OPTIMALLY EMPTY PROMISES AND ENDOGENOUS SUPERVISION

By

David A. Miller and Kareen Rozen

October 2011
Revised June 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1823

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Optimally empty promises and endogenous supervision*

David A. Miller† — UCSD

Kareen Rozen‡ — Yale

June 2012

Abstract

We study optimal contracting in team settings, featuring stylized aspects of production environments with complex tasks. Agents have many opportunities to shirk, task-level monitoring is needed to provide useful incentives, and because it is difficult to write individual performance into formal contracts, incentives are provided informally, using wasteful sanctions like guilt and shame, or slowed promotion. These features give rise to optimal contracts with "empty promises" and endogenous supervision structures. Agents optimally make more promises than they intend to keep, leading to the concentration of supervisory responsibility in the hands of one or two agents.

Keywords: Partnership, teams, moral hazard, monitoring, supervision, informal sanctions.
JEL Codes: C72, D03, D86.

---

*We are grateful to Andy Postlewaite, Larry Samuelson, Todd Sarver, Joel Sobel, Juuso Välimäki, and Joel Watson for helpful comments and suggestions. We also thank seminar participants at Northwestern and Yale. Max Dvorkin 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, and the NSF for financial support (award SES-1127643).

‡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 (award SES-0919955).

Contents

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 sanctions 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 sanction him. Empty promises arise, for example, when Bob is assigned four tasks but completes at most two of them, regardless of how many are feasible. For Bob to make those two empty promises, Carol should not sanction 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 two empty promises is small; and forgiving them saves Bob from sanctions in the more likely event that two or fewer of his tasks are feasible. Empty promises thus buffer against the sanctions Bob would incur when too few tasks are feasible. In the presence of costly sanctions, it may be socially optimal for Carol to simply accept Bob's empty promises rather than force him to complete all his feasible tasks. Bob's promises are "empty" because it is commonly known that he will not complete them, even if they are feasible.

In this worker-supervisor arrangement, it is socially optimal for Bob to shirk some of his feasible tasks even though Carol is monitoring all of them. 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 promise two tasks and monitor the other's two. Or Bob could promise one task and Carol could promise three, with each monitoring all the promises of the other. Are all these arrangements

[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., Radner, 1985; Baker, Gibbons, and Murphy, 1994; Che and Yoo, 2001; Levin, 2003; MacLeod, 2003) 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.

welfare-equivalent? No—whenever it is optimal for anyone to make empty promises, the worker-supervisor arrangement is strictly better than any mutual monitoring arrangement. So in the optimal contract an endogenous supervisor emerges. This occurs due to the statistical complementarities arising when cutoff strategies are used. If promise-keeping were optimal, it would make no difference how supervisory responsibility is allocated.

Our task-based approach with peer monitoring and informal, wasteful sanctions 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 prototypical 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…

Another is the shopfloor of an American garment producer, where workers in teams "learned all production tasks, had more information about production tasks, and … quickly caught quality problems, which allowed the team to quickly identify and correct the source of quality problems" (Hamilton, Nickerson, and Owan, 2003, p. 477).

Finally, since the output of a specialist is hard to describe, it is hard to write performance-based incentives at the individual level into a formal contract. Lacking the ability to enforce transfers through a formal contract, the agents must provide individual-level incentives informally, using wasteful instruments like guilt and shame (Kandel and Lazear, 1992; Barron and Gjerde, 1997; Carpenter, Bowles, Gintis, and Hwang, 2009).[2] When extreme sanctions 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"; moreover, "midcareer separation may signal negative attributes."

In our example, if 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, 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 theory 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 sanctions. 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,[3] or allows for exogenously specified individual-level monitoring.[4] In contrast, our approach endogenizes individual-level monitoring by putting the agents in charge of monitoring each other. We assume that assigning

[2]Linear incentives based on team-level output merely amplify the social benefits of task completion that are already a primitive of our model. In Section 6 we study a firm that hires a team of agents, offering them a formal contract that is linear in team output in concert with an informal contract of wasteful sanctions. Because the formal contract cannot distinguish among the tasks completed by different agents, the firm optimally offers a team-level output bonus that is too small to solve the moral hazard problem.

[3]For example: Legros and Matsushima (1991); Legros and Matthews (1993); d'Aspremont and Gérard-Varet (1998); Battaglini (2006); Coviello, Ichino, and Persico (2011). Free riding, of course, also arises in public goods problems (e.g., Palfrey and Rosenthal, 1984).

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

an agent to monitor her peers crowds out her own productivity.[5] This allows us to study the trade-off 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,[6] in our model each task constitutes an independent dimension of effort.[7] This assumption imposes a natural structure on the stochastic relationship among effort, output, and monitoring, and enables us to make more specific predictions about task completion and supervision than 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).[8] Instead, we rule out formal monetary transfers, and focus on providing incentives through informal sanctions that are socially wasteful. Wasteful sanctions are also studied by Kandel and Lazear (1992); Barron and Gjerde (1997); Che and Yoo (2001); Carpenter, Bowles, Gintis, and Hwang (2009) in different settings. Theoretically, a framework with sanctions may be interpreted as the reduced-form of a repeated game with a sharply kinked Pareto frontier. Such sanctions are a natural instrument in an environment in which the agents cannot commit to inspection-contingent transfers. In principle, formal monetary transfers are one polar case, while wasteful sanctions are another. Since many realistic scenarios share some features of both polar cases, a natural next step is to study the implications of wasteful sanctions. In practice, if a team of three or more agents *could* commit to inspection-contingent transfers, then promise-keeping would be optimal and

[5]Li and Zhang (2001) formalize Alchian and Demsetz (1972)'s conjecture that costly monitoring should be the responsibility of a residual claimant. Rahman (2010) and Rahman and Obara (2010) show the monitor need not be the residual claimant when a mediator can make correlated recommendations. Like Li and Zhang, we show monitoring responsibilities are optimally given to one agent, but like Rahman and Rahman and Obara we need not give the monitoring agent residual claims. There is also a literature on costly monitoring in principal-agent relationships (e.g., Townsend, 1979; Border and Sobel, 1987; Williamson, 1987; Mookherhjee and Png, 1989; Snyder, 1999).

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

[7]Matsushima, Miyazaki, and Yagi (2010) also study a model where agents have private information about the feasibility of arbitrarily many independent tasks, but assume that monitoring is exogenous and utility is transferable. Holmström and Milgrom (1991); Legros and Matsushima (1991); Legros and Matthews (1993); Miller (1997); d'Aspremont and Gérard-Varet (1998); Laux (2001) allow multi-dimensional effort, but their agents have no private information. Coviello, Ichino, and Persico (2011) study dynamic scheduling of many tasks under diseconomies of scope.

[8]For example: Alchian and Demsetz (1972); Mirrlees (1976); Holmström (1982); Holmström and Milgrom (1991); Legros and Matsushima (1991); McAfee and McMillan (1991); Legros and Matthews (1993); Miller (1997); d'Aspremont and Gérard-Varet (1998); Laux (2001); Battaglini (2006); Matsushima, Miyazaki, and Yagi (2010); Rahman and Obara (2010). Another literature sees bonuses and penalties as financially equivalent, using reference-dependent preferences to distinguish the incentive effects (e.g., Aron and Olivella, 1994; Frederickson and Waller, 2005).

socially costless to implement. In particular, *without affecting team welfare*, whenever one of his tasks failed inspection, an agent could be forced to pay arbitrarily high transfers to a third agent (not the monitor of that task). Empty promises arise in our setting because with socially costly sanctions, there is a tradeoff between performing all feasible tasks and providing the necessary incentives to do so.

