Job Scope and Motivation under Informal Incentives

Charles Angelucci and Roi Orzach*

November 25, 2024

Abstract

We model the relationship between the number of tasks assigned to an employee and a firm's ability to motivate effort through informal performance-based bonuses. We show that the assignment of multiple tasks gives the firm a greater range of performance levels that can be rewarded. The firm takes advantage of this by designing equally motivating, flatter, and hence more credible incentives.

Keywords: Relational Contract, Managerial Incentives, Relationship Scope

JEL classifications: C73, L14, L2, M52

1 Introduction

Motivating employees is a critical aspect of successful management and is often achieved through informal incentives such as discretionary performance-based bonuses (Granovetter, 1985; Gibbons, 2005). A common feature of these bonuses is the use of caps, where rewards increase with performance until reaching a maximum. For example, Murphy (2000) analyzes survey data on executive compensation packages and finds that most companies in the survey that utilize discretionary components also implement caps on bonuses. This pattern extends beyond executive compensation. At Google, junior engineers' annual bonuses are largely discretionary and capped at 20% of base salary (e.g., Francis and Mickle, 2022). Leading economic consulting firms, including Bain and McKinsey, similarly impose caps on subjective compensation (e.g., Schmitt, 2012). In this paper, we offer a novel theoretical explanation for informal bonus schemes that increase with performance up to a cap, emphasizing the role of multi-tasking and incentive design.

We consider a model of relational contracting between a principal and an agent (MacLeod and Malcomson, 1989; Baker et al., 1994, 2002; Levin, 2003). We begin the analysis by supposing that

*Angelucci: Sloan School of Management, MIT Orzach: Department of Economics, MIT.
We thank Alessandro Bonatti, Glenn Ellison, Jin Li, Dana Foarta, Robert Gibbons, Scott Duke Kominers, Juan Ortner, Kramer Quist, Michael Whinston, and Giorgio Zanarone.

the agent privately chooses whether to exert low or high effort on two identical tasks, each of which has a non-contractible stochastic outcome corresponding to either zero or one unit of output. In this scenario, we characterize the optimal self-enforcing bonus schemes that motivate high effort on both tasks, identifying those that can be implemented over the widest range of discount factors.[1] We show that optimal bonus schemes must take one of three forms: (i) A linear piece-rate scheme where the bonus equals a fixed commission rate multiplied by total output across both tasks, treating the tasks independently; (ii) A quota-based scheme that pays a fixed bonus for achieving at least one unit of output across the two tasks; or (iii) A quota-based scheme that pays a fixed bonus for achieving two units of output across the two tasks.

To understand how the presence of multiple tasks enhances the principal's ability to design credible incentives through the use of flatter bonus schemes, we compare the two-task case to a single-task benchmark. When the piece-rate bonus scheme is optimal, the presence of two tasks does not impact the principal's ability to motivate high effort. This equivalence arises because the principal's present value from defecting on the highest bonus she has promised is twice as large with two tasks compared to a single task, but so is the discounted future value of her relationship with the agent. In contrast, when quota-based bonus schemes are optimal, the presence of two tasks expands the range of discount factors over which the principal can credibly commit to informal incentives that motivate high effort. While the discounted future value of the principal's relationship with the agent is still twice as high with two tasks, the principal's present value from defecting on the promised bonus increases by less than a factor of two. Intuitively, when the employee is responsible for multiple tasks, the principal can pool these tasks to construct a single, finer performance scale and promise bonuses relative to this scale. This pooling allows the principal to offer flatter—and thus more credible—incentives by specifying the same bonus across various performance levels.

This logic strengthens substantially as the number of tasks grows. Specifically, we establish a strong limiting result: As the number of tasks the agent works on grows large, the minimum discount factor needed to credibly motivate effort on all tasks converges to the level that would be needed in an ideal first-best scenario where the agent's effort choices are fully observable. We prove this result by constructing a *capped* piece-rate bonus scheme that achieves this convergence

[1] Our analysis allows for asymmetric effort across tasks. For brevity, we here discuss only results for bonus schemes that incentivize high effort on all tasks.

while preserving incentive compatibility. The effectiveness of this incentive scheme provides a novel theoretical foundation for common incentive schemes which, as mentioned earlier, feature both an "incentive zone" where compensation is strictly increasing in performance followed by a cap where compensation remains constant regardless of further performance (c.f., Figure 1 Murphy, 2000). This perspective complements the existing literature, which often focuses on single-task environments with formal contracting, showing that factors such as limited liability (e.g., Poblete and Spulber, 2012) or risk aversion (e.g., Arnaiz and Salas-Fumás, 2008) can lead to capped incentives.

Related Literature. We are closely related to Bernheim and Whinston (1990). In their framework, firms that operate in multiple markets concurrently can leverage the fact that sustaining collusion in some markets is easy to also sustain collusion in more challenging markets. While we similarly study "multilateral" self-enforcing relationships, we focus on identical tasks, deliberately abstracting from their key mechanism. Further, Levin (2002) shows that firms benefit from pooling multiple employees' incentives into a multilateral relational contract, enabling cross-subsidization of incentives that are not self-enforcing in isolation. In our model, a single employee works on multiple tasks, and can potentially shirk on any subset of those tasks. In contrast to Levin (2002), pooling tasks will not always be optimal because doing so may lower the agent's incentives to exert high effort on all tasks simultaneously.

We are also closely related to Argyres et al. (2020) and Li and Powell (2020). Argyres et al. (2020) focus on buyer-supplier relationships and find that sourcing multiple inputs from a single supplier can enhance relational contracting by forcing the supplier and the buyer to internalize a larger fraction of the negative consequences of eventual defections. This logic finds empirical support in Argyres et al. (2022), who analyze slot exchanges between major US airlines and their regional airline partners.[2] In our setting it is the principal's temptation to defect which is reduced by the presence of multiple tasks, and not the agent's.

