results of Theorem 1 extend to any contract with *complete monitoring*: for every promise of every agent, there is another agent who monitors that promise with probability one.

**Corollary 2.** *Consider a team of size $N$. Among contracts with complete monitoring, the optimal contract is such that each agent $i$ has a cutoff strategy for task completion, doing at most $p_i^*$ feasible tasks, and each agent's cutoff $p_i^*$ is increasing in $\lambda$. Optimal penalties are again kinked linear.*

There are many possible complete monitoring contracts. For example, suppose there are two workers (Alice and Bob) who each have 8 hours to work, and each task (performing or monitoring) takes 1 hour. Aside from a worker-supervisor contract under which Alice monitors 8 tasks and Bob promises 8 tasks (or vice-versa), other possible contracts include, for example, that Alice and Bob each promise 4 tasks and monitor 4 tasks, or that Alice does 2 tasks and monitors 6 tasks while Bob does 6 tasks and monitors 2 tasks. For $N \geq 3$, more intricate possibilities exist. As the following result shows, in the presence of empty promises these contracts are not payoff-identical.

**Theorem 2.** *In an optimal complete monitoring contract, the optimal division of labor is a worker-supervisor contract when there are two agents. When $N > 2$ agents, the optimal division of labor includes at least one supervisor (who specializes in monitoring). Among the workers who promise tasks, if $i$ makes more promises than $j$ ($p_i > p_j$) then $i$ also has a higher cutoff for task completion than $j$ ($p_i^* > p_j^*$). The optimality in this result is strict if and only if $\lambda$ is such that there are empty promises under the supervision contract.*

As seen in the proof below, optimal supervision endogenously arises for statistical reasons, despite the symmetry of players and tasks. It is particularly intuitive to consider the case of $M$ is even, and compare a worker-supervisor contract in which the worker makes an even number of empty promises $p^*$, to a symmetric contract in which each player makes $M/2$ promises and $p^*/2$ empty promises. In both cases the total numbers of promises and empty promises are the same. Suppose exactly $p^*$ tasks turn out to be feasible. In the worker-supervisor contract all of them will be completed, but in the symmetric contract all of them will be completed if and only if each player turns out to have exactly half of them. The fact that each player has a separate cutoff in the symmetric contract means there are two constraints to be satisfied, rather than just one as in the worker-supervisor contract. The same issue arises when comparing any two arbitrary contracts with the same total number of empty promises—whichever is the more asymmetric is superior.

*Proof.* Consider a putative optimal complete monitoring contract in which each agent $i$ makes $p_i \in \{0, 1, \ldots, M\}$ promises and has a cutoff $p_i^*$ for task completion. By complete monitoring, $\sum_{i=1}^{N} p_i = \frac{NM}{2}$. Let $p_{\text{sum}}^* = \sum_{i=1}^{N} p_i^*$. In analogy to Eq. 4, the team welfare is given by

$$\sum_{i=1}^{N} \left( p_i^* \left( \frac{b}{N} - c \right) + \frac{N-1}{N} b \sum_{a=0}^{p_i} \binom{p_i}{a} \lambda^a (1-\lambda)^{p_i - a} \min\{a, p_i^*\} \right). \tag{6}$$

Observe that when $p_i^* = p_i$ for every worker, the expected number of promises completed per worker is $p_i\lambda$, and it is irrelevant how those promises are allocated across workers. Consider the case where $p_k^* < p_k$ for some agent $k$, and imagine a different complete monitoring contract (with corresponding optimal penalties) where each agent $i$ makes $\tilde{p}_i$ promises and has a cutoff $\tilde{p}_i^*$, also having the property that $\sum_{i=1}^{N} \tilde{p}_i = \frac{NM}{2}$ and $p_{\text{sum}}^* = \sum_{i=1}^{N} \tilde{p}_i^*$. Could this contract welfare-dominate the putative optimal contract? Note that since the sum of cutoffs is the same, the ranking of the two contracts is determined by the truncated expectations ($\sum_{a=0}^{p_i} \binom{p_i}{a} \lambda^a (1-\lambda)^{p_i-a} \min\{a, p_i^*\}$) in Eq. 6. Note that each truncated expectation is supermodular in $p$ and $p^*$. This is because the condition for increasing differences reduces to

$$\sum_{a=p^*+1}^{p+1} \binom{p+1}{a} \lambda^a (1-\lambda)^{p+1-a} - \sum_{a=p^*+1}^{p} \binom{p}{a} \lambda^a (1-\lambda)^{p-a} > 0,$$

which holds because making more promises leads to a first-order stochastic improvement in the number of feasible tasks. Since the truncated expectation is zero when $p_i = p_i^* = 0$, supermodularity implies superadditivity. In particular, when $N = 2$, superadditivity immediately implies that the putative optimal contract is dominated unless it is a worker-supervisor contract. For the case $N > 2$, note first there must be at least one pair of agents $i, j$ for whom $p_i + p_j < M$, else the contract would violate complete monitoring. This again ensures that there must be an agent specializing in monitoring. Moreover, the rearrangement inequality of Lorentz (1953) for supermodular sums implies that to maximize Eq. 6, it must be that $p_i \geq p_j$ implies $p_i^* \geq p_j^*$.[13] □

There are two different ways to interpret complete monitoring. Under one interpretation, units of capacity are not substitutable across performance and monitoring. Rather, there are $\frac{NM}{2}$ performance units and $\frac{NM}{2}$ monitoring units, and the problem is to allocate these units within the team. Under a second interpretation, units of capacity are perfectly substitutable across performance and

[13] The Lorentz-Fan rearrangement inequality for supermodular sums says if $f : \mathbb{R}^k \to \mathbb{R}$ is a supermodular function, then for any collection of vectors $(x^1, \ldots, x^n)$, the sum $\sum_{i=1}^{n} f(x^i)$ is smaller than or equal to the sum $\sum_{i=1}^{n} f(x^{*i})$, where $x^{*i}$ is the "majorized" vector which, for every dimension $k$, contains the $i$th largest component among $x_k^1, \ldots x_k^n$. That is, majorizing yields a larger sum.

monitoring, and complete monitoring arises when the two functions receive equal allocations. In this second interpretation, complete monitoring is an *ad hoc* constraint. The following section relaxes this constraint, allowing the team to monitor less in order to accomplish more.

# 4 Trading off performance and monitoring

In this section, we examine the optimality of empty promises when units of capacity are substitutable between performance and monitoring. For example, rather than monitor all the worker's performance tasks, the supervisor in the previous section could use some of his units of capacity towards performing tasks—in which case the worker would need to monitor some tasks as well. In particular, they may wish to allocate more units of capacity to performing tasks than to monitoring tasks (else, in view of Theorem 2, their original promise scheme was optimal). With too little monitoring, however, they may not be able to implement finely-tuned punishment schemes. These concerns raise several questions. How much capacity should they devote to monitoring? How much capacity should they devote to empty promises, and how much to promises that they intend to fulfill? Does it matter how monitoring responsibility is distributed?

We begin by extending the ideas of Section 3 to this setting. There, we minimized expected punishment by making all incentive constraints for a putative optimal strategy bind, and showed that the strategy must then be a cutoff strategy. Both realized and expected punishments had a kinked-linear form: they were zero up to a threshold, and increased linearly thereafter. Such an incentive scheme is achievable under complete monitoring because every task of every agent is monitored, so that there are sufficiently many degrees of freedom in the penalty scheme.

In fact, the same penalty scheme can be achieved by allocating only $M$ units of capacity towards monitoring, regardless of the number of agents in the team. The same expected punishment (and cutoff strategies) optimally arise, simply by employing a correlated randomization device to determine which player to monitor using all the other players' monitoring capacity, and scaling the punishment scheme appropriately. That is, for each player $i$ there is an $\alpha_i \in (0, 1]$ such that with probability $\alpha_i$ all of $i$'s tasks are monitored , and with probability $1 - \alpha_i$ none of her tasks are monitored. Then her new punishment schedule is $v_i$ is multiplied by $1/\alpha_i$, so that, taking expectations over whether she will be monitored, her expected punishment schedule is exactly $v_i$. We call this *probabilistically complete monitoring*, since either all of player $i$'s tasks are monitored or none of player $i$'s tasks are monitored.[14]

[14]More formally, monitoring is probabilistically complete if, for each realization $R \in \times_{i=1}^{N} \mathcal{PS}(\bigcup_{j \neq i} P_j)$ in the support of the monitoring strategy $r$, and for each $i$, there exists $\alpha_i \in (0, 1]$ such that $(\bigcup_{j \neq i} R_j) \cap P_i$ (the subset of $P_i$ that is monitored) is $P_i$ with probability $\alpha_i$ and is $\varnothing$ with probability $1 - \alpha_i$.

Theorem 3, below, uses a continuous, normal approximation to the binomial distribution to show that if probabilistically complete monitoring is feasible then having a single supervisor is approximately optimal (leaving open the conjecture that it is in fact exactly optimal).[15]

**Theorem 3.** *Consider any capacity size $M$, number of agents $N$, and cost-benefit ratio $\frac{c}{b}$. Under probabilistically complete monitoring,*

1. *The results of Corollary 2 on the optimality of kinked-linear penalty schedules, cutoff strategies, and the increasingness of $p_i$ and $p_i^*$ in $\lambda$ for each agent $i$ continue to hold.*
2. *For $N = 2$, all monitoring responsibilities should be allocated to a single agent. For $N > 2$ and any $\epsilon > 0$ there exists $\underline{M} < \infty$ such that for any $M \geq \underline{M}$ a contract with one supervisor and $N - 1$ agents each making $M$ promises and using a cutoff*

$$p^* = \left\lceil M\lambda - \sqrt{2M(1-\lambda)\lambda}\ \mathrm{erf}^{-1}\left(\frac{b + bN - 2cN}{b - bN}\right)\right\rceil \tag{7}$$

*attains a $1 - \epsilon$ fraction of the welfare of an optimal contract.*[16]

This simple observation bounds the resources devoted to monitoring when monitoring may be stochastic, regardless of whether monitoring is probabilistically complete. Devoting more resources toward monitoring cannot improve over complete monitoring with regard to incentives, while at the same time it reduces the capacity available for performing tasks.

**Corollary 3.** *When units of capacity are substitutable between performance and monitoring, the optimal contract would allocate at most $M$ units of capacity to monitoring tasks.*

Even probabilistically complete monitoring is restrictive—it is possible to reduce the resources devoted to monitoring even further, increasing the capacity devoted to promises. For the special case of promise-keeping contracts, monitoring can be minimized in the following manner. Two agents become partial supervisors, who each monitor one task and make $M - 1$ promises. A correlated randomization device determines whether, with probability $1/2$, each supervisor monitors the

[15] Depending on when capacity must be allocated, correlation effects may further favor having a single supervisor under probabilistically complete monitoring. Consider a team of three agents (Alice, Bob, and Carol), with $M = 4$. Suppose Alice promises four tasks, and Bob and Carol each promise two tasks and monitor two tasks. With probability $1/2$, Bob and Carol each monitor two (different) tasks Alice promised; and with probability $1/2$, Bob and Carol each monitor each other. But if Bob learns at $t = 1$ that he will monitor Alice, he will have no incentive to complete any tasks, since he will know that Carol will monitor Alice as well. Theorem 3 shows that a single supervisor optimally arises even if the realization of the monitoring scheme can be kept secret from the agents until $t = 3$.