Our results are organized as follows. First, we examine the basic structure of optimal contracts when there are enough supervisory resources to monitor every task. 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 only up to p^* feasible tasks. The cutoff p^* is increasing in the probability, λ, that any given task is feasible. Empty promises optimally arise for an intermediate range of λ; they 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 large.[9] 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 and are working on independent tasks, we show that a statistical complementarity arises whenever it is optimal for either of them to make empty promises. Then it is strictly optimal for one of them to specialize in performing tasks while the other specializes in monitoring. It is important to emphasize that the strict optimality of having a monitoring specialist—a supervisor—arises from the fact that it is optimal to make empty promises. If there are no empty promises, it does not matter how monitoring responsibility is allocated. Section 4 then shows how to economize on monitoring, studying the tradeoff between allocating capacity to monitoring or production. With N agents, we show that at most the capacity of one agent is used towards monitoring. Even though promise-keeping can be implemented using only two units of capacity for monitoring, there are gains from allocating more units to monitoring and employing empty promises, even if the probability of task feasibility is very high.

For expositional simplicity and tractability, we make a few stylized modeling assumptions that do not qualitatively affect our main conclusions regarding empty promises and endogenous supervision structures. For instance, our main results assume that monitoring is costless (aside from its opportunity cost, of course), and therefore take for granted that agents are willing to monitor

[9]The kinked-linear structure of the corresponding contract bears a similarity to the debt-like contracts with low-powered incentives and random verification arising in the costly state verification literature (initiated by Townsend 1979). For different reasons, low-powered incentives arise in both cases: for Townsend, it is because monitoring is costly, whereas here it is because punishments are not transfers.

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 sanction 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), reallocating capacity (Section 5.3), and to imperfections in monitoring (Section 5.4).

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 an expected benefit $\frac{b}{N}$, where $b > c > \frac{b}{N}$. Hence each task is socially beneficial, but no agent will 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.[10] 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 $\tau = 1$ 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.[11] The monitoring agent, however, cannot distinguish whether the task was infeasible or intentionally shirked.

[10] The spirit of this assumption is that agents discuss and agree on who will promise which tasks; formally it may be simpler to assume they make promises sequentially, in arbitrary order.

[11] Perfect monitoring simplifies the exposition; we discuss imperfect monitoring later.

- At $\tau = 4$, the agents reveal the results of their inspections.[12]
- At $\tau = 5$, each agent can impose unbounded sanctions on other agents, at no cost to himself.

We consider a setting in which it is not possible to commit to transfers that are contingent on inspection outcomes; instead, any sanction imposed on an agent is pure waste. We study perfect Bayesian equilibria of this game. Since the sanctions 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. Moreover, by the revelation principle it is without loss of generality to restrict attention to equilibria in which agents reveal their inspection results truthfully at time $\tau = 4$. Similarly, since monitoring is costless, we may ignore deviations from the equilibrium monitoring scheme at time $\tau = 3$. 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 $\mathcal{P}(Z)$ 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 $P \in \times_{i=1}^{N} \mathcal{P}(\mathcal{X})$, with each distinct P_i and P_j disjoint in $\mathcal{X}$, specifying which tasks should be promised by which agents;*
2. *A task completion strategy $s_i : \mathcal{P}(P_i) \to \Delta(\mathcal{P}(P_i))$ for each agent i, with every realization a subset of the argument, specifying the subset of her promises to complete among those feasible;*
3. *A monitoring scheme $r \in \Delta \times_{i=1}^{N} \mathcal{P}\big(\bigcup_{j \neq i} P_j\big)$, with every realization a vector of disjoint subsets of $\mathcal{X}$, specifying a joint distribution over which agents monitor which tasks;*
4. *A sanctioning scheme $v : \times_{i=1}^{N} \big(\mathcal{P}(P_i) \times \mathcal{P}(P_i)\big) \to \mathbb{R}_{-}^{N}$, specifying the net sanction imposed on each player $i = 1, \ldots, N$ as a function of which tasks were shown to pass and fail inspection.*

Remark 1 (Contractual interpretation)**.** *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 sanctioning scheme. Then*

[12]Whether inspection results are verifiable does not matter. In Section 5.1, we show that our analysis is robust to monitoring costs incurred to (verifiably) prove a task has failed inspection.

it should be a perfect Bayesian 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 sanctioning scheme. We investigate this principal-agent interpretation in Section 6.

A contract must respect each agent's bounded capacity. A contract is *feasible* if it satisfies $|P_i| + \max_{R \in \text{supp } r} |R_i| \leq M$ 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 sanctioning schemes; and no agent has an incentive to fail to report his inspection results truthfully. A contract is *optimal* if it maximizes the team's aggregate utility within the class of feasible and incentive compatible contracts.

Remark 2 (Randomization, noncontingent transfers, and individual rationality)**.** *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 could 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) were randomly assigned. Alternatively, by using ex ante (noncontingent) transfers, it would be easy to spread the wealth so as to make everyone willing to accept the contract, no matter how asymmetric were the roles. In light of these possibilities, 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 sanction depends only on his tasks that failed inspection, so without loss of generality $v_i : \mathcal{P}(P_i) \to \mathbb{R}_-$;*
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 $|s_i(A)| \leq |s_i(A')|$.*

It follows from part 2 of Lemma 1 that agents have no disincentive to report their inspection results truthfully, enabling us to henceforth ignore the incentive constraints for truthful revelation.

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 $f \subseteq P_i$ 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)

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

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. A 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 involves *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 $p_i^* < p_i$. In a contract, if one of these descriptors applies to all the agents' strategies, then the descriptor applies to the contract as well.

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 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 optimal worker-supervisor contracts.

Theorem 1. *Conditional on a worker-supervisor structure, there is an optimal contract such that:*

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

Theorem 1 says that the optimal worker-supervisor 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 λ, is strictly positive whenever $1 - (2 - \frac{c}{b/2})^{1/M} < \lambda$, and is strictly smaller than the number of promises made whenever $\lambda < (\frac{c}{b/2} - 1)^{1/M}$. Recall that for $N = 2$, $\frac{b}{2} < c < b$ and so $\frac{c}{b/2} \in (1, 2)$. Hence, the interval of λ's for which there are empty promises increases with both the capacity and cost-benefit ratio. Indeed:

Corollary 1. *Conditional on a worker-supervisor structure, for any $\lambda \in (0, 1)$ and $\frac{c}{b}$ there exists $M < \infty$ sufficiently large that an optimal contract has empty promises; for any $\lambda \in (0, 1)$ and $M < \infty$ there exists $\frac{c}{b} < 1$ sufficiently large that an optimal contract has empty promises.*

We prove Theorem 1 below. To understand the intuition for 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 sanctions anyway. Since sanctions are costly, it is possible to reduce the cost of sanctions 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 being sanctioned is three failures, then he will never fulfill more than eight promises, even if all ten are feasible. When λ 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 $A' \subset s(A)$,

$$\nu(P \setminus s(A)) + |s(A)|\left(\tfrac{b}{N} - c\right) \geq \nu(P \setminus A') + |A'|\left(\tfrac{b}{N} - c\right). \tag{3}$$

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 the expected sanction is minimized under the *kinked linear* sanctioning scheme $v(P \setminus s(A)) = (\frac{b}{2} - c) \max\{|P \setminus s(A)| - (p - p^*), 0\}$, which imposes no sanction when p^* or more tasks are completed, but a sanction of $(\frac{b}{2} - c)(p^* - s(A))$ whenever $|s(A)| < p^*$. This sanctioning 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 sanctioning scheme into Eq. 1, the team's welfare reduces to

$$p^* \big(\tfrac{b}{2} - c\big) + \tfrac{b}{2} \sum_{a=0}^{p} \tbinom{p}{a} \lambda^a (1-\lambda)^{p-a} \min\{a, p^*\}. \tag{4}$$

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 λ, leading to the monotone comparative statics in point (2). Given that p^* is increasing in λ, 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 algebra. □

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

$$\begin{aligned} p^* \big(\tfrac{b}{2} - c\big) &+ \tfrac{b}{2} \sum_{a=0}^{p} \tbinom{p}{a} \lambda^a (1-\lambda)^{p-a} \min\{a, p^*\} \\ &\geq (p^* + 1) \big(\tfrac{b}{2} - c\big) + \tfrac{b}{2} \sum_{a=0}^{p} \tbinom{p}{a} \lambda^a (1-\lambda)^{p-a} \min\{a, p^* + 1\}. \end{aligned} \tag{5}$$

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. 20 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 λ, and that the optimal sanctioning 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. *Conditional on complete monitoring, there is an optimal contract such that each agent i has a cutoff strategy for task completion, doing at most p_i^* feasible tasks and the sanctioning scheme is kinked-linear. Each agent's cutoff p_i^* is increasing in λ.*

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 $N > 2$, more intricate possibilities exist. As the following result shows, in the presence of empty promises these contracts are not payoff-identical.