Li and Powell (2020) show that multiple identical tasks can help sustain relational contracts when the principal is allowed to defect on large payments with an arbitrarily small probability.[3]

[2] On this topic see also Gil et al. (2022). On the breadth of buyer-supplier relationships and the role played by trust and repeated interactions, see Aral et al. (2018).

[3] In their analysis, Li and Powell (2020) provide a more comprehensive examination of repeated game settings with random fluctuations, including environments without transferable utility such as favor exchange and multi-market collusion. We discuss further differences between our analysis and Li and Powell (2020)'s in Footnote 11. For related

Instead, we show that, even under the standard solution concept of Perfect Public Equilibrium, sufficiently many identical tasks effectively eliminate moral hazard because they enable the principal to avoid promising large payments altogether.

In a team production setting, Schöttner (2008) studies optimal task-assignment across employees in the presence of both contractible and non-contractible performance measures. Schöttner (2008) shows that increasing scope can be beneficial because of the interplay between formal and informal incentives (c.f. Baker et al., 1994).

Finally, our paper is also related to the literature on the trade-offs between piece-rate bonuses and quota-based bonuses when employees work on multiple tasks (see e.g., Oyer, 1998, 2000; Dai and Jerath, 2013; Kishore et al., 2013; Schöttner, 2017; Kräkel and Schöttner, 2016). Unlike in our setting, this literature typically considers the case in which outputs are verifiable and formal contracts are available. Notable exceptions are Kvaløy and Olsen (2019, 2023), which generalize the insights from Levin (2003) that offering a maximum bonus when output is above a threshold is optimal to two new settings (i) multiple agents with correlated outputs and (ii) single agent working on multiple tasks with heterogeneous output precision.

The paper is organized as follows. The model is defined in Section 2, and the analysis is presented in Section 3. Section 4 concludes by discussing the managerial implications of our findings as well as possible other applications.

2 Model

We consider two risk-neutral players, a principal (she) and an agent (he), who have the opportunity to interact at dates $t = 0, 1, 2, \ldots$. Both players' per-period outside options are equal to zero. They face no wealth constraints and apply the interest rate r to value future payoffs. At the beginning of each date t, the principal proposes a compensation package to the agent. Compensation consists of a fixed salary w_t and a bonus b_t, to be described momentarily. The agent then chooses whether to accept ($d_t = 1$) or reject ($d_t = 0$) the principal's offer.

If $d_t = 1$, the agent works on n identical tasks. For each task $k = 1, \ldots, n$, the agent unobservably chooses an effort level $e_{k,t} \in \{0, 1\}$ and incurs a cost $ce_{k,t}$. We denote by $y_{k,t} \in \{0, 1\}$ task

work, see also Sekiguchi (2015).

k's output in period t. Denote by p_h (respectively, p_l) the probability that $y_{k,t} = 1$ conditional on high (respectively, low) effort. We assume that $p_h - p_l > c$, which is the necessary and sufficient condition for high effort to be socially optimal.

At the end of each period t, for each task k, the principal observes the output $y_{k,t}$ but not the effort level $e_{k,t}$.[4] The bonus b_t thus depends on the tasks' outputs: $b_t : \{0, 1\}^n \rightarrow \mathbf{R}$.[5]

At the beginning of date t, the principal is obligated to pay the fixed salary w_t. However, the players then choose whether to adjust this contractual payment with the bonus b_t. This decision belongs to the principal if $b_t > 0$ and to the agent if $b_t < 0$, and is made after the output has been produced.

Let $\mathbf{h}_t = (w_0, d_0, y_{1,0}, \ldots, y_{n,0}, b_0, \ldots, w_{t-1}, d_{t-1}, y_{1,t-1}, \ldots, y_{n,t-1}, b_{t-1})$ denote the history up to date t, and $\mathcal{H}_t$ the set of possible date t histories. Then for each date t and every history $\mathbf{h}_t \in \mathcal{H}_t$, a relational contract describes: (i) the compensation the principal should offer; (ii) whether the agent should accept or reject the offer; in the event of acceptance, (iii) the effort level to be chosen by the agent on each task k, (iv) and the bonus b_t to be paid as function of the output produced. Such a relational contract is self-enforcing if it describes a Perfect Public Equilibrium of the repeated game.

We focus on self-enforcing relational contracts that maximize the players' joint surplus. We further restrict our attention to stationary relational contracts. Theorem 2 in Levin (2003) extends to our setting, and we thus know that stationary relational contracts will maximize the players' joint surplus.[6] If either a player reneges on a payment, the principal fails to offer the expected compensation, or the agent unexpectedly refuses the compensation plan, we assume that the parties respond by permanently breaking off their relationship. This is without loss of generality as (i) it occurs only out-of-equilibrium and (ii) it represents the worst equilibrium outcome (Abreu, 1988). To ease exposition, we allow the principal to retain all the surplus generated by the relationship. Finally, when multiple stationary relational contracts maximize the players' joint surplus, we characterize the optimal relational contract that is self-enforcing for the highest value of the discount

[4]In Fong and Li (2016), it is shown that committing to observing only a history-dependent garbling of output (e.g., via an intermediary) decreases the maximum relational bonus needed to motivate effort.

[5]Since both parties are risk-neutral, deterministic mechanisms are without loss of optimality.

[6]Note that the presence of multiple tasks generates a mapping between effort and output which, though natural, does not satisfy the regularity conditions found in Levin (2003). For instance, while in his setting it is without loss of generality to focus on quota-based bonus schemes, we show in Proposition 1 that it can be sub-optimal in ours.

rate r (i.e., the 'critical value' of the discount rate). We refer to this relational contract as the *optimal relational contract.*

Discussion. For conciseness, we make the stark assumption that the principal observes only non-contractible subjective performance measures. In practice, firms rely on a combination of verifiable and non-verifiable performance measures when creating incentives (c.f., Baker et al., 1994; Murphy, 2000).[7]

Our terminology of piece-rate and quota-based schemes traditionally evokes manufacturing settings, where output is verifiable and formal contracting is available. We adopt this familiar language to convey our economic intuitions, while emphasizing that our model applies primarily to settings where performance is subjective or formal contracts are infeasible. Such settings include non-routine tasks with hard-to-verify quality measures or sectors in developing economies, such as agriculture, where contract enforcement is weak.