[16] The function $\mathrm{erf}^{-1}$ is the inverse of the "error function" $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} \mathrm{d}t$ obtained when integrating the normal distribution.

other; with the remaining probability, the supervisors combine their monitoring capacity to monitor a uniformly chosen worker among the $N-2$ other agents, each of whom make $M$ promises. The penalty schedule is linear so that all agents are willing keep all their promises, and all incentive constraints bind. We call this form of contract "promise keeping with minimal monitoring."

For contracts with empty promises, however, if monitoring is not probabilistically complete then it may be impossible to make all relevant incentive constraints bind. If fewer tasks are being monitored than an agent promised, the punishment schedule generally has too few degrees of freedom relative to the task-completion strategy. Compounding this problem, there is a gap between the punishment schedule and expected punishments: conditional on the tasks a player completed, inspection outcomes depend probabilistically on the monitoring distribution. Because punishments are restricted to be negative, it may be impossible to generate the expected punishments that would make a given combination of incentive constraints bind. In fact, we have numerical examples where the optimal contract does not have kinked-linear penalties and cutoff strategies. Theorem 5 in the supplemental appendix shows that these two features do remain optimal when the magnitude of punishment is restricted to be increasing and convex in the number of failed inspections, as might be the case in settings where punishments are imposed by third parties who are more inclined to exact punishment if they perceive a consistent pattern of failures.[17] Without restricting the contract space, however, it will generally be impossible to solve analytically for the optimal penalty schedule.

Nonetheless, we are able to show that optimal contracts do involve empty promises for an intermediate range of $\lambda$.

**Theorem 4.** *Consider any capacity size M and number of agents N. In an optimal contract,*

1. *If $\lambda$ is sufficiently low, agents don't complete any tasks.*
2. *If $\lambda$ is sufficiently high, agents make as many promises as possible and keep all of them, while monitoring is minimal: two "partial supervisors" each monitor one task.*
3. *For an intermediate range of $\lambda$'s, agents optimally make empty promises. In particular, for any M, b, c and N, there is $\lambda^* \in (0,1)$ such that for any $\lambda \leq \lambda^*$, empty promises under*

[17] This is proved in the supplemental appendix, where we show that player $i$'s conditional expected punishment will be convex in the number of her promises she fails to fulfill. This implies that if player $i$ prefers completing $\tilde{p}$ promises over $\tilde{p}-1$ promises, then she must also prefer completing $\tilde{p}-k$ promises over $\tilde{p}-(k+1)$ promises for all $k = 1, \ldots, \tilde{p}-1$. We then use a duality argument to show that the Lagrange multipliers for the convexity constraints imply a recursion that can be used to solve for the optimal expected punishment. That expression can be written in terms of the expected number of discovered unfulfilled promises above a threshold, and is implementable by a kinked-linear punishment schedule.

*probabilistically complete monitoring with one supervisor dominates promise-keeping with minimal monitoring (including in an interval where promise-keeping yields positive social utility). If the cost-benefit ratio of tasks is moderately high, then empty promises may be optimal even when $\lambda$ is very close to one: for any $M$ and $N$, if $\frac{c}{b} > \frac{2}{2+e} + \frac{1}{N}\frac{e}{2+e}$ ($\approx .42$ for large $N$), empty promises are optimal even in a neighborhood of $\lambda = \frac{M-2}{M-1}$. Finally, there is also an open range of parameters for which promise-keeping with minimal monitoring yields positive social utility but is dominated by making roughly half as many promises and keeping just one.*

Intuitively, when $\lambda$ is very low, players expect very few promises to be feasible. Rather than risk incurring punishments, it is better not to do any tasks at all. Therefore it is irrelevant whether agents make any promises in the first place. When $\lambda$ is very high, the players expect nearly all their promises to be feasible. Then it is optimal to maximize the number of promises made and complete as many as possible. This is implemented at lowest cost via a linear contract under which only one task is monitored by a "partial supervisor". Since he has $M-1$ units of capacity left for making promises, a second partial supervisor is needed to monitor him. When $\lambda$ is intermediate, there is generally a gap between the highest number of promises a player is willing to complete, and the number of promises she actually made. This is illustrated in Figure 1, where promise keeping is dominated for nearly any $\lambda$ by the envelope of some simple empty-promise contracts with two partial supervisors.

Why are empty promises beneficial? Without changing the number of promises that she plans to fulfill, by making more empty promises a player attains a first order stochastic improvement in the number of tasks she will complete according to her plan. However, the corresponding increase in the number of promises she leaves unfulfilled will lead her to expect a more severe punishment unless the contract is *forgiving*: if it does not punish her when the other players find only a "small" number of her unfulfilled promises. A forgiving contract allows players to use empty promises to buffer against task infeasibility, and minimizes punishment in the very likely event that players are unable to complete all of their promises. By contrast, punishments under promise-keeping are quite severe. Depending on task feasibility, a contract that enforces promise-keeping can be socially dominated by a forgiving contract under which players keep just a single promise. As seen from Figure 1, probabilistically complete monitoring—under which the maximal amount of resources is devoted to monitoring—dominates promise keeping when $\lambda$ is relatively low. This is because when $\lambda$ is low, agents are most likely to be punished due to task infeasibility.

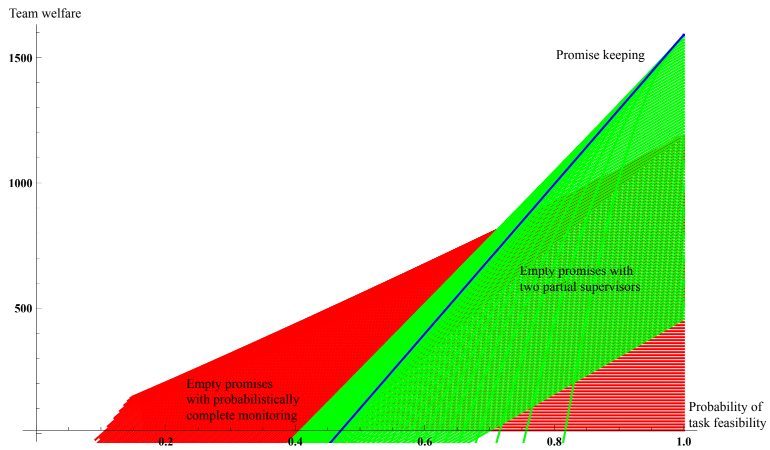


Figure 1: Consider capacity $M = 100$, number of players $N = 4$, task benefit $b = 10$, task cost $c = 6$. The blue line corresponds to the social value of promise keeping. The green region is generated from the social value functions (as a function of $\lambda$) for each $p^* \in \{1, \ldots, M - 1\}$ of a partial supervision contract with two promise-keeping supervisors who supervise workers using a cutoff of $p^*$ for task completion. The red region is generated from the social value functions of probabilistically complete monitoring with one supervisor and a cutoff of $p^*$ for workers for each $p^* \in \{1, \ldots, M - 1\}$.

# 5 Extensions

## 5.1 Costly monitoring

The analysis thus far assumed that monitoring is costless, and therefore agents are indifferent over whether to monitor each other. If, however, monitoring requires nonverifiable, costly effort, the question of "who monitors the monitor" arises. Rahman (2010) shows that to provide incentives for monitoring, agents should occasionally shirk just to "test" the monitor. Since our model already generates optimal shirking (in the form of empty promises), we set monitoring costs to zero to highlight the fact that shirking also arises from an entirely different mechanism. Adapting Rahman's argument, as follows, shows that the contracts we construct are robust to monitoring costs, without requiring any additional shirking.

Suppose that monitoring is costly. A monitor can always claim that a task passed inspection, but must exert effort to prove that a task failed his inspection. To induce him to exert monitoring effort, the team can add an additional stage, $t = 4$, to their interaction. After the penalties for failed tasks are implemented in $t = 3$, in $t = 4$ each agent reports which tasks he himself completed. Agents are not punished for these reports, and are therefore willing to report truthfully. Whenever an agent reveals an uncompleted task in $t = 4$ that was not reported as failing inspection in $t = 3$, whichever teammate (if any) was supposed to monitor that task is punished. Because task feasibility is random, even under promise-keeping there is positive probability that some tasks were not completed. Therefore a sufficiently large punishment induces faithful monitoring, and need not be incurred in equilibrium.

## 5.2 Messages that economize on monitoring

A recent literature studies the benefits of messages in contract design under private information.[18] In our model, incorporating messages can reduce the amount of monitoring needed. Consider the case of probabilistically complete monitoring, where all the tasks an agent promises are monitored. Matsushima, Miyazaki, and Yagi (2010) suggest that the principal should require an agent with private information to work on a certain number of tasks, which the agent should announce to the principal. Adapting this idea to our setting, we find that the same task completion strategies studied in earlier sections can be implemented using fewer than $M$ units of capacity for monitoring. To see this, suppose an agent promises the set of tasks $P$ and his task completion strategy is $s$. Modify the contract to allow the agent to tell the other agents which $p^*$ of his tasks to monitor, where $p^* = |s(P)|$ is the largest number of tasks he would ever complete. Clearly, the agent will include in his report all the tasks he has completed. Thus, no more than $p^*$ tasks need to be monitored. Note that if there are monitoring costs, the method in Section 5.1 for "monitoring the monitor" remains feasible since there is positive probability that fewer than $p^*$ tasks were feasible.

For probabilistically complete monitoring schemes with messages, more resources can be devoted to performing tasks while still maintaining the optimal expected punishments and cutoff strategies. The optimal cutoff balances the resulting tradeoff between reducing the amount of monitoring and increasing the number of tasks completed. However, because the opportunity cost of empty promises is reduced for any given $\lambda$, promise-keeping becomes even less attractive than before. Once again, different allocations of supervisory responsibility will not be welfare-equivalent under empty promises. With messages, a "supervisor" would have some unused units of capacity which may optimally be allocated towards completing tasks. Since another agent must monitor him, optimal supervision structures with two partial supervisors in a team of $N$ agents may arise.

## 5.3 Imperfect monitoring

We have assumed that when a shirked task is monitored, it will fail inspection with probability one. What if a shirked task that is monitored fails inspection with probability $0 < \gamma < 1$? In this case,

[18] For example: Chakraborty and Harbaugh (2007); Frankel (2010); Jackson and Sonnenschein (2005); Matsushima, Miyazaki, and Yagi (2010).

Theorem 4 continues to hold independently of $\gamma$, while the characterization of the optimal contract for the case of probabilistically complete monitoring continues to hold for $\gamma$ not too far from one.

Theorem 4 is unaffected by imperfect monitoring because the expected social welfare of the contract shown to dominate promise keeping is independent of $\gamma$. This contract is such that the penalty schedule depends on the number of failed inspections, and punishes only when the maximal number of failures is found. Letting $F$ denote the number of tasks of a player that are monitored under this contract, the probability that $F$ failures are found when $a$ tasks are completed and $p$ are promised is given by $\gamma^F \binom{p-a}{F} / \binom{p}{F}$. The expected punishment conditional on completing $a$ tasks is then given by $v(F)\gamma^F \binom{p-a}{F} / \binom{p}{F}$, which can be made independent of $\gamma$ by scaling the penalty $v(F)$ by the factor $\frac{1}{\gamma^F}$.