Theorem 2. *Conditional on complete monitoring, a worker-supervisor contract is optimal when there are two agents. When $N > 2$ agents, the division of labor in an optimal contract includes at least one supervisor (who specializes in monitoring). Among the workers who promise tasks, if agent i makes more promises than j ($p_i > p_j$) then he 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 λ is such that there are empty promises under the optimal worker-supervisor contract.*

As seen in the proof below, whenever empty promises are optimal, supervision endogenously arises for statistical reasons, despite the symmetry of players and independence of tasks. It is particularly intuitive to consider the case in which M and p^* are even, and compare a worker-supervisor contract in which the worker uses the cutoff $p^* < M$, to a symmetric contract in which both players make $M/2$ promises and use the cutoff $p^*/2$. 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. However, 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 the one in the worker-supervisor contract. The same issue arises when comparing any two arbitrary contracts with the same total number of promises and the same total (positive) 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

$$\sum\nolimits_{i=1}^N \Big(p_i^* \big(\tfrac{b}{N} - c \big) + \tfrac{N-1}{N} b \sum\nolimits_{a=0}^{p_i} \tbinom{p_i}{a} \lambda^a (1-\lambda)^{p_i - a} \min\{a, p_i^*\} \Big). \tag{6}$$

Consider the case where $p_k^* < p_k$ for some agent k, and imagine a different complete monitoring contract (with a corresponding optimal sanctioning scheme) 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 their 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

$$\sum\nolimits_{a=p^*+1}^{p+1} \tbinom{p+1}{a} \lambda^a (1-\lambda)^{p+1-a} \; - \sum\nolimits_{a=p^*+1}^{p} \tbinom{p}{a} \lambda^a (1-\lambda)^{p-a} > 0, \tag{7}$$

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] Finally, 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. □

There are two different ways to interpret complete monitoring. Under one interpretation,

[13] The Lorentz-Fan rearrangement inequality says if $f : \mathbb{R}^k \to \mathbb{R}$ is a supermodular function, then for any collection of vectors $(x^1, \ldots, x^n)$, $\sum_{i=1}^n f(x^i) \leq \sum_{i=1}^n f(x^{*i})$, where x^{*i} is the "majorized" vector which, for every dimension k, contains the ith largest component among $x_k^1, \ldots x_k^n$. Here $x^i = (p_i, p_i^*)$ and f transforms x^i into a summand in Eq. 6.

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 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 her 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 worker-supervisor arrangement was optimal). With too little monitoring, however, they may not be able to implement finely-tuned sanctioning 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?

The ability to trade off performance and monitoring generates some difficulties. If fewer tasks are being monitored than an agent promised, it will not generally be possible to compute the optimal sanctioning scheme and task completion strategies as we did in the previous section. While under complete monitoring it is simple to make all relevant incentive constraints bind, under incomplete monitoring the sanctioning scheme generally has too few degrees of freedom relative to the task-completion strategy. Compounding this problem, there is a gap between the sanctioning scheme and the expected sanction: conditional on the tasks a player completed, inspection outcomes depend probabilistically on the monitoring distribution. Because sanctions are restricted to be negative, it may be impossible to generate the expected sanctioning scheme that would make a given combination of incentive constraints bind.[14]

[14]In fact, we have numerical examples where the optimal contract does not have a kinked-linear sanctioning scheme and cutoff strategies. Theorem 4 in the supplemental appendix shows that kinked-linear sanctions and cutoff strategies do remain optimal when the magnitude of sanctions are restricted to be increasing and convex in the number of failed inspections, as might be the case in settings where sanctions are imposed by third parties who are more inclined to exact punishment if they perceive a consistent pattern of failures. We prove this by showing that player i's expected

Nonetheless, we are able to characterize several key features of optimal contracts in this general environment. The following theorem shows that our basic results on empty promises and endogenous supervision structures extend when monitoring may be incomplete and stochastic.

Theorem 3. *Consider any capacity size M, number of agents N, and cost-benefit ratio $\frac{c}{b}$. When units of capacity are substitutable between performance and monitoring, the following hold:*

1. *An optimal contract allocates at least two and at most M units of capacity to monitoring.*
2. *Conditional on promise-keeping, an optimal contract has two "partial supervisors" who each monitor one task.*
3. *An optimal contract has empty promises if at least three units of capacity are allocated to monitoring.*
4. *Conditional on allocating M units of capacity to monitoring, an optimal contract satisfies the results of Corollary 2 on sanctioning schemes, cutoff strategies, and monotonicity of p_i and p_i^* in λ for each agent i; moreover,*
   (a) *if $N = 2$, an optimal contract assigns monitoring responsibilities to a single supervisor;*
   (b) *if $N > 2$, then for any $\varepsilon > 0$ there is $\underline{M} < \infty$ such that if $M \geq \underline{M}$, a contract with one supervisor and $N - 1$ agents making M promises and using a cutoff strategy with*

$$p^* = \left\lceil M\lambda - \sqrt{2M(1-\lambda)\lambda}\ \operatorname{erf}^{-1}\left(\frac{1+N-\frac{c}{b/2}N}{1-N}\right)\right\rceil \tag{8}$$

   attains a $1-\varepsilon$ fraction of the welfare of an optimal contract.[15]

5. *For any M and N, there is $\lambda^* \in (0,1)$ such that the single-supervisor contract with empty promises in Point 4 dominates all promise-keeping contracts whenever $\lambda \leq \lambda^*$. If λ is sufficiently high given M, N, and $\frac{c}{b}$, then promise-keeping is optimal. But empty promises may be optimal even when λ is close to 1 if the cost-benefit ratio of tasks is moderately high: for any M and N, empty promises are optimal even in a neighborhood of $\lambda = \frac{M-2}{M-1}$ when $\frac{c}{b} > \frac{2+(e/N)}{2+e}$ (which converges to approximately 0.42 as $N \to \infty$).*

Proof. See the appendix. Some intuition for the proof is provided below. □

---

sanctioning scheme 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 sanctioning scheme. That expression can be written in terms of the expected number of discovered unfulfilled promises above a threshold, and is optimally implemented by a kinked-linear sanctioning scheme.

[15] erf^{-1} is the inverse of the Gaussian error function.