Further, we treat the number of tasks n as exogenous rather than deriving it from primitives. This approach allows us to focus squarely on how relationship scope affects optimal informal incentives, while holding other factors constant. For the same reason, we also abstract from technological complementarities across tasks that could arise in practice, such as effort costs on one task depending on effort choices on others, particularly when n is large. We also note that we interpret the parameter n as distinct tasks, but that an alternative interpretation of n is that it corresponds to units of output from the same task produced during the relevant time period used to determine the agent's pay.

Finally, the agent makes all effort decisions before observing any output, capturing situations with delayed learning of outputs. This timing assumption is important, as the sequential observation of outputs would diminish the effectiveness of non-linear incentives like the quota scheme in our two-task model. This timing structure arises, for example, in agricultural production where farmers make multiple planting and cultivation decisions before observing harvest yields, in R&D where scientists run several experiments in parallel before learning their results, or in software development

[7]For instance, most companies surveyed in Murphy (2000) employ a mix of objective and subjective criteria to determine executive compensation. In practice, even if a company uses objective criteria (e.g., year-on-year growth), how much weight such criteria are given when determining final pay may still be subjective (c.f., Höppe and Moers, 2011). Also, we note that, in our setting, caps would not be optimal if fully contractible performance measures were available, suggesting that caps are closely linked to the reliance on subjective, discretionary performance measures.

where programmers write multiple features before receiving user feedback.

3 Analysis

We divide the analysis into two subsections. The first subsection examines how the principal structures informal bonuses when there are two tasks. The second subsection extends the analysis to scenarios involving a large number of tasks.

3.1 Assigning the Agent Two Tasks

For ease of exposition, we drop the t subscript from all variables because of our focus on stationary relational contracts. Consider first the benchmark case with a single task. Suppose further that the principal motivates the agent to exert high effort (as low effort can be motivated absent bonuses). Without loss of generality, let $b(y) = by$. For the relational contract to be self-enforcing, the following inequalities must hold:

$$b \leq \frac{1}{r}(p_h - c) \tag{1a}$$

$$p_h b - c \geq p_l b. \tag{1b}$$

Inequality (1a) ensures that the principal is better off paying the bonus promised in case the agent produces 1 unit of output rather than terminating the relationship and incurring a loss equal to the discounted future value of her relationship with the agent, represented by $\frac{1}{r}(p_h - c)$. Additionally, inequality (1b) ensures the agent is better off exerting high effort rather than low effort.[8] Clearly, setting $b = \frac{c}{p_h - p_l}$ is optimal and inducing effort on this single task is possible under a relational contract if and only if:[9]

$$r \leq r_{\text{ST}} \equiv \frac{p_h - p_l}{c}(p_h - c). \tag{2}$$

[8] In writing inequalities (1a)-(1b) we are anticipating that, in the principal's preferred equilibrium, the fixed wage w is set so as to bind the agent's participation constraint and transfer the entirety of the relationship's surplus to the principal.

[9] As in Levin (2003), considering relational contracts in which the principal makes all the payments is without loss of generality. We thus additionally assume that $\text{supp}(b) \geq 0$ throughout the analysis.

Now, suppose that there are two identical tasks, denoted 1 and 2. If the principal motivates high effort on only one task, say task 1, she finds it optimal to offer a bonus scheme given by $b(y_1, y_2) = \frac{c}{p_h - p_l} y_1$. In other words, she promises the same bonus as in the single-task benchmark, conditional on high output from the first task, while the output of the second task does not affect the reward. As we show below, this relational contract constitutes an equilibrium if and only if r is lower than or equal to some threshold $\tilde{r}_{\text{TT}}$, where $\tilde{r}_{\text{TT}} \leq r_{\text{ST}}$.

Now, consider the case where the agent is asked to exert high effort on both tasks. As shown in the Appendix, since the tasks are symmetric, there is no loss of generality in focusing on contracts that treat both tasks symmetrically. This implies that the bonus depends only on the total output. Given that outputs y_1 and y_2 are binary variables, we can express the bonus scheme $b(y_1, y_2)$ as:

$$b\left(y_1, y_2\right) = \sum_{i=0}^{2} b_i \mathbf{1}_{y_1 + y_2 = i}, \tag{3}$$

noting that setting $b_0 = 0$ is without loss of generality and henceforth assumed.

The relational contract that induces the agent to exert high effort on both tasks must satisfy two sets of constraints. First, the bonuses promised must be credible:

$$\max\{b_1, b_2\} \leq \frac{1}{r} 2(p_h - c). \tag{4}$$

Inequality (4) ensures that the principal is better off honoring any bonus promised to the agent. Because the optimal relational contract is stationary and such that the fixed wage w binds the agent's participation constraint, the principal's continuation value–i.e, what she stands to lose if the relationship were to end–is equal to the joint surplus $\frac{1}{r} 2\left(p_h - c\right)$.