Our characterization of the optimal contract under probabilistically complete monitoring relies on being able to find a penalty schedule under which the expected punishments make all relevant incentive constraints bind. When the monitoring technology is sufficiently imperfect, such a penalty schedule may not exist. Consider the simple case in which an agent makes three promises, all of which are monitored. An uncompleted task fails inspection with probability $\frac{1}{4}$. Suppose for simplicity that $\frac{b}{N} - c = 1$. To induce a cutoff of $p^* = 2$, the optimal *expected* punishments would be 0 whenever at least two tasks are completed, $-1$ if one task is completed, and $-2$ if no task is completed. Taking into account the probability of failing inspection, the penalty schedule $(v(0), v(1), v(2), v(3))$ should satisfy

$$\begin{pmatrix} \frac{27}{64} & \frac{27}{64} & \frac{9}{64} & \frac{1}{64} \\ \frac{9}{16} & \frac{3}{8} & \frac{1}{16} & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v(0) \\ v(1) \\ v(2) \\ v(3) \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \\ 0 \\ 0 \end{pmatrix},$$

where the $a$-th row and the $f$-th column of the $4 \times 4$ matrix corresponds to the probability that $f$ failures will be found when $a$ tasks are completed. The unique solution to this system sets $v(0) = v(1) = 0$, $v(2) = -16$, and $v(3) = 16$. That is, the agent would have to receive a *reward* of 16 if the maximal number of failures are found. The difficulty here is that described by Farkas' Lemma: there is not always a negative solution to a linear system. By continuity, however, since a solution exists when the technology is perfect, one exists when the technology is not too imperfect. Indeed, in the example above one can find a penalty schedule obtaining the desired expected punishments for any $\gamma$ larger than about $1/3$.

# 6 Teams within firms

When a team is embedded within a larger firm, the agents on the team may not directly benefit from the tasks they complete. The firm may be able to offer contractual bonuses that depend on the team's performance, but not on the performance of individual team members. At the level of individual performance, only wasteful punishments (peer pressure, separation, etc.) are available. In this section we show that the firm's problem of designing an optimal contract in this environment is similar to the problem the agents would face if they were partners, as characterized in the previous sections.

The firm hires a team of $n$ agents to perform tasks and monitor each other. The firm reaps the entire benefit $B$ from each task, but cannot observe who performed it. Agents have limited liability in terms of money, but can suffer from wasteful punishments. For comparison to earlier results, we assume that $c < B < Nc$. The firm makes the agents a take-it-or-leave-it offer comprising:

1. A fixed ex ante payment $t_i \geq 0$ for each $i$;
2. A bonus $b \geq 0$, paid to the team for each completed task, and split equally among the agents so that each receives $b/N$;
3. A randomization over *contracts* (promise schemes, task completion strategies, monitoring schemes, and punishment schemes, as defined in Definition 1).

Each agent accepts the offer if and only if her expected utility from the offer is at least as high as her exogenous outside option. Each agents' incentive compatibility constraint is simply Eq. 2. As discussed in Remark 2, even if the firm's offer puts some agents into supervisory roles and others into productive roles, by randomizing the agents' roles after they accept the contract it can satisfy their individual rationality constraints as long as their expected utilities sum to at least zero. The firm's objective function is

$$\sum_{i=1}^{N}\Big(-t_i + \sum_{A\subseteq P_i} \lambda^{|A|}(1-\lambda)^{p_i-|A|}|s_i(A)|(B-b)\Big) \tag{8}$$

By the usual argument, IR must bind in an optimal contract.[19] Substituting the binding IR

[19]Consider a non-degenerate contract (in which agents complete some tasks) for which limited liability binds the bonus ($b = 0$) and IR is slack. The ex ante payments must be greater than zero (otherwise only a degenerate contract would be individually rational). But the firm can benefit from reducing the ex ante payments to zero, making it up to the agents by increasing the bonus to compensate. (When ex ante payments are zero, IR implies that the bonus satisfies $b \geq c$.) But since an increase in the bonus strengthens the agents' incentives, the firm can induce the same task performance at lower expected cost. Therefore the ex ante payments must be zero and the bonus must be nonzero. Further, IR cannot be slack, since the firm could impose marginally harsher punishments in order to marginally reduce the bonus.

constraints into Eq. 8 reduces the firm's objective function to the team's objective function (Eq. 1) but with $B$ in place of $b$. That is, when the firm hires the agents as a team, for any fixed bonus $b$ its optimal contract is exactly the same as the optimal contract for the agents if they were partners. However, now the bonus is also a choice variable rather than a parameter.

A bonus based on team output, naturally, is a crude instrument for providing incentives, since each team member receives $b/N$ whenever any team member completes a task. Since profitability for the firm requires $b < B$, and yet $B/N < c$, the bonus alone (i.e., without punishments) cannot motivate the agents to perform and still yield positive profit for the firm. Hence if the firm's optimal contract is non-degenerate, it must employ both a nonzero bonus and a non-degenerate punishment scheme. Observe, however, that as the team gets larger it becomes more and more expensive to use the bonus for motivation. Indeed, since $b$ is bounded above by $B$ regardless of $N$, the bonus loses its motivational power in the limit as $N \to \infty$. At this limit, only punishments provide incentives, so the bonus might as well be replaced by a fixed ex ante payment, since its only purpose is to meet the agents' IR constraints.

These characteristics are consistent with the stylized facts identified by Baker, Jensen, and Murphy (1987)—individual financial incentives are rare—and Oyer and Schaefer (2005)—broad-based group incentives are common. According to the model, these contractual features are optimal when the firm cannot formally monitor employees at the individual level, and must supplement its formal incentives at the team level with peer monitoring and informal punishments at the individual level. That is, industries where production is complex and requires accumulated job-specific human capital, as discussed in the Introduction.

As for the form of the firm's optimal contract, since $b \leq B$ is required for profitability, the conclusions of Theorem 4 still hold—empty promises arise for an intermediate range of $\lambda$. For settings like Figure 1, empty promises dominate promise keeping unless $\lambda$ is very high. Empirically, since punishments may be implemented informally, they may be hard to observe. In practice, since the agent is not punished at all until his results fall below a threshold, and even then he may be punished in a way that is not evident to an outside observer, we should expect formal punishments to be enacted quite rarely.

# 7 Discussion

We study a model of teams in which agents optimally make empty promises, and are "forgiven" for having done so. Empty promises buffer against the potential infeasibility of tasks. This equilibrium phenomenon is robust to a tradeoff between performance and monitoring, even though empty promises use up a unit of capacity that could otherwise be allocated towards having a finer, more attenuated monitoring scheme.

Our model endogenously gives rise to optimal supervisory structures, despite the fact that all agents and tasks are identical. Although there is no inherent complementarity in task completion, statistical complementarities arise from the optimal task completion strategies. Simply stated, there are increasing returns to a worker's task load when he is making empty promises: doubling his number of promises and cutoff for task completion more than doubles his social contribution. Consequently, it is best to have one agent do all the "working" and the other agent all the "supervising" rather than have mixed roles. Under the assumption of unbounded liability, this intuition implies that there should be at most two supervisors, no matter how large the team. More realistically, a bound on liability would yield a lower bound on the ratio of supervisors to workers.

Introducing asymmetries into the model, even with complete information, may lead to additional interesting predictions. Suppose, for example, that the probability of task feasibility $\lambda$ is player-specific. Then the least capable player should be performing as few tasks as possible, and using his resources towards supervision instead. This accords with the "Dilbert principle," which suggests that less productive team members should become supervisors (Adams, 1996).

While the capacity constraints in our model serve the technical purpose of ensuring an optimal solution, they are also amenable to a bounded rationality interpretation. Although it is commonly assumed in contract theory that an agent's memory has unlimited capacity and perfect recall, evidence from psychology shows that working memory is both sharply bounded and imperfect.[20] One interpretation for the limiting resource is a bound on the number of tasks an agent can remember. A task in this view contains detailed information, such as a decision tree, that is necessary to complete it properly.[21] Imperfect task feasibility may arise from being unable to remember all the necessary details for proper task completion. When tasks are complex, it may be impossible

[20] A seminal paper by Miller (1956) suggests that the capacity of working memory is approximately $7 \pm 2$ "chunks." A chunk is a set of strongly associated information—e.g., information about a task. More recently, Cowan (2000) suggests a grimmer view of $4 \pm 1$ chunks for more complex chunks. The economic literature studying imperfect memory includes Dow (1991), Piccione and Rubinstein (1997), Hirshleifer and Welch (2001), Benabou and Tirole (2002) and Wilson (2004). Mullainathan (2002) and Bodoh-Creed (2010), in particular, study updating based on data recalled from long-term memory. There is also a literature on repeated games with finite automata which can be interpreted in terms of memory constraints (e.g., Cole and Kocherlakota, 2005; Compte and Postlewaite, 2008; Piccione and Rubinstein, 1993; Romero, 2010), as well as work on self-delusion in groups (e.g., Benabou, 2008).

[21] Al-Najjar, Anderlini, and Felli (2006) characterize finite contracts regarding "undescribable" events, which can be fully understood only using countably infinite statements. In this interpretation, to carry out an undescribable task properly, a player must memorize and recall an infinite statement. The related literature considers contracts with bounded rationality concerns relating to complexity—such as limitations on thinking through or foreseeing contingencies (e.g., Bolton and Faure-Grimaud, Forthcoming; Maskin and Tirole, 1999; Tirole, 2009), communication complexity (e.g., Segal, 1999), and contractual complexity (e.g., Anderlini and Felli, 1998; Battigalli and Maggi, 2002).

to fully specify their details in a convenient written form, such as a contract. As noted by (Aoki, 1988, p. 15), “the experience-based knowledge shared by a team of workers on the shopfloor may be tacit and not readily transferable in the form of formal language.” Without a convenient way to fully specify a task, an agent who promises to perform the task must expend memory resources to store the relevant details. Moreover, another agent may need to expend resources to store those details in order to be able to monitor him, leading to a tradeoff between performance and monitoring as in Section 4. Coping with multiple complex tasks “may require more versatile workers’ skills (deeper and broader information-processing capacities), which have not been considered essential in traditional hierarchies” (Aoki, 1988, p. 31).

# Appendix

*Proof of Lemma 1.* We prove each point below.

1. Suppose there is an optimal contract in which the promise scheme is not deterministic. However, since the promise scheme is realized publicly, there is an equally good contract that assigns probability 1 to whichever realization yields the highest welfare.
2. First, an agent has no influence over whether other agents' tasks pass or fail inspection. So for any punishment scheme that depends on other agents' outcomes, it is equally effective to employ a modified punishment scheme in which the agent's penalty is conditioned only on his own outcomes, where this penalty is set to offer the same conditional expected penalty as the original contract. Second, conditional on which of his promised tasks fail inspection, an agent has no influence over which of his tasks pass inspection—passed inspections depend entirely on how the other agents monitor him. Specifically, fix a set of tasks that fail inspection, and suppose the agent considers completing an additional task. For monitoring realizations in which that task is monitored, he reduces the number of tasks that fail inspection. For monitoring realizations in which that task is not monitored, he does not affect how many tasks fail or pass inspection. Therefore the agent's incentives under a contract that depends on both failed and passed inspections can be replicated by a contract that, conditional on failed inspections, offers the same penalty regardless of passed inspections, where this penalty is set to offer the same conditional expected penalty as the original contract.
3. Suppose to the contrary that there is an optimal contract in which, when $A \subseteq P_i$ tasks are feasible, an agent is supposed to complete $s_i(A) \subset A$ tasks, but is indifferent between completing $s_i(A)$ tasks and completing $A' \subseteq A$ tasks, with $|A'| > |s_i(A)|$. But then there exists a superior contract, otherwise unchanged, in which he simply completes $A'$ tasks whenever $A$ tasks are feasible—he is no worse off himself, and his team members are strictly better off. □
4. By revealed preference, $s_i(s_i(A)) = s_i(A)$, so it suffices to show that $A \subset A'$ implies $|s_i(A)| \leq |s_i(A')|$. Suppose to the contrary that $A \subset A'$ and $|s_i(A)| > |s_i(A')|$. Since upward incentive constraints are slack, the agent must strictly prefer $s(A')$ over $s(A)$. Furthermore, by revealed preference, $s_i(A') \nsubseteq s_i(A)$. But then a superior contract can be constructed by slightly relaxing the penalty when tasks in $s_i(A') \setminus s_i(A)$ fail inspection. The agent, facing a strictly milder punishment schedule, is strictly better off. For a sufficiently small relaxation of the penalty, incentive compatibility still holds.