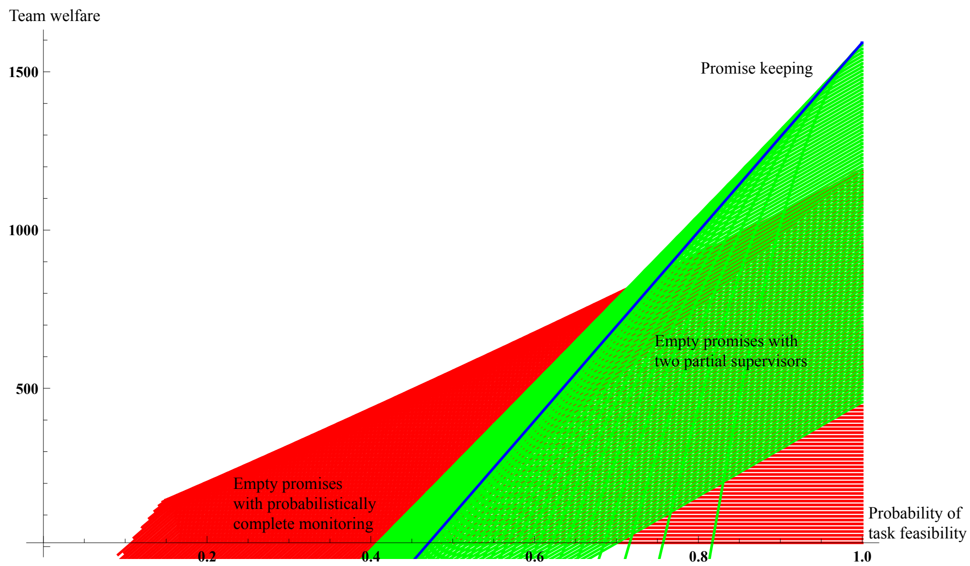

FIGURE 1: Capacity $M = 100$, number of players $N = 4$, task benefit $b = 10$, task cost $c = 6$. The blue line is the social value of promise keeping for each λ. As a function of λ, for workers using a cutoff of $p^* \in \{1, \ldots, M-1\}$, each green curve is the social value of a partial supervision contract with two promise-keeping supervisors who each monitor one task; while red is the social value of a contract with one supervisor who monitors all the tasks of a given worker with probability $\frac{1}{N-1}$.

We begin by drawing some insights from our analysis of complete monitoring. Note that the same expected sanctions (and cutoff strategies) available under complete monitoring are available even when only M units of capacity are allocated to monitoring, simply by employing a correlated randomization device to determine which player to monitor using all the other players' monitoring capacity, and then scaling the sanctioning scheme appropriately. That is, for each player i there is an $\alpha_i \in (0, 1]$ such that with probability α_i all of i's tasks are monitored, and with probability $1 - \alpha_i$ none of her tasks are monitored. Then her new sanctioning scheme is v_i multiplied by $1/\alpha_i$, so that, taking expectations over whether she will be monitored, her *expected sanctioning scheme* is exactly v_i.[16] Hence devoting more resources toward monitoring reduces capacity available for

[16]Depending on when capacity must be allocated, correlation effects may further favor having a single supervisor under this type of 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(4) shows that a single supervisor optimally arises even if the realization of the monitoring scheme is kept secret until $\tau = 3$.

performing tasks without improving over complete monitoring with regard to incentives. This yields point (1).

Moreover, our previous analysis extends to show that cutoff strategies are optimal, conditional on optimally allocating M units of capacity to monitoring in the manner described above. As in the previous section, it is best to assign all monitoring to one supervisor when $N = 2$. When $N > 2$, we use a continuous, normal approximation to the binomial distribution to show that having a single supervisor is approximately optimal (leaving open the conjecture that it is in fact exactly optimal). This yields point (4).

Conditional on implementing promise keeping, monitoring can be minimized (and therefore the number of tasks completed can be maximized) 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 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 sanctioning scheme is linear so that all agents are willing to complete all their feasible promises, and all incentive constraints bind. Therefore, this is the optimal way to implement promise keeping, yielding point (2). Point (3) then follows from point (2): whenever more than two units capacity are devoted to monitoring in an optimal contract, it must be the case that promise keeping is suboptimal. When two units of capacity are used, the optimal task completion cutoff may still be strictly smaller than p even if λ is very high (see Figure 1), as the same two units of monitoring may be used to provide a forgiving contract instead.

Point (5) shows that contracts with empty promises are optimal for a range of parameters. First, if λ is moderately low, the promise keeping is dominated by assigning a single supervisor to use his full capacity to monitor the other agents. Even for λ very close to 1, empty promises dominate promise keeping if M and $\frac{c}{b}$ are sufficiently large; in particular, promise keeping is dominated by a forgiving, empty promises contract with two partial supervisors, each allocating one unit of capacity to monitoring. Figure 1 illustrates point (5) for the case of $N = 4$, $M = 100$, and $\frac{c}{b} = \frac{3}{5}$, showing that promise keeping is dominated for nearly any λ by the envelope of some simple contracts with empty promises. For λ between about 0.1 and about 0.7, a single-supervisor contract as in point (4) dominates promise keeping. This is because when λ is low, agents are likely to find many of their tasks to be infeasible. For λ above 0.7, tasks are too valuable to "waste" M units of capacity on monitoring, but even with only two units of capacity allocated to monitoring it is still optimal to implement a forgiving contract with empty promises unless λ is extremely close to 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 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 show that a task failed his inspection. To induce him to exert monitoring effort, the team can add an additional stage, $\tau = 6$, to their interaction. After the sanctions for failed tasks are implemented in $\tau = 5$, in $\tau = 6$ 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 $\tau = 6$ that was not reported as failing inspection in $\tau = 4$, 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 sanction 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.[17] In our model, incorporating messages can reduce the amount of monitoring needed. Consider the case in which 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

[17] For example: Jackson and Sonnenschein (2005); Chakraborty and Harbaugh (2007); Matsushima, Miyazaki, and Yagi (2010); Frankel (2011).

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.

Monitoring schemes with messages allow more resources to be devoted to performing tasks while still maintaining the optimal expected sanctioning scheme 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 λ, 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 unused units of capacity which could be allocated towards completing tasks. Since another agent must monitor him, optimal supervision structures with two partial supervisors in a team of N agents are again likely to arise.

5.3 Reallocating tasks

We have assumed that if an agent promises a task that ends up being infeasible, she cannot reallocate that unit of capacity toward monitoring. Recall that the optimal way to implement promise-keeping in our original setting is via a linear contract that makes all incentive constraints bind. As with point (2) of Theorem 3, for a linear contract the same schedule of expected sanctions can be maintained by scaling the actual sanctions inversely with the number of monitoring slots employed. Consequently, the opportunity to allocate additional capacity towards monitoring is useful for reducing expected sanctions only if empty promises are optimal.

In our model, a task is either feasible or infeasible, regardless of whom it is assigned to. But there could be interesting team settings where the feasibility of any given task is idiosyncratic to each agent. This raises the possibility that agents could exchange tasks, to see whether the tasks that are infeasible for one might be feasible for another. Allowing for this would significantly alter the model, since it would require inserting both a task-trading phase and an additional task performance phase in between $\tau = 2$ and $\tau = 3$. At an intuitive level, however, allowing task trading would simply change the distribution over how many feasible tasks an agent might find, and an agent would still face a positive probability of finding fewer than he is willing to perform. Moreover, task-trading introduces the new incentive constraint that an agent should not want to trade away a feasible task unless that he is supposed to fulfill. So there is still the same incentive problem of motivating him to perform tasks rather than claim they are infeasible for him. Qualitatively, the same kinds of results on empty promises and concentrated supervision would arise.

5.4 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 $\gamma \in (0,1)$? In this case, the characterization of the optimal contract, conditional on M units of monitoring, continues to hold for γ not too far from one; and the other points in Theorem 3 continue to hold independently of γ.

The optimality of empty promises in Theorem 3 is unaffected by imperfect monitoring because the expected social welfare of the contract shown to dominate promise keeping is independent of γ. This contract is such that the sanctioning scheme 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 sanction 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 γ by scaling the sanction $v(F)$ by the factor $\frac{1}{\gamma^F}$.