The second set of constraints that the bonus scheme $b(y_1, y_2)$ must satisfy ensures that the agent is better off exerting high effort on both tasks as compared to exerting high effort on only one task or on neither task:

$$b_2 p_h^2 + b_1 2 p_h (1 - p_h) - 2c \geq b_2 p_l p_h + b_1 \big(p_l (1 - p_h) + p_h (1 - p_l)\big) - c \tag{5a}$$

$$b_2 p_h^2 + b_1 2 p_h (1 - p_h) - 2c \geq b_2 p_l^2 + b_1 2 p_l (1 - p_l) \tag{5b}$$

In the appendix, we show that the optimal values of b_1 and b_2 result in $b(y_1, y_2)$ resembling either a quota-based bonus scheme (with the quota set either at 1 or 2) or a piece-rate bonus scheme. The intuition for this result lies in the fact that at least one of the constraints, Inequality (5a) or Inequality (5b), must bind. When both constraints bind, the agent has equal incentives to exert effort on both tasks, resulting in a piece-rate bonus scheme. If only Inequality (5b) binds, the agent has a stronger incentive to deviate by exerting low effort on both tasks. This implies that the optimal bonus scheme resembles a quota-based scheme with a quota of 2, as under this scheme, the agent would indeed find it optimal to deviate by exerting low effort on both tasks rather than on only one. Conversely, if only Inequality (5a) binds, the optimal bonus scheme takes the form of a quota-based scheme with a quota of 1, as the agent is more strongly incentivized to deviate by exerting low effort on one task while maintaining high effort on the other.

We begin by comparing the maximum payments under each of the three incentive schemes. In the piece-rate bonus scheme, the maximum payment is $\frac{2c}{p_h - p_l}$, paid when the agent produces two units of output. For the quota-based scheme, the maximum payment is $\frac{c}{(1-p_h)(p_h-p_l)}$ when the quota is set to 1, and $\frac{2c}{p_h^2 - p_l^2}$ when the quota is set to 2. The bonus scheme with the lowest maximum payment will be implementable for the largest range of discount factors. Figure 1 identifies which of the three schemes has the lowest maximum payment. Since the maximum payment under each of the three incentive schemes is linear in c, the figure varies only p_h and p_l.

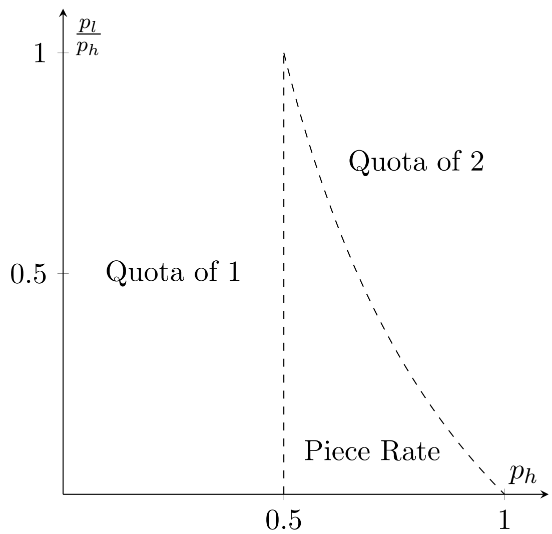

Figure 1: Bonus scheme with lowest maximum payment

The figure shows the bonus scheme with the lowest maximum payment for all combinations of p_h and p_l such that $p_h > p_l$, comparing the piece-rate scheme, the quota-based scheme with a quota of 1, and the quota-based scheme with a quota of 2.

The first key insight from the figure is that when tasks are relatively difficult (i.e., $p_h \leq \frac{1}{2}$), the probability of the agent producing two units of output is low. Consequently, the principal finds it optimal to implement a quota-based scheme with a quota of 1 to minimize the maximum payment required. In contrast, when $p_h > \frac{1}{2}$, this approach becomes strictly suboptimal. At higher probabilities, the agent is sufficiently likely to produce one unit of output by exerting effort on a single task, meaning that motivating effort on both tasks would require a bonus exceeding that of the quota-based scheme with a quota of 1. For $p_h > \frac{1}{2}$, the principal chooses between a piece-rate scheme and a quota-based scheme with a quota of 2. In the quota-based scheme with a quota of 2, the agent is most tempted to exert low effort on both tasks. Increasing the bonus for producing two units of output helps relax the agent's incentive constraint at a rate of $p_h^2 - p_l^2$. In the piece-rate scheme, tasks are independent, and increasing the piece-rate payment relaxes the incentive constraints (one for each task) at a rate of $p_h - p_l$. As a result, the quota-based scheme with a quota of 2 is more effective at providing incentives (and thus requires a strictly lower maximum payment) when $p_h + p_l > 1$, as illustrated in Figure 1.

The following proposition formalizes these results, specifying the bonus scheme and effort levels associated with the optimal relational contract.

Proposition 1. *There exists an equilibrium in which the agent exerts high effort on both tasks in every period if and only if:*

$$r \leq r_{TT} := r_{ST} \cdot \max\{1, 2(1-p_h), p_h + p_l\}. \tag{6}$$

If Inequality (6) *holds, it is optimal to motivate high effort on both tasks, and the bonus scheme in the optimal relational contract is given by:*

$$\begin{aligned} b_2 = b_1 = \frac{c}{(1-p_h)(p_h-p_l)} \quad & \text{if } p_h \leq \frac{1}{2}, \\ b_2 = 2b_1 = \frac{2c}{p_h - p_l} \quad & \text{if } p_h \geq \frac{1}{2}, p_h + p_l \leq 1, \\ b_2 = \frac{2c}{p_h^2 - p_l^2} \text{ and } b_1 = 0 \quad & \text{if } p_h \geq \frac{1}{2}, p_h + p_l \geq 1. \end{aligned} \tag{7}$$

Finally, an equilibrium exists in which the agent exerts high effort on one task and low effort on the other in every period if and only if:

$$r \leq \tilde{r}_{TT} := r_{ST} + \frac{p_l(p_h - p_l)}{c}. \tag{8}$$

If $r \in (r_{TT}, \tilde{r}_{TT}]$, it is optimal to motivate high effort on only one task. The principal is indifferent about which task is selected and offers a bonus of $\frac{c}{p_h - p_l}$ conditional on achieving high output in the task where high effort is incentivized.