*Proof of Theorem 3.* Part (i) is clear from the discussion in the text. For part (ii), when $N = 2$, Theorem 2 on complete monitoring applies. We henceforth consider the case $N > 2$. By the

De Moivre–Laplace theorem (Johnson, Kemp, and Kotz, 2005, Eq. 3.20), the normal distribution with mean $p\lambda$ and variance $p(1-\lambda)\lambda$ approximates the binomial distribution with $p$ tasks each with $\lambda$ probability of being feasible. The approximation error in the CDF at any point is no greater than the order of $\sqrt{p\lambda(1-\lambda)}$.

Using this approximation, we define the *continuous problem* of choosing promises $\tilde{p}_i \in \mathbb{R}$ and earnest promises $\tilde{p}_i^* \in \mathbb{R}$ to solve

$$\max_{\{\tilde{p}_i, \tilde{p}_i^* \in \mathbb{R}\}_{i=1}^N} \sum_{i=1}^n \mathbb{E}\big((b-c)\min\{a, \tilde{p}_i^*\} + (\frac{b}{N} - c)\max\{\tilde{p}_i^* - a, 0\}\big)$$
$$\text{s.t. } \tilde{p}_i^* \leq \tilde{p}_i \leq M \text{ for all } i \text{ and } \sum_i \tilde{p}_i \leq M(N-1),$$

where expectation of $\tilde{p}_i^*$ is taken with respect to the normal distribution $\mathcal{N}\big(\tilde{p}_i\lambda, \tilde{p}_i(1-\lambda)\lambda\big)$. We write the objective function as $\sum_i E_i$, where $E_i \equiv \mathbb{E}\big((b-c)\min\{a, \tilde{p}_i^*\} + (\frac{b}{N} - c)\max\{\tilde{p}_i^* - a, 0\}\big)$.

First, we solve the inner part of the continuous problem—optimizing $\tilde{p}_i^*$ given $\tilde{p}_i$. The first order condition is

$$\frac{\partial E_i}{\partial \tilde{p}_i^*} = \frac{b + bN - 2cN + b(N-1)\operatorname{erf}\Big(\frac{\lambda\tilde{p}_i - \tilde{p}_i^*}{\sqrt{2\tilde{p}(1-\lambda)\lambda}}\Big)}{2N} = 0, \tag{9}$$

where $\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\mathrm{d}t$. The first order condition is solved at

$$\tilde{p}_i^* = \tilde{P}^*(\tilde{p}_i) \equiv \tilde{p}_i\lambda - \sqrt{2\tilde{p}_i(1-\lambda)\lambda}\,\operatorname{erf}^{-1}\left(\frac{b + bN - 2cN}{b - bN}\right), \tag{10}$$

and the welfare arising from each agent $i$ is

$$\tilde{p}_i\lambda(c-b) - \frac{be^{-\operatorname{erf}^{-1}\left(\frac{b+bN-2cN}{b-bN}\right)^2}(N-1)}{N\sqrt{2\pi}}\sqrt{\tilde{p}_i\lambda(1-\lambda)}. \tag{11}$$

The strict second order condition is satisfied globally:

$$\frac{\partial^2 E_i}{\partial(\tilde{p}_i^*)^2} = -\frac{b(N-1)e^{-\frac{(\lambda\tilde{p}_i - \tilde{p}_i^*)^2}{2\tilde{p}(1-\lambda)\lambda}}}{N\sqrt{2\pi\tilde{p}(1-\lambda)\lambda}} < 0.$$

We now move to the outer part of the continuous problem: choosing $p_i$. By the envelope theorem,

$$\frac{dE_i}{d\tilde{p}_i}\Big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)} = \frac{\partial E_i}{\partial \tilde{p}_i}\Big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)},$$

$$\frac{d^2E_i}{d\tilde{p}_i^2}\Big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)} = \Big(\frac{\partial^2 E_i}{\partial \tilde{p}_i^2} + \frac{\partial^2 E_i}{\partial p_i \partial \tilde{p}_i^*}\frac{d\tilde{P}^*}{d\tilde{p}_i}\Big)\Big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)}.$$

Solving the closed form of $d^2E_i/dp_i^2$ at $\tilde{p}_i^* = \tilde{P}^*(\tilde{p}_i)$ demonstrates that the objective is strictly convex in each $\tilde{p}_i$:

$$\frac{d^2E_i}{dp_i^2}\Big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)} = \frac{be^{-\mathrm{erf}^{-1}\left(\frac{b+bN-2cN}{N-bN}\right)^2}(N-1)\sqrt{\tilde{p}_i(1-\lambda)\lambda}}{4N\tilde{p}_i^2\sqrt{2\pi}} > 0.$$

Since in addition $dE_i/d\tilde{p}_i > 0$ at $\tilde{p}_i^* = \tilde{P}^*(\tilde{p}_i)$, and probabilistically complete monitoring requires $\sum_i \tilde{p}_i \leq (N-1)M$, it follows that the optimal promise scheme in the continuous problem is for $N-1$ agents each to promise $\tilde{p}_i = M$ tasks and complete $\tilde{p}^* \equiv \tilde{P}^*(M)$ of them, while the $N$th agent promises zero tasks (i.e., specializes in monitoring).

Now we construct a contract for the true (discrete) model, using the same promise scheme: $N-1$ agents each make $M$ promises, while the $N$th agent supervises. The number of earnest promises must of course be an integer, so we round $\tilde{p}^*$ up to the next integer, $\lceil \tilde{p}^* \rceil$. Let $\hat{V}$ be the welfare attained by this discrete contract, and let $\tilde{V}$ be the value of the continuous problem. The difference $\hat{V} - \tilde{V}$ arises from four issues.

1. The "tail benefit": The discrete contract applies to a distribution with a lower bound of zero feasible tasks, and so does not involve the harsh punishments that arise in the long lower tail of the continuous problem.
2. The "integer benefit": The maximum number of tasks accomplished is greater under the discrete contract than in the solution to the continuous problem, leading to higher social payoffs for realizations with many feasible tasks.
3. The "integer deficit": Because only whole tasks can be performed under the discrete contract, when fewer than $\lceil \tilde{p}^* \rceil$ tasks are performed the actual punishment may be greater than in the solution to the continuous problem.
4. Approximation error: The CDF of the normal distribution at $a + \frac{1}{2}$ is only an approximation of the binomial CDF at $a$.

Let $\delta = \frac{N-1}{N}b$ and $\rho = (b-c)$. Let $\Phi$ and $\varphi$ be the CDF and PDF of the normal distribution, and $\hat{\Phi}$ and $\hat{\varphi}$ be the CDF and PDF of the binomial. The tail benefit (which is not affected by approximation error) is

$$X = -\int_{-\infty}^{-1/2} \big((\delta-\rho)(\lceil\tilde{p}^*\rceil - \tilde{p}^*) + \delta a\big)\varphi(a)\,da.$$

The integer benefit, accounting for approximation error, is at least

$$Y = \rho\Big(\big(1-\hat{\Phi}(\lceil\tilde{p}^*\rceil - 1)\big)\lceil\tilde{p}^*\rceil - \big(1-\Phi(\lceil\tilde{p}^*\rceil - \tfrac{1}{2})\big)\tilde{p}^*\Big).$$

The integer deficit, accounting for approximation error, is

$$Z = \sum_{a=0}^{\lceil\tilde{p}^*\rceil-1}\left(\int_{a-1/2}^{a+1/2}\big((\delta-\rho)(\lceil\tilde{p}^*\rceil - \tilde{p}^*) + \delta\tilde{a}\big)\varphi(\tilde{a})\,d\tilde{a} - \delta a\hat{\varphi}(a)\right).$$

Combining terms and collecting $(\lceil\tilde{p}^*\rceil - \tilde{p}^*)$ yields the deficit in welfare yielded by each of the $N-1$ task-performing agents under the discrete contract, compared to the value of the continuous problem:

$$\begin{aligned}
-\frac{\hat{V}-\tilde{V}}{N-1} = Z - X - Y = &-(\lceil\tilde{p}^*\rceil - \tilde{p}^*)\big(\rho - \delta\Phi(\lceil\tilde{p}^*\rceil - \tfrac{1}{2})\big) \\
&- \rho\big(\Phi(\lceil\tilde{p}^*\rceil - \tfrac{1}{2}) - \hat{\Phi}(\lceil\tilde{p}^*\rceil - 1)\big)\lceil\tilde{p}^*\rceil \\
&- \sum_{a=0}^{\lceil\tilde{p}^*\rceil-1}\delta a\hat{\varphi}(a) + \int_{-\infty}^{\lceil\tilde{p}^*\rceil-1/2}\delta a\varphi(a)\,da
\end{aligned}$$

Notice that the first line is bounded by $[\rho-\delta,\rho]$ regardless of $\tilde{p}^*$, while second and third lines are on the order of $\tilde{p}^*$ times the approximation error between $\Phi$ and $\hat{\Phi}$. By the De Moivre–Laplace theorem, the approximation error is on the order of $M^{-1/2}$. Since by Eq. 10 and Eq. 11 both $\tilde{p}^*$ and the value of the continuous problem are on the order of $M$, the ratio of the welfare under the discrete contract and the value of the continuous problem converges to 1 as $M\to\infty$.

Finally, consider the true optimal contract in the discrete problem, and let $p_i$ be the number of promises and $p_i^*$ be the number of earnest promises of each agent $i$. The value of this contract can be approximated by evaluating the objective of the continuous problem at $p_i + \frac{1}{2}$ and $p_i^* + \frac{1}{2}$. By a similar argument, the deficit per agent of this approximation compared to the true value of the optimal discrete contract is no more than

$$\frac{1}{N}\sum_{i}\left(\gamma\big(\Phi(\lceil\tilde{p}^*\rceil-\tfrac{1}{2})-\hat{\Phi}(\lceil\tilde{p}^*\rceil-1)\big)\lceil\tilde{p}^*\rceil+\sum_{a=0}^{\lceil\tilde{p}^*\rceil-1}\delta a\hat{\varphi}(a)-\int_{-\infty}^{\lceil\tilde{p}^*\rceil-1/2}\delta a\varphi(a)\,da\right),$$

which is on the order of $M^{1/2}$. Therefore the ratio of $\hat{V}$ and the value of true optimal contract converges to 1 as $M\to\infty$. □

We now introduce some notation and lemmas, allowing the monitoring technology to be imperfect: an uncompleted task fails inspection with probability $\gamma\in(0,1]$. Let the number of monitoring slots used to monitor agent $i$ be $F_i$ , and the number of agent $i$'s promises be $p_i$. If agent $i$ fulfills $x$ of her promises, and tasks are drawn uniformly for monitoring, then the probability that agent $i$ will have $y$ failed inspections is given by the compound hypergeometric-binomial distribution

$$g(y,x)=\sum_{k=y}^{F_i}\frac{\binom{p_i-x}{k}\binom{x}{F_i-k}}{\binom{p_i}{F_i}}\binom{k}{y}\gamma^y(1-\gamma)^{k-y}. \tag{12}$$

To interpret Eq. 12, observe that in order to discover $y$ unfulfilled promises of agent $i$, the monitor(s) must have drawn $k\geq y$ promises from the $p_i-x$ promises agent $i$ failed to fulfill, and $F_i-k$ promises from the $x$ promises agent $i$ fulfilled; this is described by a hypergeometric distribution. Of these $k$ promises, the monitor(s) must then identify exactly $y$ failed inspections; this is described by a binomial distribution. This distribution is studied by Johnson and Kotz (1985) and shown by Stefanski (1992) to have a monotone likelihood ratio property: $g(y,x)/g(y,x-1)<g(y-1,x)/g(y-1,x-1)$ for all $x,y$. Hence an increase in the number of tasks completed yields a first-order stochastic improvement in the number of unfulfilled promises discovered.