Our characterization of the optimal contract, conditional on M units of monitoring, relies on being able to find a sanctioning scheme under which the expected sanctioning scheme makes all relevant incentive constraints bind. When the monitoring technology is sufficiently imperfect, such a 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 sanction would be zero 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 sanctioning scheme $(v(0), v(1), v(2), v(3))$ should satisfy

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

where the a-th row and the f-th column of the 4×4 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 sanctioning scheme that generates the desired expected sanctioning scheme for any γ 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 sanctions (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 sanctions. 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 sanctioning schemes; see 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, which is normalized to zero. 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

$$\sum_{i=1}^{N} \left(-t_i + \sum_{A \subseteq P_i} \lambda^{|A|} (1-\lambda)^{p_i - |A|} |s_i(A)| (B - b) \right) \tag{10}$$

For the usual reasons, IR must bind in an optimal contract.[18] Substituting the binding IR constraints into Eq. 10 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 informal sanctions) 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 sanctioning 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 sanctions 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. As for the form of the firm's optimal contract, since $b \leq B$ is required for profitability, our previous conclusions still hold—empty promises arise for an intermediate range of λ. For settings like Figure 1, empty promises dominate promise keeping unless λ is very high.

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 sanctions at the individual level. That is, industries where production is complex and requires accumulated job-specific human capital, as discussed in the introduction. For a striking example, Knez and Simester (2001) show that introducing a firm-level bonus scheme, complemented by peer monitoring and informal sanctions, increased on-time performance at Continental Airlines in the mid-1990s. The firm-wide bonus, coupled with the highly interdependent nature of on-time

[18]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 sanctions in order to marginally reduce the bonus.

performance, provided each workgroup sufficient incentives to collectively prefer a high-effort equilibrium. At the individual level, high effort was supported by informal sanctions, where the members of each workgroup would "monitor and sanction their colleagues to enforce the group decision" (p. 746).

7 Discussion

We study a model of teams in which agents optimally make empty promises, and are "forgiven" for having done so. Doing so buffers against the potential infeasibility of tasks, thereby minimizing costly sanctions. Because such buffering reduces the number of completed tasks, empty promises would not be optimal if formal, performance-contingent transfers could be made. This optimal equilibrium phenomenon is robust to a tradeoff between performance and monitoring, even though empty promises use up 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 socially optimal task completion strategies. Simply stated, there are increasing returns to a worker's task load when he makes empty promises: doubling both his number of promises and his 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 λ 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). Of course, if an agent who is better at performing tasks could also train other agents, and if supervising and teaching are complementary, then it may be optimal for the most productive team members to supervise.

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.[19] 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.[20] 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 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).

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.

[19] A seminal paper by Miller (1956) suggests working memory capacity is about 7 ± 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 ± 1 chunks for more complex chunks. The economic literature studying imperfect memory includes Dow (1991); Piccione and Rubinstein (1997); Hirshleifer and Welch (2001); Bénabou and Tirole (2002); Wilson (2004); Kocer (2010). Mullainathan (2002) and Bodoh-Creed (2010) study updating based on data 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., Piccione and Rubinstein, 1993; Cole and Kocherlakota, 2005; Compte and Postlewaite, 2008; Romero, 2011), as well as work on self-delusion in groups (e.g., Bénabou, 2008).

[20] 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., Maskin and Tirole, 1999; Tirole, 2009; Bolton and Faure-Grimaud, 2010), communication complexity (e.g., Segal, 1999), and contractual complexity (e.g., Anderlini and Felli, 1998; Battigalli and Maggi, 2002).

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 Bénabou. Groupthink: Collective delusions in organizations and markets. *Working Paper*, December 2008. URL http://goo.gl/oEt3K.

Roland Bénabou 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://goo.gl/jrfhe.

Patrick Bolton and Antoine Faure-Grimaud. Satisficing contracts. *Review of Economic Studies*, 77(3):937–971, July 2010.

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.

Decio Coviello, Andrea Ichino, and Nicola Persico. Time allocation and task juggling. Working paper, November 2011. URL http://goo.gl/8bUuV.

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, September 2011. URL http://goo.gl/kA5wj.

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.

Barton H. Hamilton, Jack A. Nickerson, and Hideo Owan. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497, June 2003.

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.

Marc Knez and Duncan Simester. Firm-wide incentives and mutual monitoring at Continental Airlines. *Journal of Labor Economics*, 19(4):743–772, October 2001.

Yilmaz Kocer. Endogenous learning with bounded memory. Working paper, November 2010. URL http://goo.gl/KdwaL.

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://goo.gl/lwtBE.

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

Julian Romero. Bounded rationality in repeated games. Working paper, March 2011. URL http://goo.gl/HrSom.

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.

Robert M. Townsend. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293, October 1979.

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://goo.gl/SkarN.

Appendix A Proofs

Proof of Lemma 1. (*Part 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.

(*Part 2*) First, for agents to reveal their inspection results truthfully, their sanction must not depend on their announcements at time $\tau = 4$. Moreover, conditional on the monitoring scheme, an agent has no influence over whether other agents' tasks pass or fail inspection. So for any sanctioning scheme that depends on other agents' outcomes, it is equally effective to employ a modified sanctioning scheme in which the agent's sanction is conditioned only on his own outcomes, where the sanctioning scheme is set to offer the same expected sanctioning scheme 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 sanction regardless of passed inspections, where the sanctioning scheme is set to offer the same expected sanctioning scheme as the original contract.

(*Part 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.

(*Part 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, $s_i(A') \not\subseteq s_i(A)$. But then a superior contract can be constructed by slightly relaxing the sanctions when tasks in $s_i(A') \backslash s_i(A)$ fail inspection. The agent, facing a strictly milder sanctioning scheme, is strictly better off. For a sufficiently small relaxation of the sanctioning scheme, incentive compatibility still holds. $\square$

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 promises, and tasks are drawn uniformly for monitoring, then the probability agent i will have y failed inspections is given by the compound hypergeometric-binomial distribution

$$g(y,x) = \sum\nolimits_{k=y}^{F_i} \frac{\binom{p_i-x}{k}\binom{x}{F_i-k}}{\binom{p_i}{F_i}} \binom{k}{y} \gamma^y (1-\gamma)^{k-y}. \tag{11}$$

To interpret Eq. 11, 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 sanction 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 sanctioning 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 $(p_i - |a|)\frac{\gamma F}{p_i}$. Conditional expected sanctions are independent of F, for $F \geq 1$. This contract gives expected social utility

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{x=0}^{M}\tbinom{p_i}{x}\lambda^x(1-\lambda)^{p_i-x}[(b-c)x + (p_i - x)(\frac{b}{N} - c)] \\
&\quad = \left(\tfrac{b}{N} - c\right)\sum_{i=1}^{N} p_i \sum_{x=0}^{M}\tbinom{p_i}{x}\lambda^x(1-\lambda)^{p_i-x} + \tfrac{N-1}{N}b\sum_{x=0}^{p_i}\tbinom{p_i}{x}\lambda^x(1-\lambda)^{p_i-x}x \\
&\quad = \left(\tfrac{N-1}{N}b\lambda + \tfrac{b}{N} - c\right)\sum_{i=1}^{N} p_i. \qquad (12)
\end{aligned}$$

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 $p_{N-1} = p_N = M-1$. $\square$

Lemma 3. *Consider a sanctioning scheme 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*

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

when the cutoff p_i^* is induced. The value of Eq. 13 is strictly increasing and concave in λ. 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 λ.

