Artificial Intelligence in Team Dynamics: Who Gets Replaced and Why?

Xienan Cheng* Mustafa Dogan** Pinar Yildirim‡

June 2025

Abstract

This study investigates the effects of artificial intelligence (AI) adoption in organizations. We ask: (1) How should a principal optimally deploy limited AI resources to replace workers in a team? (2) In a sequential workflow, which workers face the highest risk of AI replacement? (3) How does substitution with AI affect both the replaced and non-replaced workers' wages? We develop a sequential team production model in which a principal can use peer monitoring—where each worker observes the effort of their predecessor—to discipline team members. The principal may replace some workers with AI agents, whose actions are not subject to moral hazard. Our analysis yields four key results. First, the optimal AI strategy involves the stochastic use of AI to replace workers. Second, the principal replaces workers at the beginning and at the end of the workflow, but does not replace the middle worker, since this worker is crucial for sustaining the flow of information obtained by peer monitoring. Third, the principal may choose not to fully exhaust the AI capacity at her discretion. Fourth, the optimal AI adoption increases average wages and reduces intra-team wage inequality.

Keywords: *Artificial intelligence (AI), automation, technology, peer monitoring, teams, organizational structure*

JEL Codes: L20, M21, J31, M54, O33

We are grateful to Jose Apesteguia, Yuxin Chen, Lin William Cong, Rahul Deb, Wouter Dessein, Anthony Dukes, Miguel Espinosa, Nima Haghpanah, Lukas Hensel, Gaoji Hu, Jin Li, Xing Li, Claudio Mezzetti, Weicheng Min, Mariann Ollar, Xianwen Shi, Feng Tian, Yiqing Xing, Yang You, Jidong Zhou, Li-An Zhou in addition to audiences at the Marketing and Technology Conference at Columbia University for their feedback about this project. Yildirim thanks the Wharton Dean's Research Fund and Mack Institute for their generous financial support for the research. Names are listed alphabetically. All errors are our own. All correspondence can be sent to the first author.

*Peking University, e-mail: xienancheng@gsm.pku.edu.cn.

**University of East Anglia, e-mail: m.dogan@uea.ac.uk

‡Wharton School of the University of Pennsylvania and NBER, e-mail: pyild@wharton.upenn.edu.

1 Introduction

Over the past decade, artificial intelligence (AI) made profound changes to how organizations work and how tasks are assigned to workers. Organizations worldwide are adopting AI at unprecedented scales, owing to benefits such as increased consistency of output and reduced production costs. According to a McKinsey survey, approximately 78% of respondents reported that their organizations utilize AI in at least one work function (Singla et al., 2025). Despite its rapid adoption and projected growth, integrating AI into existing organizational structures comes with challenges (Glynn et al., 2024). Specifically, traditional workflows are designed and optimized for human workers, typically operating in teams, working on connected tasks with linked incentives. Thus, when AI is introduced into existing workflows, its effects are unlikely to remain localized. Workers whose tasks are not directly impacted by AI are still likely to be indirectly impacted. At a time when AI is reshaping work and workflows (Bughin et al., 2018; Brynjolfsson et al., 2022) and organizations are thinking about how best to integrate AI into their existing systems, it is essential to ask what the implications of AI are for teams. Yet, the literature to date largely reported effects of AI focusing on individual workers, overlooking team dynamics.

In this study, we focus on this overlooked area and ask the following research questions: How should a principal optimally deploy limited AI resources to replace workers in a team? How does the position of a worker in the production sequence impact his risk of being replaced with AI and his earnings? Would the principal always prefer to fully utilize all available AI resources, or are there any strategic benefits to keeping some slack AI capacity? What are the effects of optimal AI deployment on the intra-team wage inequality, i.e, the wage gap between the highest and the lowest paid workers in a team?