**Lemma 2.** *Promise keeping is optimally implemented by a linear contract with $N-2$ agents making $M$ promises and two partial supervisors each monitoring one slot and making $M-1$ promises.*

*Proof.* Let $P_i$ be the set of promises made by agent $i$ and $p_i=|P_i|$. By incentive-compatibility, to ensure that $P_i$ rather than $a\subset P_i$ promises are fulfilled when $P_i$ are feasible, we need $h_{v_i}(A)\leq h_{v_i}(P_i)+(p_i-|A|)(\frac{b}{N}-c)$, where $h_{v_i}(\cdot)$ is the expected punishment conditional on the set of tasks completed. This means that $h_{v_i}(A)$ can be at best $(p_i-|A|)(\frac{b}{N}-c)$. We claim this can be achieved as in the statement of the lemma. Suppose each of two supervisors (agents $N-1$ and $N$) monitor $F$ tasks. We divide the entire set of agents into two, each assigned to a different supervisor's responsibility for monitoring (clearly each supervisor must be assigned to the group of the other supervisor). Each supervisor randomizes uniformly over which of their agents to monitor, and then uniformly over which task of that agent to monitor. Let $N_i$ be the number of agents in the group to which $i$ belongs. Penalties depend on the number $y$ of failed inspections. Let agent $i$'s penalty scheme be $v_i(y) = N_i y \frac{p_i}{\gamma F}(\frac{b}{N} - c)$, so that

$$h_{v_i}(A) = \frac{1}{N_i}\sum_{y=0}^{F} v_i(y)g(y,|A|) = \frac{p_i}{\gamma F}(\frac{b}{N} - c)\sum_{y=0}^{F} yg(y,|A|) = (p_i - |A|)(\frac{b}{N} - c)$$

because the expectation of the compound hypergeometric-binomial is $(p_i - |a|)\frac{\gamma F}{p_i}$. Expected punishments are independent of $F$, so long as $F \geq 1$. Moreover, this contract gives expected social utility

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{x=0}^{M}\binom{p_i}{x}\lambda^x(1-\lambda)^{p_i-x}[(b-c)x + (p_i - x)(\frac{b}{N} - c)] \\
&= (\frac{b}{N} - c)\sum_{i=1}^{N} p_i \sum_{x=0}^{M}\binom{p_i}{x}\lambda^x(1-\lambda)^{p_i-x} + \frac{N-1}{N}b\sum_{x=0}^{p_i}\binom{p_i}{x}\lambda^x(1-\lambda)^{p_i-x}x \\
&= (\frac{N-1}{N}b\lambda + \frac{b}{N} - c)\sum_{i=1}^{N} p_i.
\end{aligned}$$

This is positive if $\lambda > \frac{c-\frac{b}{N}}{\frac{N-1}{N}b}$ and largest when the maximal number of promises are made, using $F = 1$ for each of the two supervisors: $p_i = M$ for $i = 1, 2, \ldots, N-2$ and $p_{N-1} = p_N = M - 1$. □

**Lemma 3.** *Consider a penalty schedule $v_i$ where agent $i$ promises $M$ tasks, with $F < M$ uniformly monitored. If $v_i(y) = 0$ for $y < F$ and $v_i(F) < 0$ then a cutoff strategy is implemented, with agent $i$'s contribution to the social welfare given by*

$$\sum_{x=0}^{M}\binom{M}{x}\lambda^x(1-\lambda)^{M-x}\Big((b-c)\min\{x,p_i^*\} + \frac{(c-\frac{b}{N})g\big(F,\min\{x,p_i^*\}\big)}{g(F,p_i^*) - g(F,p_i^*-1)}\Big) \tag{13}$$

*when the cutoff $p_i^*$ is induced. The value of Eq. 13 is strictly increasing and concave in $\lambda$. Moreover, if $p_i^* < \tilde{p}_i^* \leq M - F + 1$, the value of Eq. 13 for $\tilde{p}_i^*$ strictly single crosses the value of Eq. 13 for $p_i^*$ from below, as a function of $\lambda$.*

*Proof of Lemma 3.* That a cutoff strategy is induced follows from Theorem 5. Eq. 13 follows from choosing $F$ to make the incentive constraint for doing $p_i^*$ versus $p_i^* - 1$ tasks bind. Let $\beta(x) \equiv (b-c)\min\{x,p_i^*\} + \frac{(c-\frac{b}{N})g(F,\min\{x,p_i^*\})}{g(F,p^*)-g(F,p^*-1)}$ The value of Eq. 13 is the expectation of $\beta(x)$ with respect to the binomial distribution over $x$. For any cutoff, the first term is concave. The second term of

$\beta(a)$ is a negative constant times $g(F, \min\{x, p_i^*\})$, where $g(F, \min\{x, p_i^*\}) = \lambda^F \binom{M-\min\{x,p_i^*\}}{F} \Big/ \binom{M}{F}$ is convex:

$$\begin{aligned}
&\binom{M - \min\{x+1, p_i^*\}}{F} - 2\binom{M - \min\{x, p_i^*\}}{F} + \binom{M - \min\{x-1, p_i^*\}}{F} \\
&= \begin{cases} \binom{M-x}{F}\left(\frac{F}{M-(x+1)-F} - \frac{F}{M-x}\right) & \text{if } x \leq p_i^* - 1, \\ \binom{M-(p_i^*-1)}{F} - \binom{M-p_i^*}{F} & \text{if } x = p_i^*, \\ 0 & \text{if } x \geq p_i^* + 1. \end{cases}
\end{aligned}$$

which is positive because $F \geq 1$, and $M - p_i^* + 1 \geq F$. Hence $\beta(x)$ is concave. Finally, the binomial distribution satisfies double-crossing, since

$$\frac{\partial^2}{\partial \lambda^2}\left(\binom{M}{x}\lambda^x(1-\lambda)^{M-x}\right) = \binom{M}{x}(1-\lambda)^{M-2-x}\lambda^{x-2}\left(x^2 - \left(1 + 2(M-1)\lambda\right)x + M(M-1)\lambda^2\right)$$

is negative if and only if $x^2 - \left(1 + 2(M-1)\lambda\right)x + M(M-1)\lambda^2 < 0$. Hence by Lemma 4, Eq. 13 is concave in $\lambda$. To see that Eq. 13 is increasing in $\lambda$, observe that the benefit of each task is linear in $x$, increasing in $p_i^*$ and independent of $\lambda$, which is a parameter of first-order stochastic dominance for the binomial distribution. The expected punishment for completing $\min\{x, p_i^*\}$ tasks is

$$\frac{(c - \frac{b}{N})g(F, \min\{x, p_i^*\})}{g(F, p_i^*) - g(F, p_i^* - 1)}.$$

Since $\lambda$ cancels out of the above, we need only check that this expression has increasing differences in $x$ and $p_i^*$ (by Corollary 10 of Van Zandt and Vives, 2007). Let us denote a $p_i^*$-cutoff strategy by $s_{p_i^*}$. Since $c - \frac{b}{N} > 0$, the sign of the second difference depends on

$$\begin{aligned}
&\frac{g(F, s_{p_i^*+1}(x+1)) - g(F, s_{p_i^*+1}(x))}{g(F, p_i^*+1) - g(F, p_i^*)} - \frac{g(F, s_{p_i^*}(x+1)) - g(F, s_{p_i^*}(x))}{g(F, p_i^*) - g(F, p_i^*-1)} \\
&= \begin{cases} 0 & \text{if } x \geq p_i^* + 1 \\ 1 & \text{if } x = p_i^* \\ \frac{g(F,x+1)-g(F,x)}{g(F,p_i^*+1)-g(F,p_i^*)} - \frac{g(F,x+1)-g(F,x)}{g(F,p_i^*)-g(F,p_i^*-1)} & \text{if } x \leq p_i^* - 1. \end{cases}
\end{aligned} \tag{14}$$

Concentrating on the third case, since $g(F, x)$ is decreasing in $x$, it suffices to show that

$$\binom{M - p_i^*}{F} - \binom{M - p_i^* + 1}{F} > \binom{M - p_i^* + 1}{F} - \binom{M - p_i^* + 2}{F}. \tag{15}$$

But this is exactly analogous to the earlier calculation. □

*Proof of Theorem 4.* We prove each part separately.

**$\lambda$ sufficiently low.** Fix $M$, $b$, and $c$. The value of a contract is continuous in $\lambda$, $P$, $s$, and $v$. Without loss of generality, $\mathcal{X}$ can be taken to be finite, with at least $NM$ tasks. Then both $P$ and $s$ are defined on compact spaces, and $v$ can without loss of generality take values from the extended non-positive real numbers $[-\infty, 0]$. Since the IR and IC constraints are weak inequalities that are continuous in $\lambda, b, c, P, v$, and $s$, the constraint set is compact-valued for each $\lambda$, $b$, and $c$. Finally, the constraint set is nonempty because it always contains the contract in which no tasks are promised and no punishments are imposed. Therefore, by Berge's Theorem of the Maximum (Aliprantis and Border, 2006, Theorem 17.31), the value of an optimal contract is continuous in $\lambda$ and the correspondence mapping $\lambda$ to the set of optimal contracts ($v$, $s$, and $\rho$) is upper hemicontinuous.

At $\lambda = 0$, in any optimal contract either $s(A) = \emptyset$ for all $A$ or $v(f) = 0$ for all $f$, so as $\lambda \to 0$, the optimal contracts must converge to either $s(A) = \emptyset$ for all $A$ or $v(f) = 0$ for all $f$. If punishments converge to zero, then it is incentive compatible only for the agents to choose $s(A) = \emptyset$ for all $A$, in which case it is optimal to set the punishments to exactly $v(f) = 0$ for all $f$. If the strategies converge to anything other than $s(A) = \emptyset$ for all $A$, then for incentive compatibility the punishments must diverge ($v(f) \to -\infty$ for some $f$)—but the value of such contracts does not converge to zero, a contradiction. Hence for $\lambda$ sufficiently low, $s(A) = \emptyset$ for all $A$ and $v(f) = 0$ for all $f$.

**$\lambda$ sufficiently high.** At $\lambda = 1$, in every optimal contract each of $N-2$ agents must promise $M$ tasks, and two partial supervisors each promise $M-1$ tasks, with all agents fulfilling all of them. The contract must impose severe enough punishments to make it incentive compatible for them to do so, but the punishments may be arbitrarily severe since they are not realized on the equilibrium path. The value of any such contract is $(NM-2)(b-c)$. For $\lambda \to 1$, the value of the contract must converge to $(NM-2)(b-c)$, and so must have the same number of promises per agent as above for $\lambda$ sufficiently high. To minimize the cost of punishments, all the downward constraints for completing that number of promises should bind, which is achieved by a linear contract with uniform randomization over monitored tasks. Finally, given a linear contract, $s(A) = A$ for all $A$ is optimal. Now apply Lemma 2.