Proof of Lemma 3. That a cutoff strategy is induced follows from Theorem 4. 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 $g(F, \min\{x, p_i^*\}) = \lambda^F \binom{M-\min\{x,p_i^*\}}{F} \Big/ \binom{M}{F}$, which is convex:

$$\binom{M-\min\{x+1,p_i^*\}}{F} - 2\binom{M-\min\{x,p_i^*\}}{F} + \binom{M-\min\{x-1,p_i^*\}}{F}$$
$$= \begin{cases} \binom{M-x}{F}\left(\frac{F}{M-(x+1)-F} - \frac{F}{M-x}\right) & \text{if } x \leq p_i^* - 1, \\ \binom{M-(p_i^*-1)}{F} - \binom{M-p_i^*}{F} & \text{if } x = p_i^*, \\ 0 & \text{if } x \geq p_i^* + 1. \end{cases} \quad (14)$$

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

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

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

$$\frac{(c - \frac{b}{N})g(F, \min\{x, p_i^*\})}{g(F, p_i^*) - g(F, p_i^* - 1)}. \quad (15)$$

Since λ 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

$$\frac{g(F, s_{p_i^*+1}(x+1)) - g(F, s_{p_i^*+1}(x))}{g(F, p_i^*+1) - g(F, p_i^*)} - \frac{g(F, s_{p_i^*}(x+1)) - g(F, s_{p_i^*}(x))}{g(F, p_i^*) - g(F, p_i^*-1)}$$
$$= \begin{cases} 0 & \text{if } x \geq p_i^* + 1 \\ 1 & \text{if } x = p_i^* \\ \frac{g(F,x+1)-g(F,x)}{g(F,p_i^*+1)-g(F,p_i^*)} - \frac{g(F,x+1)-g(F,x)}{g(F,p_i^*)-g(F,p_i^*-1)} & \text{if } x \leq p_i^* - 1. \end{cases} \quad (16)$$

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

$$\binom{M-p_i^*}{F} - \binom{M-p_i^*+1}{F} > \binom{M-p_i^*+1}{F} - \binom{M-p_i^*+2}{F}. \quad (17)$$

But this is exactly analogous to the earlier calculation. □

Proof of Theorem 3. (*Part 1*) This is clear from the discussion in the text.

(*Parts 2 and 3*) These follow from Lemma 2.

(*Part 4*) 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 λ 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

$$\begin{gathered}\max_{\{\tilde{p}_i, \tilde{p}_i^* \in \mathbb{R}\}_{i=1}^N} \sum\nolimits_{i=1}^n \mathbb{E}\Big((b-c)\min\{a, \tilde{p}_i^*\} + \left(\tfrac{b}{N} - c\right)\max\{\tilde{p}_i^* - a, 0\}\Big) \\ \text{s.t. } \tilde{p}_i^* \leq \tilde{p}_i \leq M \text{ for all } i \text{ and } \sum\nolimits_i \tilde{p}_i \leq M(N-1),\end{gathered} \tag{18}$$

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

$$\frac{\partial E_i}{\partial \tilde{p}_i^*} = \frac{1}{2N}\Big(b + bN - 2cN + b(N-1)\operatorname{erf}\big(\tfrac{\lambda\tilde{p}_i - \tilde{p}_i^*}{\sqrt{2\tilde{p}(1-\lambda)\lambda}}\big)\Big) = 0, \tag{19}$$

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

$$\tilde{p}_i^* = \tilde{P}^*(\tilde{p}_i) \equiv \tilde{p}_i\lambda - \sqrt{2\tilde{p}_i(1-\lambda)\lambda}\operatorname{erf}^{-1}\big(\tfrac{b+bN-2cN}{b-bN}\big), \tag{20}$$

and the welfare arising from each agent i is

$$\tilde{p}_i\lambda(c-b) - be^{-\operatorname{erf}^{-1}\left(\frac{b+bN-2cN}{b-bN}\right)^2}\frac{N-1}{N}\sqrt{\frac{\tilde{p}_i\lambda(1-\lambda)}{2\pi}}. \tag{21}$$

The strict second order condition is satisfied globally:

$$\frac{\partial^2 E_i}{\partial(\tilde{p}_i^*)^2} = -be^{-\frac{(\lambda\tilde{p}_i - \tilde{p}_i^*)^2}{2\tilde{p}(1-\lambda)\lambda}}\frac{N-1}{N}\sqrt{\frac{1}{2\pi\tilde{p}(1-\lambda)\lambda}} < 0. \tag{22}$$

We now move to the outer part of the continuous problem: choosing p_i. By the envelope theorem, $\frac{dE_i}{d\tilde{p}_i}\big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)} = \frac{\partial E_i}{\partial \tilde{p}_i}\big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)}$ and $\frac{d^2E_i}{d\tilde{p}_i^2}\big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)} = \big(\frac{\partial^2 E_i}{\partial \tilde{p}_i^2} + \frac{\partial^2 E_i}{\partial p_i \partial \tilde{p}_i^*}\frac{d\tilde{P}^*}{d\tilde{p}_i}\big)\big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)}$. Solving the closed form of d^2E_i/dp_i^2 at $\tilde{p}_i^* = \tilde{P}^*(\tilde{p}_i)$ shows the objective is strictly convex in each $\tilde{p}_i$:

$$\frac{d^2E_i}{dp_i^2}\Big|_{\tilde{p}_i^*=\tilde{P}^*(\tilde{p}_i)} = \frac{1}{4\tilde{p}_i^2} be^{-\text{erf}^{-1}\left(\frac{b+bN-2cN}{N-bN}\right)^2} \frac{N-1}{N}\sqrt{\frac{\tilde{p}_i\lambda(1-\lambda)}{2\pi}} > 0. \tag{23}$$

Since in addition $dE_i/d\tilde{p}_i > 0$ at $\tilde{p}_i^* = \tilde{P}^*(\tilde{p}_i)$, and M units of 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 $\tilde{p}^* \equiv \tilde{P}^*(M)$, while the Nth agent promises zero tasks (i.e., only monitors). Now we construct a contract for the true (discrete) model, using the same promise scheme: $N-1$ agents each make M promises, while one agent supervises. The number of earnest promises must 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 sanctions 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 sanctions may be harsher 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 Φ and ϕ be the CDF and PDF of the normal distribution, and $\hat{\Phi}$ and $\hat{\phi}$ be the CDF and PDF of the binomial. The tail benefit (which is not affected by approximation error) is

$$X = -\int_{-\infty}^{-1/2} \big((\delta - \rho)(\lceil \tilde{p}^* \rceil - \tilde{p}^*) + \delta a\big)\phi(a)\, da. \tag{24}$$

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

$$Y = \rho\Big(\big(1 - \hat{\Phi}(\lceil \tilde{p}^* \rceil - 1)\big)\lceil \tilde{p}^* \rceil - \big(1 - \Phi(\lceil \tilde{p}^* \rceil - \tfrac{1}{2})\big)\tilde{p}^*\Big). \tag{25}$$

The integer deficit, accounting for approximation error, is

$$Z = \sum_{a=0}^{\lceil\tilde{p}^*\rceil-1}\Big(\int_{a-1/2}^{a+1/2}\big((\delta-\rho)(\lceil\tilde{p}^*\rceil-\tilde{p}^*)+\delta\tilde{a}\big)\phi(\tilde{a})\,d\tilde{a}-\delta a\hat{\phi}(a)\Big). \tag{26}$$

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:

$$\begin{aligned}-\tfrac{1}{N-1}(\hat{V}-\tilde{V}) = Z-X-Y = &-(\lceil\tilde{p}^*\rceil-\tilde{p}^*)\big(\rho-\delta\Phi(\lceil\tilde{p}^*\rceil-\tfrac{1}{2})\big)\\ &-\rho\big(\Phi(\lceil\tilde{p}^*\rceil-\tfrac{1}{2})-\hat{\Phi}(\lceil\tilde{p}^*\rceil-1)\big)\lceil\tilde{p}^*\rceil\\ &-\sum_{a=0}^{\lceil\tilde{p}^*\rceil-1}\delta a\hat{\phi}(a)+\int_{-\infty}^{\lceil\tilde{p}^*\rceil-1/2}\delta a\phi(a)\,da\end{aligned} \tag{27}$$

Notice that the right hand side of 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 Φ and $\hat{\Phi}$. By the De Moivre–Laplace theorem, the approximation error is on the order of $M^{-1/2}$. Since by Eq. 20 and Eq. 21 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 $M\to\infty$. 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

$$\frac{1}{N}\sum_i\begin{pmatrix}\gamma\big(\Phi(\lceil\tilde{p}^*\rceil-\frac{1}{2})-\hat{\Phi}(\lceil\tilde{p}^*\rceil-1)\big)\lceil\tilde{p}^*\rceil\\ +\sum_{a=0}^{\lceil\tilde{p}^*\rceil-1}\delta a\hat{\phi}(a)-\int_{-\infty}^{\lceil\tilde{p}^*\rceil-1/2}\delta a\phi(a)\,da\end{pmatrix}, \tag{28}$$

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 $M\to\infty$.