Since exerting high effort on each task is socially efficient, the optimal relational contract motivates high effort on both tasks whenever the necessary payments are credible (i.e., $r \leq r_{\mathrm{TT}}$). Compared to the single-task benchmark, the range of discount rates over which motivating high effort on all tasks is feasible is weakly larger (i.e., $r_{TT} \geq r_{ST}$), and strictly so whenever the optimal bonus scheme is quota-based. If the optimal bonus scheme with two tasks is piece-rate, motivating high effort on both tasks is feasible over the same range of discount rates as motivating high effort on a single task. This is because the maximum payment in the two-task case is twice as large as in the single-task case, but so is the continuation value of the relationship. In contrast, if the optimal bonus scheme with two tasks is quota-based, the continuation value with two tasks remains twice as large as with one task, but the maximum payment is less than twice that of the single-task case.

Further, if motivating high effort on both tasks is not feasible, the principal may nevertheless be able to induce high effort on a single task and will optimally choose this approach whenever she can. In particular, motivating high effort on a single task requires a payment that is credible over a weakly wider range of discount rates compared to the maximum payment needed to motivate high effort on both tasks (i.e., $r_{TT} \leq \tilde{r}_{TT}$). This occurs because the payment scheme only needs to incentivize effort on one task, while the principal's continuation value is still determined by both tasks. Even if the agent exerts low effort on the second task, its contribution increases the continuation value of the relationship as long as $p_l > 0$. This, in turn, enhances the principal's credibility in committing to promised payments for the task where high effort is incentivized. This reasoning implies that having two tasks, rather than just one, allows the principal to motivate high effort on at least one task over a wider range of discount rates when $p_l > 0$.

We have identified two distinct reasons why an increase in scope, measured by the number of tasks, may be beneficial. First, when effort is motivated on each task, greater scope allows for more flexible pay schemes (quota-based schemes), enabling effort to be incentivized for a wider range of discount factors. Second, an increase in scope allows for asymmetric equilibrium effort levels, which can enhance the principal's ability to credibly promise payments for the tasks where high effort is induced. These insights are formalized in the following proposition.

Proposition 2. *There exists an optimal relational contract that induces the agent to exert high effort on at least one task over a weakly larger range of discount rates when two tasks are present compared to a single task (i.e.,* $\max\{r_{TT}, \tilde{r}_{tt}\} \geq r_{ST}$*), and over a strictly larger range of discount rates when either* $p_h < \frac{1}{2}$ *or* $p_l > 0$*.*

Proposition 2 supports the idea that a broader relationship scope can make it easier for an employer to motivate effort. When $p_h < \frac{1}{2}$, the principal optimally offers a quota-based scheme with a quota of 1, making it strictly easier to motivate high effort on two tasks compared to motivating high effort on a single task. Additionally, if $p_l > 0$, the presence of a second task allows the principal to induce asymmetric effort levels across tasks, making it easier to motivate high effort on just one task. Together, these results imply that the average profitability per task (weakly) increases when the number of tasks, n, increases from one to two, and increases strictly for certain parameter values.[10]

3.2 The Agent is Assigned a Large Number of Tasks

We showed that the presence of two identical tasks can help the principal to motivate effort because it enables her to pool these tasks and create a single but finer range of performance levels (i.e., total output), which, in turn, allows for flatter and hence more credible incentives. When the number of tasks n exceeds 2, finding a closed-form solution for the optimal relational contract becomes complex due to the large number of potential contracts and the presence of integer constraints.[11] Nevertheless, we prove an extension to Proposition 2 which shows that the presence of sufficiently many tasks necessarily helps the principal to motivate high effort on each task.

Intuitively, aggregate uncertainty about total output $\sum_{i=1}^{n} y_i$ decreases as n gets large. At the same time, the relational contract must specify credible payments for all possible realizations of the outputs, including for the vanishingly-small-probability event in which $\sum_{i=1}^{n} y_i = n$. We show that a *capped* piece-rate bonus scheme helps the principal credibly commit to a relational contract that motivates high effort on all tasks. If n grows large enough, the range of discount rates under which the principal can motivate high effort on all tasks can be made arbitrarily close to the corresponding range of discount rates if effort choices were directly observable.

Consider a capped piece-rate bonus scheme such that the principal pays the agent a fixed amount, $\frac{c}{p_h - p_l} + \epsilon$, where $\epsilon > 0$, for each unit of output produced, up to a maximum of $\kappa(n) \leq n$

[10]In contrast, if the employer had access to formal contracts, the average profitability per task would be independent of n, as treating tasks independently would be optimal.

[11]Li and Powell (2020) make progress towards a solution by showing that it is without loss of optimality for the principal to promise a bonus that takes up to three values. Instead, we show that a simple relational contract characterized by features commonly observed in practice—including a bonus scheme that strictly increases with performance up to a maximum cap beyond which additional performance no longer affects the bonus—achieves the same surplus as the optimal relational contract as $n \to \infty$. This is established by showing that the maximum reneging temptation under the simple relational contract approaches the maximum reneging temptation of the optimal relational contract in the absence of moral hazard.

units.[12] Specifically:

$$b(y_1, \ldots, y_n) = \begin{cases} (\frac{c}{p_h - p_l} + \epsilon) \sum y_i, & \text{if } \sum y_i \leq \kappa(n) \equiv np_h + \mu\sqrt{n} \\ (\frac{c}{p_h - p_l} + \epsilon)(np_h + \mu\sqrt{n}), & \text{otherwise.} \end{cases} \tag{9}$$

If the agent exerts high effort on all tasks, the expected total output produced is equal to np_h and, by the Central Limit Theorem, the probability that the agent produces more than $\kappa(n) = np_h + \mu\sqrt{n}$ approaches a constant that is a function of μ, p_h and p_l.