To answer these questions, we develop a model that builds on the framework of Winter (2010). In our setting, a principal (*she*) manages a team of workers (*he*), each performing complementary tasks in sequential order, contributing equally to the team's output. Workers can choose to exert effort or shirk, and their effort choices determine the success of the production process. Within this framework, we examine how the principal should optimally allocate limited AI resources to replace human workers. To not trivially advantage one resource over another, we assume that performing a task is equally costly for AI agents and human workers. However, in line with prior studies (Dogan and Yildirim, 2022; Dogan et al., 2024), an AI agent's effort is not subject to moral hazard—AI always exerts effort when assigned a task. In our setting, while each worker can observe the effort of his predecessor, he cannot tell whether the effort is exerted by a human worker or an AI agent. Winter (2010) shows that, in all-human teams, the optimal compensation scheme takes advantage of peer monitoring by incentivizing workers to shirk upon observing their predecessor shirk, which creates a *domino effect* of shirking among all subsequent workers. Mutual monitoring thus deters workers from shirking and helps the principal to maintain effort within the team. Introducing AI into teamwork may interrupt this domino effect, and thus create trade-offs for the principal.

Specifically, there are three trade-offs that a manager needs to consider when deciding *whether* to replace a human worker. The first and most intuitive trade-off is the *direct cost savings* achieved when replacing a high-wage worker with an AI agent, fixing everything else. However, this benefit may be offset by other counter effects. A second effect, the *direct incentive cost*, arises when an increased probability of replacement heightens a worker's temptation to shirk, making it more expensive or difficult to maintain productivity. The third trade-off, the *indirect incentive cost,* emerges as the higher replacement probability of a worker makes it costlier to incentivize the workers who precede him, when shirking becomes more attractive to them. These trade-offs are resolved differently for team members based on their positions in the production sequence. For clarity, where convenient, we will describe these trade-offs focusing on a team of three workers distinguished by their positions as *front-most*, *middle*, and *end-most*.

Considering the interplay of these three effects, our study yields four key insights. First, we demonstrate that, in the setting we study, the principal prefers a mixed strategy where she *randomizes* the deployment of AI to replace human workers. Such randomization can be practically achieved by replacing workers with AI probabilistically across various team projects or shifts. Put differently, teams where AI replacement is randomized are preferred over deterministic team compositions, where some workers are replaced with certainty. This finding alleviates the concerns raised by pundits about widespread labor displacement effects due to AI (e.g., OECD, 2023; Smith, 2025). Specifically, when deployed optimally, AI affects human workers at the intensive margin, by altering the intensity or frequency of their work, rather than at the extensive margin, or by displacing them (Jiang et al., 2025).

Second, we address *which* team members face the highest and the lowest risk of AI replacement. Our findings show that in the optimal replacement strategy, the middle worker faces the lowest (zero) risk of replacement, so that the information flow among the members of the team can be maintained. Conversely, both the front-most and the end-most workers face a positive replacement risk, with the end-most worker facing a higher risk than the front-most worker.

Third, we observe that the optimal replacement strategy may not fully utilize the available AI capacity. The principal may benefit strategically from keeping some slack AI resources to create an environment of uncertainty. This finding implies that in some cases, a manager will choose to maintain an all-human team with positive probability, despite having unused AI capacity at her disposal to create a human-AI team.

Fourth, we examine how optimal AI adoption impacts workers' wages and how the impact varies by position within the team members. Without AI adoption, the wages of workers increase monotonically along the production sequence, with the end-most worker earning the highest wage. Our results show that optimal AI deployment maintains this wage hierarchy but reduces the wage gap between team members. Specifically, the front-most and middle workers see their wages increase due to AI adoption, whereas the wage of the (highest-paid) end-most worker remains unchanged. This leads to decreased intra-team wage inequality, aligning with Bessen et al. (2025), who finds minimal wage scarring and even wage gains

following technological adoption.

In Section 4, we introduce three extensions to the main model. Section 4.1 focuses on a critical characteristic of team production: synergy among workers (Che and Yoo, 2001). We proxy synergies by the degree of complementarity between the tasks workers carry out. We propose a specific production function that parameterizes this complementarity and find that greater task complementarity enhances the principal's ability to leverage peer monitoring, reducing the wage incentives necessary to deter shirking. However, higher complementarity also implies that a worker's shirking is more detrimental, and this implies the risk of replacing the front-most worker increases and that of the end-most worker decreases. We find that AI's potential to mitigate wage inequality is more pronounced in teams with low task complementarity. In Section 4.2, we focus on task-based automation (Acemoglu and Restrepo, 2018) and introduce an alternate model where AI substitutes for a fraction of a worker's tasks, rather than substituting the worker. We find that replacing a fraction of the tasks of a worker is never optimal, and the principal prefers to replace either all or none of the tasks of a worker. Moreover, as in the main model, only the front-most and the end-most workers face risks of task replacement. Finally, in Section 4.3, we examine how the optimal deployment of AI varies with the underlying network structure of the team. We compare the baseline chain network—whose optimality for human teams is established in Winter (2010)—to a star configuration. In the star network, the central worker—the worker who facilitates the team's information flow—again faces the lowest (i.e., zero) risk of AI replacement, while that for the peripheral workers may be positive. These findings highlight how the network structure among the team members may shape the principal's AI adoption strategy.