(*Part 5*) First observe that 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 harsh enough sanctions to make it incentive compatible for them to do so, but the sanctions may be arbitrarily severe since they are not realized on the equilibrium path. The value of any such contract is $(NM-2)(b-c)$. Consider fixed M, b, and c. The value of a contract is continuous in λ, P, s, and ν. 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 ν 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 λ, b, c, P, ν, and s, the constraint set is compact-valued for each λ, b, and c. Finally, the constraint set is nonempty because it always contains the contract in which no tasks are promised and no sanctions 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 λ and the correspondence mapping λ to the set of optimal contracts (ν, s, and ρ) is upper hemicontinuous. Hene as $\lambda \to 1$, 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 λ sufficiently high. To minimize the cost of sanctions, 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.

Next, we show that empty promises with M units of 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 contract in which one agent devotes M units of capacity to monitoring while $N-1$ agents each make M promises and keep $M-1$ promises. The value of promise keeping, $(MN-2)\left(\frac{N-1}{N}b\lambda + \frac{b}{N} - c\right)$, is zero at λ^*. The value of this single-supervisor contract is

$$(N-1)\Big((M-1)\left(\tfrac{b}{N}-c\right) + \tfrac{N-1}{N}b\sum_{a=0}^{M}\tbinom{M}{a}\lambda^a(1-\lambda)^{m-a}\min\{a, M-1\}\Big)$$
$$= (N-1)\Big((M-1)\left(\tfrac{b}{N}-c\right) + \tfrac{N-1}{N}b(M\lambda - \lambda^M)\Big). \quad (29)$$

This expression is zero at the solution to $\frac{M\lambda-\lambda^M}{M-1} = \lambda^*$. 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 the solution $\hat{\lambda}$ to Eq. 29 satisfies $\hat{\lambda} < \lambda^*$. Since the contract's value strictly increases in λ, empty promises with a single supervisor dominates promise keeping on a neighborhood of λ^*. By single crossing, this is also the case on $[0, \lambda^*]$.

Finally, 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

$$2(M-1)\left(\tfrac{N-1}{N}b\lambda + \tfrac{b}{N} - c\right) + (N-2)\sum_{a=0}^{M}\lambda^a(1-\lambda)^{M-a}\begin{pmatrix}\min\{a,p^*\}(b-c)\\ +\ \frac{\frac{b}{N}-c}{M-p^*}\binom{M-\min\{a,p^*\}}{2}\end{pmatrix} \tag{30}$$

By Lemma 3 this is a single-crossing family and the optimal p^* increases with λ. Note that the binomial term is quadratic, and $\sum_{a=0}^{M}\lambda^a(1-\lambda)^{M-a}a^2 = M\lambda(1-\lambda)+(M\lambda)^2$. When $p^* = M-1$, the binomial in the second term of Eq. 30 may be replaced with $\binom{M-a}{2}$ and Eq. 30 reduces to

$$2(M-1)\left(\tfrac{N-1}{N}b\lambda + \tfrac{b}{N} - c\right) + (N-2)\left((M\lambda-\lambda^M)(b-c) + \left(\tfrac{b}{N}-c\right)\frac{M(M-1)(1-\lambda)^2}{2}\right). \tag{31}$$

Simplifying terms and factoring, Eq. 31 dominates promise-keeping when

$$\begin{aligned}
&2N(M\lambda-\lambda^M)(b-c) + (b-cN)\left(M(M-1)(1-\lambda)^2\right) > 2M\left((N-1)b\lambda + b - cN\right)\\
\iff\quad & 2N\lambda^M(c-b) - 2M\lambda\left(cN + (N-(N-1))b\right) > \left(2M - M(M-1)(1-\lambda)^2\right)(b-cN)\\
\iff\ & 2(b-c)\lambda^M N < (1-\lambda)(M-\lambda(M-1)-3)M(b-cN).
\end{aligned} \tag{32}$$

The last LHS in Eq. 32 is positive, so the RHS must be positive too, for Eq. 32 to hold. So $b < cN$ means $M - \lambda(M-1) - 3 < 0$. For $\lambda^* = \frac{M-2}{M-1}$, Eq. 32 is

$$2(b-c)\left(\tfrac{M-2}{M-1}\right)^M N < \tfrac{M}{M-1}(cN-b). \tag{33}$$

The RHS of Eq. 33 is bounded below by $cN-b$. Consider $z_M = \left(\frac{M-2}{M-1}\right)^M$. 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 $z_M \to \frac{1}{e}$ from below. The LHS of Eq. 33 is then bounded above by $\frac{2(b-c)N}{e}$, and a sufficient condition for Eq. 33 is $\frac{2(b-c)N}{e} < cN - b$, which rearranges to $\frac{c}{b} > \frac{2}{e+2} + \frac{1}{N}\frac{e}{e+2}$. $\square$

Appendix B Optimal convex contracts (for online publication only)

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, the sanction depends on the number of failures f of inspection, where $f \in \{0, 1, \ldots, F\}$.

Within this class, contracts which deliver increasingly large sanctions 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 sanctions are imposed by third parties who are more inclined to exact sanctions 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 the sanction linearly thereafter.

Theorem 4. *For any M, cutoff strategies with a kinked linear sanctioning scheme 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 $\phi : \mathcal{Z} \to \mathbb{R}$, where $\mathcal{Z} \subseteq \mathbb{R}$, is *double crossing* if there is a (possibly empty) convex set $A \subset \mathbb{R}$ such that $A \cap \mathcal{Z} = \{z \in \mathcal{Z} : \phi(z) < 0\}$.

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 $z \in \mathcal{Z} = \{0, 1, \ldots, Z\}$.*[21] *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 $\sum_{r=0}^{R} (mr + b) q_z(r) = \frac{m}{k} z + \hat{b} + c$ for any real m and b. Hence, for any m and b,

$$\sum_{r=0}^{R} (mr+b)\big(q_{z+1}(r) - 2q_z(r) + q_{z-1}(r)\big) = \frac{m}{k}\big(z+1-2z+z-1\big) = 0, \tag{34}$$

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

for all $z = 1, \ldots, Z-1$. Therefore, for any m and b, the second difference of $\Psi(z)$ is

$$\begin{aligned}\Psi(z+1) - 2\Psi(z) + \Psi(z-1) &= \sum\nolimits_{r=0}^{R} \psi(r)\big(q_{z+1}(r) - 2q_z(r) + q_{z-1}(r)\big) \\ &= \sum\nolimits_{r=0}^{R} \big(\psi(r) - mr - b\big)\big(q_{z+1}(r) - 2q_z(r) + q_{z-1}(r)\big). \quad (35)\end{aligned}$$

By assumption, $q_{z+1}(r) - 2q_z(r) + q_{z-1}(r)$, as a function of r, is double crossing. Furthermore, since ψ 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. 35 we may conclude $\Psi(z)$ is concave. □

The next lemma says that expected sanctioning scheme will be decreasing convex in the number of tasks completed. Recall the definition of $g(f, a)$ from Eq. 11.