If the bonus scheme were completely linear and with a commission fee equal to $\frac{c}{p_h - p_l} + \epsilon$, the agent would treat each task independently and put in high effort on all tasks. By contrast, under a capped piece-rate scheme, an agent exerts high effort on a given task only if he thinks that it is sufficiently unlikely that he can achieve a total output of $\kappa(n)$ or more by exerting effort only on the remaining tasks. However, by setting the parameter μ appropriately, the principal can choose a large enough cap $\kappa(n)$ to ensure that exerting high effort on each individual task increases the agent's expected payment enough to cover the cost of effort c. Moreover, because of the cap $\mu\sqrt{n} + n(p_h - p_l) \in \mathcal{O}(n)$, the principal is able to design a capped piece-rate bonus scheme whose associated maximum bonus is on the order of $nc + n\epsilon$. This implies that the critical discount rate approaches $\frac{p_h - c}{c + \epsilon}$. Because ϵ can be made arbitrarily small, the principal's ability to motivate high effort is thus (almost) identical to her ability to motivate high effort if she could observe effort choices.

Proposition 3. *For any tuple $\{\epsilon, p_h, p_l, c\}$, there exists a threshold n^* and a constant μ such that, if $n > n^*$, the bonus scheme outlined in Equation (9) induces high effort on all tasks. Further, as $n \to \infty$, the critical discount rate required to induce high effort on all tasks approaches $\frac{p_h - c}{c}$.*

This result provides insight into the popularity of capped piece-rate contracts, which combine the effort incentives of linear contracts with the assurance that the principal will fulfill all bonus promises by limiting their potential temptation to renege. In other words, these contracts offer a way to balance the need for effort incentives with the risk of opportunistic behavior by the principal.

[12] In the version of the model with two tasks, the commission fee in the optimal piece-rate scheme, $\frac{c}{p_h - p_l}$, made the agent indifferent whether to exert high effort on a task. Here, because the agent can produce a total output equal to or higher than the cap $\kappa(n)$ with some probability, the commission needs to be higher by a positive epsilon. When taking n to infinity, we can show that, with an appropriately chosen cap $\kappa(n)$, we can take ϵ to zero.

When $r > \frac{p_h - c}{c}$, the principal is unable to motivate high effort on all tasks, even under the hypothetical scenario where effort levels are directly observable. Consequently, it is also impossible to achieve this outcome using the capped piece-rate contract described above when faced with a large number of tasks. However, if $p_l > 0$, the principal can motivate high effort on a share α of tasks. Specifically, up to integer constraints, and in the hypothetical scenario where effort levels are observable, the principal could motivate high effort on a share α of the tasks if and only if:

$$\alpha \cdot n \cdot c \leq \frac{1}{r} n\Big(\alpha(p_h - c) + (1-\alpha)p_l\Big) \iff \alpha \leq \frac{p_l}{cr + c - p_h + p_l}. \tag{10}$$

Thus, for any n, the principal cannot motivate high effort on a greater share of tasks than the bound established above when effort levels are unobservable. However, asymptotically, the share of tasks on which high effort can be motivated converges to this bound. To see why, note that, for any share α of the tasks, $\alpha \cdot n$ diverges to infinity, implying that the logic of Proposition 3 applies: The principal can design a capped piece-rate contract which effectively removes moral hazard on the tasks where high effort is motivated. This intuition is formalized below.

Proposition 4. *If $p_l > 0$, then, for any $r > 0$, there exists a capped piece-rate contract which induces high effort on $\alpha(n) \cdot n$ tasks, where $\alpha(n) \to \frac{p_l}{cr+c-p_h+p_l}$.*

This section shows how the qualitative features of the previous section with two tasks continue to hold for a large number of tasks. Specifically, as scope increases, the principal gains the ability to design more flexible and, consequently, more credible informal bonus schemes (Proposition 3). Additionally, as scope increases, the principal gains flexibility in motivating asymmetric effort levels across the tasks, which also allows for more credible informal bonus schemes (Proposition 4).

Finally, as mentioned in Section 1, our analysis provides a new rationale for the prevalence of capped bonus plans (for evidence on executive bonus plans see, for instance, Murphy, 2000; Chen et al., 2022). The emphasis is typically on how capped bonus plans impact employee incentives in terms of risk-taking, effort provision, or gaming behavior. However, this paper shifts the focus to the effects of capped bonus plans on managers, and specifically on their ability to make credible promises about discretionary bonuses.

4 Conclusion

Our main finding is that a broad scope of tasks and responsibilities given to employees helps maintain relational incentives. In particular, a manager uses this scope to design relational incentives such as quotas or capped piece-rate incentives which provide equally motivating and flatter incentives. These flatter incentives are more credible, enabling them to constitute an equilibrium for a broader range of discount rates.

Further, to keep things tractable, we focused on situations where there were either two tasks or a large number of tasks (i.e., n going to infinity). However, by using the simplex representation to solve for the optimal bonus, one can show that linear bonus schemes are not optimal when $n \geq 4$. This means that when there are at least four tasks, relationship scope always strictly expands the range of discount rates for which high effort can be motivated on all tasks under relational contracting.

Finally, although the focus of the analysis was on employment settings, our insights apply more generally to other contexts in which informal incentives are prevalent, utility is transferable, and the scope of activities varies, such as buyer-supplier relationships in manufacturing and in R&D or collusion between firms across multiple markets.

References

Abreu, Dilip, "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 1988, *56* (2), 383–396.

Aral, Sinan, Yannis Bakos, and Erik Brynjolfsson, "Information Technology, Repeated Contracts, and the Number of Suppliers," *Management Science*, 2018, *64* (2), 592–612.

Argyres, Nicholas, Janet Bercovitz, and Giorgio Zanarone, "The role of relationship scope in sustaining relational contracts in interfirm networks," *Strategic Management Journal*, 2020, *41* (2), 222–245.

_ , Ricard Gil, and Giorgio Zanarone, "Does Supply Concentration Encourage Cooperation? Evidence From Airlines," *Working Paper*, 2022.

Arnaiz, Oscar Gutierrez and Vicente Salas-Fumás, "Performance standards and optimal incentives," *Journal of Accounting and Economics*, 2008, *45* (1), 139–152.