**Intermediate $\lambda$.**

1. We show that empty promises with probabilistically complete monitoring are strictly better than promise-keeping at $\lambda^* = \frac{cN-b}{(N-1)b}$, and thus (by continuity) for an open neighborhood.

First, since $c < b < cN$, $\lambda^* \in (0,1)$. Now consider a probabilistically complete monitoring contract in which one agent supervises while $N-1$ agents each make $M$ promises and keep $M-1$ promises. To start, note that the value of promise keeping, $(MN-2)\left(\frac{N-1}{N}b\lambda + \frac{b}{N} - c\right)$, is zero at $\lambda^*$. The value of this probabilistically complete monitoring contract, in contrast, is

$$
\begin{aligned}
(N-1)\left((M-1)\left(\frac{b}{N}-c\right) + \frac{N-1}{N}b\sum_{a=0}^{M}\binom{M}{a}\lambda^a(1-\lambda)^{m-a}\min\{a, M-1\}\right) \\
= (N-1)\left((M-1)\left(\frac{b}{N}-c\right) + \frac{N-1}{N}b(M\lambda - \lambda^M)\right).
\end{aligned}
\tag{16}
$$

This expression is zero at the solution to

$$
\frac{M\lambda - \lambda^M}{M-1} = \lambda^*. \tag{17}
$$

By Descartes' Rule of Signs, the only real roots of $\frac{M\lambda-\lambda^M}{M-1} = \lambda$ are $\lambda = 0$ and $\lambda = 1$, and that $\frac{d}{d\lambda}\left(\frac{M\lambda-\lambda^M}{M-1}\right)\Big|_{\lambda=0} = \frac{M}{M-1} > 1$. These facts imply that the solution $\hat{\lambda}$ to Eq. 16 satisfies $\hat{\lambda} < \lambda^*$. Since the value of any contract is strictly increasing in $\lambda$, probabilistically complete monitoring with empty promises dominates promise keeping on a neighborhood of $\lambda^*$. By single crossing, this is also the case on $[0, \lambda^*]$.

2. We show that when $\frac{c}{b} > \frac{2}{e+2} + \frac{1}{N}\frac{e}{e+2}$, empty promises are strictly better than promise keeping for $\lambda = \frac{M-2}{M-1}$, and thus (by continuity) for an open neighborhood. Scale the promise-keeping contract with minimal monitoring by $2(N-2)$ to account for the probability of being monitored, we employ a maximally forgiving contract against each of the $N-2$ workers which enforces a cutoff $p^*$ as in Lemma 3. Combining Lemma 2 and Lemma 3, and applying $F = 2$, the social value of this contract when the cutoff is $p^*$ is given by

$$
\begin{aligned}
&2(M-1)\left(\frac{N-1}{N}b\lambda + \frac{b}{N} - c\right) \\
&\quad + (N-2)\sum_{a=0}^{M}\lambda^a(1-\lambda)^{M-a}\left(\min\{a,p^*\}(b-c) + \left(\frac{b}{N}-c\right)\frac{\binom{M-\min\{a,p^*\}}{2}}{M-p^*}\right)
\end{aligned}
\tag{18}
$$

By Lemma 3 this is a single-crossing family and the optimal $p^*$ increases with $\lambda$. Note that the binomial term is quadratic, and recall that $\sum_{a=0}^{M}\lambda^a(1-\lambda)^{M-a}a^2 = M\lambda(1-\lambda) + (M\lambda)^2$. When $p^* = M-1$, the binomial in the second term of Eq. 18 may be replaced with $\binom{M-a}{2}$

and Eq. 18 reduces to

$$2(M-1)\Big(\frac{N-1}{N}b\lambda+\frac{b}{N}-c\Big) \\ +(N-2)\bigg((M\lambda-\lambda^M)(b-c)+\Big(\frac{b}{N}-c\Big)\frac{M(M-1)(1-\lambda)^2}{2}\bigg). \quad (19)$$

Simplifying terms and factoring, Eq. 19 dominates promise-keeping, which has value $(MN-2)(\frac{N-1}{N}b\lambda+\frac{b}{N}-c)$, when

$$\begin{aligned} &2N(M\lambda-\lambda^M)(b-c)+(b-cN)\big(M(M-1)(1-\lambda)^2\big) \\ &\qquad > 2M\big((N-1)b\lambda+b-cN\big) \\ \iff &2N\lambda^M(c-b)-2M\lambda\big(cN+(N-(N-1))b\big) \\ &\qquad > \big(2M-M(M-1)(1-\lambda)^2\big)(b-cN) \\ \iff &2(b-c)\lambda^M N<(1-\lambda)(M-\lambda(M-1)-3)M(b-cN). \end{aligned} \quad (20)$$

Observe that the LHS of the last line in Eq. 20 is positive, so the RHS must be positive for Eq. 20 to be satisfied. Since $b<cN$, this requires $M-\lambda(M-1)-3<0$. Let $\lambda^*=\frac{M-2}{M-1}$. Then Eq. 20 reduces to

$$2(b-c)\Big(\frac{M-2}{M-1}\Big)^M N<\frac{M}{M-1}(cN-b). \quad (21)$$

The RHS of Eq. 21 is bounded below by $cN-b$. Consider $z_M=(\frac{M-2}{M-1})^M$. Taking logarithms, $\ln z_M=M\ln\frac{M-2}{M-1}$. Using l'Hôpital's rule, $\lim_{M\to\infty}\ln z_M=\lim_{M\to\infty}\frac{-M^2}{(M-2)(M-1)}=-1$, hence $z_M$ converge from below to $\frac{1}{e}$. The LHS of Eq. 21 is thus bounded above by $\frac{2(b-c)N}{e}$. Thus a sufficient condition for Eq. 21 is $\frac{2(b-c)N}{e}<cN-b$, which rearranges to $\frac{c}{b}>\frac{2}{e+2}+\frac{1}{N}\frac{e}{e+2}$.

3. We show that for any $M$, $N$, the optimal equilibrium (including asymmetries, nonuniform monitoring, etc) involves empty promises for an open range of the parameters $\lambda$ and $\frac{c}{b}$ by showing a range where having all but one promise empty dominates promise keeping. By Lemma 2, the social value of promise keeping is given by $(MN-2)(\frac{N-1}{N}b\lambda+\frac{b}{N}-c)$. If each player would do just one promise out of $p=\frac{M+1}{2}$ promises, with $F=\frac{M-1}{2}$, the payoff using a maximally forgiving contract would be

$$N\cdot\Big((1-(1-\lambda)^{\frac{M+1}{2}})(\frac{2(\frac{b}{N}-c)}{M-1}+b-c)+(1-\lambda)^{\frac{M+1}{2}}(\frac{b}{N}-c)\frac{M+1}{M-1}\Big).$$

This dominates promise-keeping when (after some algebra)

$$\lambda(M-\frac{2}{N})+(1-\lambda)^{\frac{M+1}{2}} < \frac{c}{b}(\frac{N}{N-1}(M-\frac{2}{M-1}-\frac{2}{N}-1)-\frac{1}{N-1}(M-\frac{2}{M-1}-\frac{2}{N}-1)+1.$$

Let $\varphi(\lambda) := \lambda(M-\frac{2}{N})+(1-\lambda)^{\frac{M+1}{2}}$. Note that $\varphi'(\lambda) \geq 0$ for $M \geq 3$ and $N \geq 2$, so we want to check the above condition for $\tilde{\lambda} = \frac{c-\frac{b}{N}}{b-\frac{b}{N}} = \frac{N}{N-1}\frac{c}{b} - \frac{1}{N-1}$, which is the point at which promise keeping begins to have positive social value. Let $x = \frac{c}{b}$. Note that $x \in (\frac{1}{N}, 1)$. Then the condition reduces to

$$(xN-1)(\frac{2}{N}+1)+(N-1)(\frac{N}{N-1}-\frac{N}{N-1}x)^{\frac{M+1}{2}} < N-1.$$

This is always satisfied at $x = \frac{1}{N}$, for each $M$. Indeed, the above works for any $x < \frac{N^2+2}{N^2+2N}$ when $M$ goes to infinity. Similarly, as $N$ goes to infinity this goes to $x + (1-x)^{\frac{M+1}{2}} < 1$, which always works for $x < 1$.

$\square$

# References


Scott Adams. *The Dilbert Principle: A Cubicle's-Eye View of Bosses, Meetings, Management Fads & Other Workplace Afflictions*. HarperCollins Publishers, Inc, New York, NY, 1996.

Nabil I. Al-Najjar, Luca Anderlini, and Leonardo Felli. Undescribable events. *Review of Economic Studies*, 73(4):849–868, 2006.

Armen A. Alchian and Harold Demsetz. Production, informatin costs, and economic organization. *American Economic Review*, 62(5):777–795, December 1972.

Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis*. Springer, New York, NY, 2006.

Luca Anderlini and Leonardo Felli. Describability and agency problems. *European Economic Review*, 42(1):35–59, January 1998.

Masahiko Aoki. *Information, incentives, and bargaining in the Japanese economy*. Cambridge University Press, Cambridge, UK, 1988.

Masahiko Aoki. The continent governance of teams: Analysis of institutional complementarity. *International Economic Review*, 35(3):657–676, August 1994.

Debra J. Aron and Pau Olivella. Bonus and penalty schemes as equilibrium incentive devices, with application to manufacturing systems. *Journal of Law, Economics, and Organization*, 10(1): 1–34, April 1994.

George Baker, Robert Gibbons, and Kevin J. Murphy. Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics*, 109(4):1125–1156, November 1994.

George P. Baker, Michael C. Jensen, and Kevin J. Murphy. Compensation and incentives: Practice vs. theory. *Journal of Finance*, 43(3):593–616, July 1987.

John M. Barron and Kathy Paulson Gjerde. Peer pressure in an agency relationship. *Journal of Labor Economics*, 15(2):234–254, April 1997.

Marco Battaglini. Joint production in teams. *Journal of Economic Theory*, 130:138–167, 2006.

Pierpaolo Battigalli and Giovanni Maggi. Rigidity, discretion, and the costs of writing contracts. *American Economic Review*, 92(4):798–817, September 2002.

Roland Benabou. Groupthink and ideology. *Working Paper*, 2008.

Roland Benabou and Jean Tirole. Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117:871–915, 2002.

Aaron L. Bodoh-Creed. Mood, associative memory, and the formation and dynamics of belief. Working Paper, April 2010. URL http://www.stanford.edu/%7Eacreed/Associative%20Memory.pdf.

Patrick Bolton and Antoine Faure-Grimaud. Satisficing contracts. *Review of Economic Studies*, Forthcoming. URL http://dx.doi.org/10.1111/j.1467-937X.2009.00597.x.

Brent Boning, Casey Ichniowski, and Kathryn L. Shaw. Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics*, 25(4):613–650, October 2007.

Kim C. Border and Joel Sobel. Samurai accountant: A theory of auditing and plunder. *The Review of Economic Studies*, 54(4):525–540, October 1987.

Jeffrey Carpenter, Samuel Bowles, Herbert Gintis, and Sung-Ha Hwang. Strong reciprocity and team production: Theory and evidence. *Journal of Economic Behavior & Organization*, 71: 221–232, 2009.

Archishman Chakraborty and Rick Harbaugh. Comparative cheap talk. *Journal of Economic Theory*, 132:70–94, 2007.

Yeon-Koo Che and Seung-Weon Yoo. Optimal incentives for teams. *American Economic Review*, 91(3):525–541, June 2001.