Lemma 5. *If ν is decreasing convex, then $h_\nu \equiv \sum_{f=0}^{F} \nu(f) g(f, \cdot)$ 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 $h_\nu(A)$ as follows:

$$\begin{aligned} h_\nu(A) &= \sum\nolimits_{f=0}^{F} g(f, a)\nu(f) \\ &= \sum\nolimits_{f=0}^{F} \Big(\sum\nolimits_{k=0}^{F} \frac{\binom{p-a}{k}\binom{a}{F-k}}{\binom{p}{F}} \binom{k}{f} \gamma^f (1-\gamma)^{k-f} \Big) \nu(f) \qquad (36) \\ &= \sum\nolimits_{k=0}^{F} \frac{\binom{p-a}{k}\binom{a}{F-k}}{\binom{p}{F}} \Big(\sum\nolimits_{f=0}^{F} \binom{k}{f} \gamma^f (1-\gamma)^{k-f} \nu(f) \Big). \end{aligned}$$

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 γk, 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

$$\binom{k}{f} \gamma^f (1-\gamma)^{k-f} \Big(\tfrac{(k+1)(1-\gamma)}{k+1-f} - 2 + \tfrac{k-f}{k(1-\gamma)} \Big). \qquad (37)$$

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

$$\frac{\binom{p-a}{k}\binom{a}{F-k}}{\binom{p}{F}}\left(\tfrac{p-a-k}{p-a}\cdot\tfrac{a+1}{a+1-F+k}-2+\tfrac{p-a+1}{p-a+1-k}\cdot\tfrac{a-F+k}{a}\right). \tag{38}$$

It can be shown that the term in parentheses has either no real roots or exactly two real roots.[22] 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 4. Fix any p, F, λ. Suppose 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 (or FOSD in a) implies the expected sanction decreases in the number of completed tasks: $h(a) > h(a-1)$ for all a. By contradiction, suppose 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 any a with $s(a) = a$, and every $a' < a$, the downward constraint $h(a) - h(a') = \sum_{k=a'}^{a-1} h(k+1) - h(k) \geq (a-a')(c-b)$ is slack. Some constraint must bind at the optimum, else the strategy is implementable for free, so the downward constraint for p^* versus p^*-1 must bind. 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.

Suppose we look for the optimal convex contract with p promises, F monitoring slots, and cutoff strategy s with cutoff p^*. By the above, the only binding incentive constraint is the downward constraint for completing p^* promises. Since $v(0) = 0$, convexity implies monotonicity. The constraint $v(0) \geq 0$ does not bind,[23] so the cost minimization problem in primal form is

$$\begin{gathered}
\max_{(-v)\geq\vec{0}} \sum_{f=0}^{F}\left(-(-v(f))\sum_{a=0}^{p} -g(f,a)t_s(a)\right) \text{ subject to} \\
\sum_{f=0}^{F}(-v(f))[g(f,p^*) - g(f,p^*-1)] \leq -(c-b), \\
2(-v(f)) - (-v(f+1)) - (-v(f-1)) \leq 0 \text{ for all } f = 1,\ldots,F-1,
\end{gathered} \tag{39}$$

[22] 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.

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

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 incentive compatibility constraint, z_f the 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,

$$A = \begin{pmatrix} g(0,p^*) - g(0,p^*-1) & -1 & & \\ \vdots & 2 & \ddots & \\ \vdots & -1 & \ddots & -1 \\ \vdots & & \ddots & 2 \\ g(F,p^*) - g(F,p^*-1) & & & -1 \end{pmatrix}. \tag{40}$$

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

$$\min_{r \geq \vec{0}} \ (b-c)x \quad \text{s.t. } (Ar)_f \geq -\sum_{a=0}^{p} g(f,a) t_s(a) \text{ for all } f = 0, 1, \ldots, F, \tag{41}$$

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

$$(Ar)_f = x[g(f,p^*) - g(f,p^*-1)] - z_{f-1} + 2z_f - z_{f+1}, \tag{42}$$

where we define $z_0 \equiv 0$, $z_F \equiv 0$, and $z_{F+1} \equiv 0$. Let $\hat{f}$ be the smallest f with $v(f) < 0$. It must be that $v(f) < 0$ for all $f \geq \hat{f}$, so by duality, $(A \cdot r)_f \geq -\sum_{a=0}^{p} g(f,a) t_s(a)$ binds for all $f \geq \hat{f}$. Hence

$$x = \frac{\sum_{a=0}^{p} g(f,a) t_s(a) - z_{f-1} + 2z_f - z_{f+1}}{g(f,p^*-1) - g(f,p^*)} \text{ for all } f = \hat{f}, \ldots, F. \tag{43}$$

In particular, this means that if $z_{F-1} = 0$ (which is implied when $\hat{f} = F$) the optimal contract (which would have expected sanction $-x(c-b)$) has the same value as that derived in Lemma 3), completing the claim. In the remainder we assume $z_{F-1} > 0$.

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: $-z_{F-1} + (2z_{F-1} - z_{F-2}) + (-z_{F-3} + 2z_{F-2} - z_{F-1}) = z_{F-2} - z_{F-3}$. Iterating, the sum of the z-terms in $(A \cdot r)_f$ from any $\tilde{f} \geq \hat{f}$

to F is $z_{\tilde{f}} - z_{\tilde{f}-1}$. Summing the equalities in Eq. 43 thus yields a recursive system for $z_{\tilde{f}}, \tilde{f} = \hat{f}, \ldots, F$:

$$z_{\tilde{f}} = z_{\tilde{f}-1} - \sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a) + x \sum_{f=\tilde{f}}^{F} \big(g(f,p^*-1) - g(f,p^*)\big). \tag{44}$$

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

$$z_{f'} = -\sum_{\tilde{f}=\hat{f}}^{f'} \sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a) + x \sum_{\tilde{f}=\hat{f}}^{f'} \sum_{f=\tilde{f}}^{F} \big(g(f,p^*-1) - g(f,p^*)\big). \tag{45}$$

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 $\hat{f}$:

$$x = \frac{\sum_{\tilde{f}=\hat{f}}^{F} \sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a)}{\sum_{\tilde{f}=\hat{f}}^{F} \sum_{f=\tilde{f}}^{F} \big(g(f,p^*-1) - g(f,p^*)\big)}. \tag{46}$$

The expectation of a random variable X on $\{0, \ldots, n\}$, is $\sum_{j=1}^{n} j \Pr(X=j)$, which also equals $\sum_{j=1}^{n} \Pr(X \geq j)$. 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. 46 equals

$$\begin{aligned}\sum_{\tilde{f}=\hat{f}}^{F} \sum_{f=\tilde{f}}^{F} \sum_{a=0}^{p} g(f,a)t_s(a) &= \sum_{\tilde{f}=\hat{f}}^{F} \Pr(f \geq \tilde{f}) = \sum_{\tilde{f}=\hat{f}}^{F} (\tilde{f} - \hat{f} + 1) \Pr(f = \tilde{f}) \\ &= \sum_{\tilde{f}=1}^{F} (\tilde{f} - \hat{f} + 1)_+ \Pr(f = \tilde{f}) = \mathbb{E}\big[(f - \hat{f} + 1)_+\big] \equiv \mathbb{E}[\phi(\hat{f})],\end{aligned} \tag{47}$$

where $(y)_+ \equiv \max\{y, 0\}$ and ϕ is the random function $\phi(\hat{f}) \equiv (f - \hat{f} + 1)_+$. In words, $\phi(\hat{f})$ is the number of discovered unfulfilled promises that exceed the threshold for sanctions $(\hat{f})$. The denominator of Eq. 46 can be rewritten similarly, yielding

$$x = \frac{\mathbb{E}[\phi(\hat{f})]}{\mathbb{E}[\,\phi(\hat{f}) \mid a = p^* - 1\,] - \mathbb{E}[\,\phi(\hat{f}) \mid a = p^*\,]}. \tag{48}$$

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

$$v(f) = -\frac{(c-b)(f - \hat{f} + 1)_+}{\mathbb{E}[\,\phi(\hat{f}) \mid a = p^* - 1\,] - \mathbb{E}[\,\phi(\hat{f}) \mid a = p^*\,]} \text{ for all } f = 0, 1, \ldots, F. \tag{49}$$

□