Baker, George, Robert Gibbons, and Kevin J. Murphy, "Subjective Performance Measures in Optimal Incentive Contracts," *The Quarterly Journal of Economics*, 1994, *109* (4), 1125–1156.

_ , _ , and _ , "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 2002, *117* (1), 39–84.

Bernheim, B Douglas and Michael D Whinston, "Multimarket contact and collusive behavior," *The RAND Journal of Economics*, 1990, pp. 1–26.

Chen, Clara Xiaoling, Minjeong (MJ) Kim, Laura Yue Li, and Wei Zhu, "Accounting Performance Goals in CEO Compensation Contracts and Corporate Risk Taking," *Management Science*, 2022, *68* (8), 6039–6058.

Dai, Tinglong and Kinshuk Jerath, "Salesforce Compensation with Inventory Considerations," *Management Science*, 2013, *59* (11), 2490–2501.

Fong, Yuk-Fai and Jin Li, "Information revelation in relational contracts," *The Review of economic studies*, 2016, *84* (1), 277–299.

Francis, Theo and Tripp Mickle, "Tech Giants Turn to a Classic Recruitment Tool: Cash," *The Wall Street Journal*, 2022.

Gibbons, Robert, "Incentives between firms (and within)," *Management Science*, 2005, *51* (1), 2–17.

Gil, Ricard, Myongjin Kim, and Giorgio Zanarone, "Relationships under stress: Relational outsourcing in the US airline industry after the 2008 financial crisis," *Management Science*, 2022, *68* (2), 1256–1277.

Granovetter, Mark, "The Problem of Embeddedness," *American Journal of Sociology*, 1985, *91* (3), 481–510.

Höppe, Felix and Frank Moers, "The choice of different types of subjectivity in CEO annual bonus contracts," *The Accounting Review*, 2011, *86* (6), 2023–2046.

Kishore, Sunil, Raghunath Singh Rao, Om Narasimhan, and George John, "Bonuses versus Commissions: A Field Study," *Journal of Marketing Research*, 2013, *50* (3), 317–333.

Kräkel, Matthias and Anja Schöttner, "Optimal sales force compensation," *Journal of Economic Behavior & Organization*, 2016, *126*, 179–195.

Kvaløy, Ola and Trond E Olsen, "Relational contracts, multiple agents, and correlated outputs," *Management Science*, 2019, *65* (11), 5360–5370.

_ **and** _ , "Balanced scorecards: A relational contract approach," *Journal of Accounting Research*, 2023, *61* (2), 619–652.

Levin, Jonathan, "Multilateral contracting and the employment relationship," *The Quarterly Journal of Economics*, 2002, *117* (3), 1075–1103.

_ , "Relational Incentive Contracts," *American Economic Review*, June 2003, *93* (3), 835–857.

Li, Jin and Michael Powell, "Multilateral interactions improve cooperation under random fluctuations," *Games and Economic Behavior*, 2020, *119*, 358–382.

MacLeod, W. Bentley and James M. Malcomson, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 1989, *57* (2), 447–480.

Murphy, Kevin J, "Performance standards in incentive contracts," *Journal of Accounting and Economics*, 2000, *30* (3), 245–278.

Oyer, Paul, "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *The Quarterly Journal of Economics*, 1998, *113* (1), 149–185.

_ , "A Theory of Sales Quotas with Limited Liability and Rent Sharing," *Journal of Labor Economics*, 2000, *18* (3), 405–426.

Poblete, Joaquín and Daniel Spulber, "The form of incentive contracts: agency with moral hazard, risk neutrality, and limited liability," *The RAND Journal of Economics*, 2012, *43* (2), 215–234.

Schmitt, Jeff, "Consulting Pay: What MBAs Earned In 2023," *Poets and Quants*, 2012.

Schöttner, Anja, "Relational contracts, multitasking, and job design," *The Journal of Law, Economics, & Organization*, 2008, *24* (1), 138–162.

Schöttner, Anja, "Optimal Sales Force Compensation in Dynamic Settings: Commissions vs. Bonuses," *Management Science*, 2017, *63* (5), 1529–1544.

Sekiguchi, Tadashi, "Multimarket contact under demand fluctuations," *International Journal of Game Theory*, 2015, *44*, 1033–1048.

Appendix

Proof of Proposition 1. First note that it is without loss to consider a symmetric bonus. If $b(y_1, y_2)$ is not symmetric, then one can construct an alternative scheme,

$$\tilde{b}(y_1, y_2) = \frac{1}{2}b(y_1, y_2) + \frac{1}{2}b(y_2, y_1), \tag{11}$$

which is symmetric. Further, $\tilde{b}(y_1, y_2)$ has a (weakly) smaller maximum bonus than $b(y_1, y_2)$ and motivates high effort on both tasks whenever $b(y_1, y_2)$ does so.

Thus, let b_0, b_1, b_2 correspond to the bonuses where the subscript denotes the number of units of output produced as in Equation (3). By an identical proof to Levin (2003), it is without loss of generality to consider bonus schemes where all bonuses are weakly positive and $b_0 = 0$. Therefore, the optimal relational contract which motivates high effort on both tasks satisfies:

$$\min_{b_1, b_2} \quad \max\{b_1, b_2\}$$

$$\text{subject to: } b_2 p_h^2 + b_1 2p_h(1-p_h) - 2c \geq b_2 p_l^2 + b_1 2p_l(1-p_l) \tag{12}$$

$$b_2 p_h^2 + b_1 2p_h(1-p_h) - 2c \geq b_2 p_h p_l + b_1\big(p_l(1-p_h) + p_h(1-p_l)\big) - c \tag{13}$$

There are four cases to consider based on whether Equations (12) and (13) bind at the optimum.

Case 1: Suppose neither constraint binds. One could decrease b_1 and b_2 by ϵ and still satisfy the constraints, implying this never occurs at the optimum.