Harold L. Cole and Narayana R. Kocherlakota. Finite memory and imperfect monitoring. *Games and Economic Behavior*, 53:59–72, 2005.

Olivier Compte and Andrew Postlewaite. Repeated relationships with limits on information processing. *Working Paper*, 2008.

Nelson Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24:87–185, 2000.

Claude d'Aspremont and Louis-André Gérard-Varet. Linear inequality methods to enforce partnerships under uncertainty: An overview. *Games and Economic Behavior*, 25:311–336, 1998.

James Dow. Search decisions with limited memory. *Review of Economic Studies*, 58:1–14, 1991.

Peter C. Fishburn. Moment-preserving shifts and stochastic dominance. *Mathematics of Operations Research*, 7:629–634, 1982.

Alexander Frankel. Aligned delegation. Working Paper, November 2010. URL http://www.stanford.edu/~afrankel/jmp/frankel.pdf.

James R. Frederickson and William Waller. Carrot or stick? Contract frame and use of decision-influencing information in a principal-agent setting. *Journal of Accounting Research*, 43(5): 709–733, December 2005.

David Hirshleifer and Ivo Welch. An economic approach to the psychology of change: Amnesia, inertia, and impulsiveness. *Working Paper*, 2001.

Bengt Holmström. Moral hazard in teams. *Bell Journal of Economics*, 13(2):324–340, 1982.

Bengt Holmström and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics & Organization*, 7(Sp):24–52, 1991.

Matthew O. Jackson and Hugo F. Sonnenschein. Overcoming incentive constraints by linking decisions. Forthcoming, May 2005.

Norman L. Johnson and Samuel Kotz. Some distributions arising as a consequence of errors in inspection. *Naval Research Logistics Quarterly*, 32:35–43, 1985.

Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, New Jersey, third edition, 2005.

Eugene Kandel and Edward P. Lazear. Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–817, August 1992.

Ola Kvaløy and Trond E. Olsen. Team incentives in relational employment contracts. *Journal of Labor Economics*, 24(1):139–169, 2006.

Christian Laux. Limited-liability and incentive contracting with multiple projects. *RAND Journal of Economics*, 32(3):514–526, Autumn 2001.

Edward P. Lazear and Kathryn L. Shaw. Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, 21(4):91–114, Fall 2007.

Patrick Legros and Hitoshi Matsushima. Efficiency in partnerships. *Journal of Economic Theory*, 55:296–322, 1991.

Patrick Legros and Steven A. Matthews. Efficient and nearly-efficient partnerships. *Review of Economic Studies*, 60(3):599–611, July 1993.

Jonathan Levin. Relational incentive contracts. *American Economic Review*, 93(3):835–857, June 2003.

Shuhe Li and Weiying Zhang. Optimal assignment of principalship in teams. *Journal of Economic Behavior & Organization*, 44:105–127, 2001.

G. G. Lorentz. An inequality for rearrangements. *American Mathematical Monthly*, 60(3):176–179, March 1953.

W. Bentley MacLeod. Optimal contracting with subjective evaluation. *American Economic Review*, 92(1):216–240, March 2003.

Eric Maskin and Jean Tirole. Unforeseen contingencies and incomplete contracts. *Review of Economic Studies*, 66(1):83–114, January 1999.

Hitoshi Matsushima, Koichi Miyazaki, and Nobuyuki Yagi. Role of linking mechanisms in multitask agency with hidden information. *Journal of Economic Theory*, 145:2241–2259, 2010.

R. Preston McAfee and John McMillan. Optimal contracts for teams. *International Economic Review*, 32(3):561–577, August 1991.

George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63:81–97, 1956.

Nolan H. Miller. Efficiency in partnerships with joint monitoring. *Journal of Economic Theory*, 77:285–299, 1997.

James A. Mirrlees. The optimal structure of incentives and authority within an organization. *Bell Journal of Economics*, 7(1):105–131, Spring 1976.

Dilip Mookherhjee and Ivan Png. Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics*, 104(2):399–415, May 1989.

Sendhil Mullainathan. A memory-based model of bounded rationality. *Quarterly Journal of Economics*, 117:735–774, 2002.

Paul Oyer and Scott Schaefer. Why do some firms give stock options to all employees?: An empirical examination of alternative theories. *Journal of Financial Economics*, 76:99–133, 2005.

Thomas R. Palfrey and Howard Rosenthal. Participation and the provision of discrete public goods: A strategic analysis. *Journal of Public Economics*, 24(2):171–193, 1984.

Michele Piccione and Ariel Rubinstein. Finite automata play a repeated extensive game. *Journal of Economic Theory*, pages 160–168, 1993.

Michele Piccione and Ariel Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, pages 3–24, 1997.

Roy Radner. Repeated principal-agent games with discounting. *Econometrica*, 53(5):1173–1198, September 1985.

David Rahman. But who will monitor the monitor? Working paper, September 2010. URL http://www.econ.umn.edu/~dmr/monitor.pdf.

David Rahman and Ichiro Obara. Mediated partnerships. *Econometrica*, 78(1):285–308, January 2010.

Julian Romero. Bounded rationality in repeated games. Working paper, January 2010. URL http://jnromero.com/research/papers/Bounded%20Rationality%20in%20Repeated%20Games.pdf.

Ilya Segal. Complexity and renegotiation: A foundation for incomplete contracts. *Review of Economic Studies*, 66(1):57–82, January 1999.

Christopher M. Snyder. Bounding the benefits of stochastic auditing: The case of risk-neutral agents. *Economic Theory*, 14(1):247–253, July 1999.

Leonard Stefanski. Monotone likelihood ratio of a 'faulty-inspection' distribution. *The American Statistician*, 46:110–114, 1992.

Jean Tirole. Cognition and incomplete contracts. *American Economic Review*, 99(1):265–294, March 2009.

Timothy Van Zandt and Xavier Vives. Monotone equilibria in Bayesian games of strategic complementarities. *Journal of Economic Theory*, 134(1):339–360, May 2007.

Stephen D. Williamson. Costly monitoring, loan contracts, and equilibrium credit rationing. *The Quarterly Journal of Economics*, 102(1):135–145, 1987.

Andrea Wilson. Bounded memory and biases in information processing. Working Paper, October 2004. URL http://qed.econ.queensu.ca/paper/memorybiases1.pdf.

# **Supplemental appendix** (not for publication)

In this supplemental appendix we consider contracts that are symmetric with respect to task names and for which the amount of monitoring to be accomplished (denoted $F$) is public. In this case, a penalty schedule provides punishment based on the number of failures $f$ of inspection, where $f \in \{0, 1, \ldots, F\}$. Within this class, contracts which deliver increasingly large punishments for larger numbers of inspection failures may be a focal class to consider. Such *decreasing convex (DC)* contracts satisfy the restriction $v(f) - v(f+1) \geq v(f-1) - v(f) \geq 0$. Convex contracts may be natural in settings where punishments are imposed by third parties who are more inclined to exact punishment if they perceive a consistent pattern of failures. Conversely, a non-convex contract may be particularly difficult to enforce via an affected third party, since it would require leniency on the margin for relatively large injuries. For arbitrary memory size $M$, we show that DC contracts optimally induce task-completion strategies that have a cutoff form. Furthermore, the optimal such contract forgives empty promises up to some failure threshold, and increases punishments linearly thereafter.

**Theorem 5.** *For any $M$, cutoff strategies and kinked linear penalties are optimal in the class of DC contracts.*

We first prove several lemmas. The first provides a sufficient condition on a one-parameter family of probability distributions for the expectation of a concave function to be concave in the parameter. Though it can be derived from a theorem of Fishburn (1982), we provide a simpler statement of the condition along with a more direct proof. We say that a function $\psi : \{0, 1, \ldots, R\} \to \mathbb{R}$ is *concave* if $\psi(r+1) - \psi(r) \leq \psi(r) - \psi(r-1)$ for all $r = 1, \ldots, R-1$. A function $\varphi : \mathcal{Z} \to \mathbb{R}$, where $\mathcal{Z} \subseteq \mathbb{R}$, is *double crossing* if there is a (possibly empty) convex set $A \subset \mathbb{R}$ such that $A \cap \mathcal{Z} = \left\{z \in \mathcal{Z} : \varphi(z) < 0\right\}$.

**Lemma 4.** *Let $\mathcal{R} = \{0, 1, \ldots, R\}$, and let $\{q_z\}_{z \in \mathcal{Z}}$ be a collection of probability distributions on $\mathcal{R}$ parameterized by $z \in \mathcal{Z} = \{0, 1, \ldots, Z\}$.*[22] *The function $\Psi(z) = \sum_{r=0}^{R} \psi(r) q_z(r)$ is concave if*

1. *There exists $k, c \in \mathbb{R}$, $k \neq 0$, such that $z = k \sum_{r=0}^{R} r q_z(r) + c$ for all $z \in \mathcal{Z}$;*
2. *$q_{z+1}(r) - 2q_z(r) + q_{z-1}(r)$ for all $z = 1, \ldots, Z-1$, as a function of $r$, is double crossing;*
3. *$\psi : \{0, 1, \ldots, R\} \to \mathbb{R}$ is concave.*

*Proof.* Since $z = k \sum_{r=0}^{R} r q_z(r) + c$, there exists $\hat{b} \in \mathbb{R}$ such that $\sum_{r=0}^{R} (mr + b) q_z(r) = \frac{m}{k} z + \hat{b} + c$

[22] A similar result holds if $z \in \mathcal{Z} = [0, 1]$.

for any real $m$ and $b$. Hence, for any $m$ and $b$,

$$\sum_{r=0}^{R}(mr+b)\big(q_{z+1}(r)-2q_z(r)+q_{z-1}(r)\big) = \frac{m}{k}\big(z+1-2z+z-1\big) = 0,$$

for all $z = 1, \dots, Z-1$. Therefore, for any $m$ and $b$, the second difference of $\Psi(z)$ can be written as

$$\begin{aligned}\Psi(z+1)-2\Psi(z)+\Psi(z-1) &= \sum_{r=0}^{R}\psi(r)\big(q_{z+1}(r)-2q_z(r)+q_{z-1}(r)\big) \\ &= \sum_{r=0}^{R}\big(\psi(r)-mr-b\big)\big(q_{z+1}(r)-2q_z(r)+q_{z-1}(r)\big).\end{aligned} \tag{22}$$

By assumption, $q_{z+1}(r)-2q_z(r)+q_{z-1}(r)$, as a function of $r$, is double crossing. Furthermore, since $\psi$ is concave, we can choose $m$ and $b$ such that, wherever $\big(q_{z+1}(r)-2q_z(r)+q_{z-1}(r)\big)$ or $\frac{\partial^2}{\partial z^2}q_z(r)$ is nonzero, $\psi(r)-mr-b$ either has the opposite sign or is zero. From Eq. 22 we may conclude $\Psi(z)$ is concave. □

The next lemma says that expected punishments will be decreasing convex in the number of tasks completed. Recall the definition of $g(f,a)$ from Eq. 12.