Case 2: Suppose both constraints bind. These constraints result in two linear equations (which are not co-linear) with two variables implying a unique solution to b_1, b_2. Further, the piece-rate contract is a contract where these solutions bind, implying the only way to satisfy both constraints is if $b_2 = 2b_1 = 2c/(p_h - p_l)$.

Case 3: Suppose only Equation (12) binds. If $b_1 \neq b_2$ and both are strictly positive, then one can increase $\min\{b_1, b_2\}$ and decrease $\max\{b_1, b_2\}$ to preserve Equation (12) and maintain slack on the second constraint, thus contradicting optimality. If instead $b_1 = b_2$, then to bind Equation (12) it must be that $b_1 = b_2 = 2c/(2p_h - p_h^2 - 2p_l + p_l^2)$. However, this solution never leads to the second equation being slack. Thus if Equation (12) is tight it must be that $b_1 \neq b_2$ and one of these values is equal to zero. One can show $b_2 = 0$ is never optimal, thus $b_1 = 0$. In this case $b_2 = 2c/(p_h^2 - p_l^2)$ and this value satisfies the second constraint.

Case 4: Suppose only Equation (13) is tight. Again, if $b_1 \neq b_2$ and both are strictly positive, then one can decrease the maximum of these two values and increase the minimum. If $b_1 = b_2$ and by assumption Equation (13) binds, then $b_1 = b_2 = c/\big((1-p_h)(p_h - p_l)\big)$, and these value satisfies Equation (12). If $b_1 \neq b_2$, then one of these values is equal to zero. Again, $b_2 = 0$ is never optimal. Hence, $b_1 = 0$ and b_2 binds the second inequality. However, one can show that the value which binds the second inequality never satisfies the first.

Thus, the maximum possible payments are $2c/(p_h - p_l)$, $2c/(p_h^2 - p_l^2)$, and $c/\big((1-p_h)(p_h - p_l)\big)$. Algebraic manipulations complete the proof when high effort is incentivized on both tasks.

Finally, if the agent exerts high effort on exactly one task, we can assume without loss of generality that this task is task one. Moreover, in such a case, making the payment contingent on the output of task two is suboptimal. Therefore, the bonus following high output is $\frac{c}{p_h - p_l}$. Given this maximum bonus, algebraic manipulation provides the characterization of $\tilde{r}_{tt}$. □

Proof of Proposition 2. Proposition 2 follows immediately from Proposition 1. □

Proof of Proposition 3. We need to show that for any tuple $\{\epsilon, p_h, p_l, c\}$, there exists a cutoff n^* and a constant μ such that the bonus scheme described in Equation (9) induces high effort on all n tasks if $n \geq n^*$. Fix a $\{p_h, p_l, c, \epsilon\}$ tuple. Because (i) below the cap all tasks are identical and (ii) with sufficiently many tasks the probability of hitting the cap is small, we only need to show that the agent is better off exerting high effort on n tasks rather than $n-1$ tasks. The corresponding condition is given by:

$$c \overset{?}{>} (p_h - p_l)(\frac{c}{p_h - p_l} + \epsilon) * \mathbf{P}\left(\sum_{i=1}^{n-1} y_i \leq np_h + \mu\sqrt{n} - 1\right) \tag{14}$$

$$\iff \mathbf{P}\left(\sum_{i=1}^{n-1} y_i \leq np_h + \mu\sqrt{n} - 1\right) \leq \frac{c}{c + (p_h - p_l)\epsilon}. \tag{15}$$

The left-hand side corresponds to the benefit of exerting low effort on one task and the right-hand side to its cost, namely the reduction in the expected total bonus given by the probability of producing a positive unit of output, multiplied by the piece rate, multiplied by the probability the agent has not hit the cap by exerting high effort on the $n-1$ remaining tasks. We know from the Central Limit Theorem that for any δ there exists an n_1 such that, for all $n > n_1$,

$$|\mathbf{P}\left(\sum_{i=1}^{n-1} y_i \leq np_h + \mu\sqrt{n} - 1\right) - \Phi(\frac{p_h + \mu\sqrt{n}}{\sqrt{p_h(1-p_h)(n-1)}})| \leq \delta. \tag{16}$$

Substituting (16) into (15) yields inequality:

$$\Phi(\frac{p_h + \mu\sqrt{n}}{\sqrt{p_h(1-p_h)(n-1)}}) \leq \frac{c}{c + (p_h - p_l)\epsilon} + \delta. \tag{17}$$

Note that, for a given μ, as $n \to \infty$ the term inside the CDF approaches $\frac{\mu}{\sqrt{p_h(1-p_h)}}$. Further, as μ goes to ∞ this term approaches ∞ and hence the CDF approaches 1. Also, the right-hand side can be chosen to be less than 1 so long as $\delta \leq \frac{(p_h - p_l)\epsilon}{c + (p_h - p_l)\epsilon}$. Because we are free to choose δ, this is satisfied. Thus, for any ϵ, we can choose $\delta = \frac{1}{2}\frac{(p_h - p_l)\epsilon}{c + (p_h - p_l)\epsilon}$ and a μ sufficiently large. This completes the proof that the piece rate bonus with a cap can implement high effort on all tasks.

Further, the maximum payment corresponds to the cap:

$$\left(\frac{c}{p_h - p_l} + \epsilon\right)(np_h + \mu\sqrt{n}), \tag{18}$$

and thus the principal's reneging constraint is given by:

$$\left(\frac{c}{p_h - p_l} + \epsilon\right)\left(np_h + \mu\sqrt{n}\right) \leq \frac{1}{r}n(p_h - c). \tag{19}$$

Solving for the value of r that binds this inequality and taking $\epsilon \to 0$ yields the expression listed in the proposition. □

Proof of Proposition 4. Proposition 4 follows immediately from Proposition 3 and Equation (10).

□