**Lemma 5.** *If $v$ is decreasing convex, then $h_v \equiv \sum_{f=0}^{F} v(f)g(f,\cdot)$ is decreasing convex.*

*Proof.* By letting $a \equiv |A|$, reversing the order of summation, and using fact that $\binom{k}{f} = 0$ when $k < f$, we can write $h_v(A)$ as follows:

$$\begin{aligned}h_v(A) = \sum_{f=0}^{F} g(f,a)v(f) = &\sum_{f=0}^{F}\left(\sum_{k=0}^{F}\frac{\binom{p-a}{k}\binom{a}{F-k}}{\binom{p}{F}}\binom{k}{f}\gamma^f(1-\gamma)^{k-f}\right)v(f) \\ = &\sum_{k=0}^{F}\frac{\binom{p-a}{k}\binom{a}{F-k}}{\binom{p}{F}}\left(\sum_{f=0}^{F}\binom{k}{f}\gamma^f(1-\gamma)^{k-f}v(f)\right).\end{aligned}$$

Therefore, the expectation is first with respect to the binomial, and then with respect to the hypergeometric. Using Lemma 4 twice gives the result. First, note that the expectation of the binomial is $\gamma k$, a linear function of $k$, while the expectation of the hypergeometric is $\frac{F}{p}(p-a)$, a linear function of $a$. Hence it suffices to show that the binomial second-difference in $k$ is double-crossing in $f$ (hence the inside expectation is decreasing convex in $k$) and the hypergeometric second-difference in $a$ is double-crossing in $k$. To see this is true for the binomial, note that we may write the binomial second-difference in $k$ as

$$\binom{k}{f}\gamma^f(1-\gamma)^{k-f}\left(\frac{(k+1)(1-\gamma)}{k+1-f}-2+\frac{k-f}{k(1-\gamma)}\right).$$

It can be shown that the term in parentheses is strictly convex in $f$ and therefore double crossing in $f$, so the whole expression is double-crossing in $f$. To see this is true for the hypergeometric, note that we may write the hypergeometric second-difference in $a$ as

$$\frac{\binom{p-a}{k}\binom{a}{F-k}}{\binom{p}{F}}\left(\frac{p-a-k}{p-a}\cdot\frac{a+1}{a+1-F+k}-2+\frac{p-a+1}{p-a+1-k}\cdot\frac{a-F+k}{a}\right).$$

It can be shown that the term in parentheses has either no real roots or exactly two real roots.[23] If there are no real roots, then the term in parentheses is double-crossing in $k$ (recall that the region in which it is negative must be convex, but may be empty), and therefore the whole expression is double-crossing in $k$. If there are two real roots, it can be shown that the derivative with respect to $k$ is negative at the smaller root, and that therefore both the term in parentheses and the whole expression are double-crossing in $k$. □

*Proof of Theorem 5.* Fix any $p, F, \lambda$. Suppose that the strategy $s$, with $p^* > 0$ the maximal number of tasks completed, is optimal. Consider the decreasing convex contract $v$ that implements $s$ at minimum cost. Because $v$ is decreasing, MLRP (more weakly, FOSD in $a$) implies expected punishments are decreasing in the number of tasks completed: $h(a) > h(a-1)$ for all $a$. By contradiction, suppose that the downward constraint for $p^*$ versus $p^*-1$ is slack: $h(p^*)-h(p^*-1) > c-b$. By Lemma 5 and monotonicity, for any $k > 1$, $h(p^*-k+1)-h(p^*-k) > c-b$. But then for every $a$ such that $s(a) = a$, and every $a' < a$, the downward constraint $h(a)-h(a') = \sum_{k=a'}^{a-1} h(k+1)-h(k) \geq (a-a')(c-b)$ must be slack. However, some constraint must bind at the optimum, else the strategy is implementable for free, so it must be that the downward constraint for $p^*$ versus $p^*-1$ binds. Again, each downward constraint is satisfied, and for any $a > p^*$, $h(a)-h(p^*) < (a-p^*)(c-b)$. So the strategy $s$ is a $p^*$-cutoff.

Now, suppose that we look for the optimal convex contract implementing $p$ promises, $F$ monitoring slots, and cutoff strategy $s$ with cutoff $p^*$. By the argument above, the only incentive constraint that binds is the downward constraint for completing $p^*$ promises. Since $v(0) = 0$, convexity implies monotonicity. Furthermore the constraint $v(0) \geq 0$ does not bind,[24] so the cost minimization

[23] The term in parentheses does not account for the fact that the entire expression equals zero whenever $k > p-a$ or $F-k > a$. However, on the closure of these regions the second difference cannot be negative, and so these regions may be ignored.

[24] Although $v(0) \geq 0$ is satisfied with equality, the binding constraint on $v(0)$ is actually $v(0) \leq 0$.

problem (dropping all other incentive constraints) in primal form is

$$\max_{(-v)\geq\vec{0}} \sum_{f=0}^{F}\Big(-(-v(f))\sum_{a=0}^{p} -g(f,a)t_s(a)\Big) \text{ subject to}$$
$$\sum_{f=0}^{F}(-v(f))[g(f,p^*)-g(f,p^*-1)] \leq -(c-b),$$
$$2(-v(f))-(-v(f+1))-(-v(f-1)) \leq 0 \text{ for all } f=1,\ldots,F-1,$$

where $t_s(a)=\sum_{a'=a}^{p}\mathbb{I}\big(s(a')=a\big)\binom{p}{a'}\lambda^{a'}(1-\lambda)^{p-a'}$ denotes the probability of completing $a$ tasks given task-completion strategy $s$.

Let $x$ be the Lagrange multiplier for the lone incentive compatibility constraint, $z_f$ the Lagrange multiplier for the convexity constraint $2(-v(f))-(-v(f+1))-(-v(f-1)) \leq 0$, and $\vec{z}$ the vector $(z_1,\ldots,z_{F-1})$. The constraint set can be written as $A^\top \cdot (-v(0),\ldots,-v(F))$, where, in sparse form,

$$A=\begin{pmatrix} g(0,p^*)-g(0,p^*-1) & -1 & & & & \\ \vdots & 2 & \ddots & & & \\ \vdots & -1 & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & \ddots & \\ \vdots & & & \ddots & \ddots & -1 \\ \vdots & & & & \ddots & 2 \\ g(F,p^*)-g(F,p^*-1) & & & & & -1 \end{pmatrix}.$$

Let $r$ be the vector of dual variables: $r=(x,z_1,\ldots,z_{F-1})$. The dual problem is

$$\min_{r\geq\vec{0}} (b-c)x \quad \text{s.t. } (Ar)_f \geq -\sum_{a=0}^{p} g(f,a)t_s(a) \text{ for all } f=0,1,\ldots,F,$$

where $(Ar)_f$ is the $(f)$th component of $A\cdot r$; i.e.,

$$(Ar)_f = x[g(f,p^*)-g(f,p^*-1)] - z_{f-1} + 2z_f - z_{f+1},$$

where we define $z_0 \equiv 0$, $z_F \equiv 0$, and $z_{F+1} \equiv 0$.

Let $\hat{f}$ be the smallest $f$ such that $v(f) < 0$. Then it must be that $v(f) < 0$ for all $f \geq \hat{f}$, so by duality theory the constraint $(A \cdot r)_f \geq -\sum_{a=0}^{p} g(f,a)t_s(a)$ binds for all $f \geq \hat{f}$. Hence

$$x = \frac{\sum_{a=0}^{p} g(f,a)t_s(a) - z_{f-1} + 2z_f - z_{f+1}}{g(f,p^*-1) - g(f,p^*)} \text{ for all } f = \hat{f},\ldots,F. \tag{23}$$

In particular, this means that if $z_{F-1} = 0$ (which is implied when $\hat{f} = F$) the optimal contract (which would have expected punishment $-x(c-b)$) has the same value as that derived in **??**), completing the claim. In the remainder we assume $z_{F-1} > 0$.

Observe that the sum of the $z$-terms over $(A \cdot r)_{F-1}$ and $(A \cdot r)_F$ is $-z_{F-1} + (2z_{F-1} - z_{F-2}) = z_{F-1} - z_{F-2}$. Note also the corresponding sum of $z$-terms over $F-2$, $F-1$, and $F$: $-z_{F-1} + (2z_{F-1} - z_{F-2}) + (-z_{F-3} + 2z_{F-2} - z_{F-1}) = z_{F-2} - z_{F-3}$. Continuing in this manner, the sum of the $z$-terms in $(A \cdot r)_f$ from any $\tilde{f} \geq \hat{f}$ to $F$ is $z_{\tilde{f}} - z_{\tilde{f}-1}$. Therefore, summing the equalities in Eq. 23 yields the following recursive system for $z_{\tilde{f}}, \tilde{f} = \hat{f},\ldots,F$

$$z_{\tilde{f}} = z_{\tilde{f}-1} - \sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a) + x \sum_{f=\tilde{f}}^{F} \big(g(f,p^*-1) - g(f,p^*)\big).$$

Also, by definition of $\hat{f}$ the convexity constraint is slack at $\hat{f}-1$, so $z_{\hat{f}-1} = 0$. Then induction yields, for $f' = \hat{f},\ldots,F$,

$$z_{f'} = -\sum_{\tilde{f}=\hat{f}}^{f'} \sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a) + x \sum_{\tilde{f}=\hat{f}}^{f'} \sum_{f=\tilde{f}}^{F} \big(g(f,p^*-1) - g(f,p^*)\big).$$

Plugging this equation for $f' = F$ into the binding constraint $(Ar)_F \geq -\sum_{a=0}^{p} g(F,a)t_s(a)$ provides solution for $x$ in terms of $\hat{f}$:

$$x = \frac{\sum_{\tilde{f}=\hat{f}}^{F} \sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a)}{\sum_{\tilde{f}=\hat{f}}^{F} \sum_{f=\tilde{f}}^{F} \big(g(f,p^*-1) - g(f,p^*)\big)}. \tag{24}$$

Note that for a random variable $X$ on $\{0,\ldots,n\}$, the expectation of $X$ is $\sum_{j=1}^{n} j \Pr(X=j)$ but this is also equal to $\sum_{j=1}^{n} \Pr(X \geq j)$. Since $\sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a) = \Pr(f \geq \tilde{f})$, the numerator of

Eq. 24 can be rewritten as

$$\begin{aligned}\sum_{\tilde{f}=\hat{f}}^{F}\sum_{f=\tilde{f}}^{F}\sum_{a=0}^{p} g(f,a)t_s(a) &= \sum_{\tilde{f}=\hat{f}}^{F} \Pr(f \geq \tilde{f}) = \sum_{\tilde{f}=\hat{f}}^{F} (\tilde{f}-\hat{f}+1)\Pr(f=\tilde{f}) \\ &= \sum_{\tilde{f}=1}^{F} (\tilde{f}-\hat{f}+1)_{+}\Pr(f=\tilde{f}) = \mathbb{E}\big[(f-\hat{f}+1)_{+}\big] \equiv \mathbb{E}[\varphi(\hat{f})],\end{aligned}$$

where $(y)_+ \equiv \max\{y,0\}$ and $\varphi$ is the random function $\varphi(\hat{f}) \equiv (f-\hat{f}+1)_+$. In words, $\varphi(\hat{f})$ is the number of discovered unfulfilled promises that exceed the threshold for punishment $(\hat{f})$. The denominator of Eq. 24 can be rewritten similarly, yielding

$$x = \frac{\mathbb{E}[\varphi(\hat{f})]}{\mathbb{E}[\,\varphi(\hat{f}) \mid a = p^* - 1\,] - \mathbb{E}[\,\varphi(\hat{f}) \mid a = p^*\,]}. \tag{25}$$

The minimized expected punishment is $\mathbb{E}[v(f)] = (b-c)x$, and hence is implemented by the kinked-linear punishment schedule

$$v(f) = -\frac{(c-b)(f-\hat{f}+1)_+}{\mathbb{E}[\,\varphi(\hat{f}) \mid a = p^* - 1\,] - \mathbb{E}[\,\varphi(\hat{f}) \mid a = p^*\,]} \text{ for all } f = 0,1,\ldots,F.$$

$\square$