For practitioners considering the deployment of AI technologies and their impact on the workforce, we highlight several key strategic considerations. First, short-sighted AI replacement strategies, particularly those primarily driven by cost-cutting motives such as replacing the highest-paid workers, can backfire. A focus on direct costs of replacing workers alone overlooks the *indirect costs* AI introduces in the form of moral hazard. Our findings suggest that, in some cases, it may be optimal to replace a lower-paid worker if doing so minimizes the combined direct and indirect costs. Second, we find that a stochastic AI-human teaming strategy, where work is assigned to AI on a randomized basis, can outperform deterministic assignment strategies. This implies that retaining workers while reallocating their work to AI in a stochastic manner may be more effective than full replacement of the workers. Such strategies not only preserve human capital, but may enhance productivity within teams. Third, managers should be prepared to adjust wages for all team members, including those who are not directly impacted by AI. The wages of all workers in a team can be impacted by AI adoption, even when a worker himself is not substituted by AI. The principal may find herself having to pay a higher wage for the remaining workers after some team members are substituted with AI, despite benefiting from some savings in the form of foregone wages. This suggests that it is imperative for managers to understand that even when AI adoption is localized (i.e, replacing a small set of workers or tasks), its spillover effects can permeate the broader team.

Our research contributes to the literature on the role of peer monitoring in incentivizing effort within teams (e.g., Che and Yoo, 2001; Gibbons and Roberts, 2013; Villas-Boas, 2020; Upton, 2024), as well as the literature modeling moral hazard in production settings (e.g., Sun and Tian, 2018; Tian et al., 2023). We extend this body of work by investigating how the introduction of technology in general, and AI more specifically, reshapes the information structure within teams, with an emphasis on the risk of undermining peer monitoring. We demonstrate that AI can alter the incentives and the compensation of workers. In this sense, our work also complements the recent studies that endogenize information flow within teams when designing incentive structures (e.g., Zhou, 2016; Au and Chen, 2021; Lu and Song, 2025).

This study is also relevant to the literature studying AI effects on worker and firm productivity. A growing number of studies experimentally estimate the effect of AI adoption on worker and business productivity (Dell'Acqua et al., 2023a,b; Otis et al., 2024; Brynjolfsson et al., 2025), with a particular focus on the adoption of generative AI (GenAI) tools. Overall, these studies demonstrate heterogeneous effects of AI adoption on productivity. For instance, while novice workers typically show a higher improvement in productivity compared to experienced workers (Brynjolfsson et al., 2025), these effects may be reversed for complex tasks (Dell'Acqua et al., 2023b) or for businesses with unique needs (Otis et al., 2024). Workers may also strategically shift their focus to tasks that cannot be carried out by AI (Yiu et al., 2025). What is common between these studies and ours is the focus on productivity. Similar to these studies, we recognize that AI—either as a substitute or a complementary production technology—may alter a worker's effort decision. We provide a theoretical lens to explain the mixed findings from the earlier studies. In particular, our study shows two factors which may play a role: (i) the effects of AI do not remain localized—adopting AI to replace a worker may generate negative externalities on other workers; (ii) AI may be deployed suboptimally—e.g., AI may be used indiscriminately to replace (the tasks of) the wrong workers. Our paper highlights that, given its negative externalities, without careful consideration of how to optimally deploy AI, empirical studies may falsely conclude that AI does not improve worker or organizational productivity.

Another stream of the literature focuses on estimating the macro effects of technology adoption, such as effects on employment and wages. A substantial body of research documents the nonuniform effects of automation on employment and wages, with automation disproportionately displacing lower-skilled and less-educated workers (Acemoglu and Restrepo, 2019, 2020; Petrova et al., 2024). In contrast, AI-related displacement appears to be more task- or occupation-specific (Acemoglu et al., 2022). On wages, the evidence is similarly mixed: while some studies report declines due to automation (Acemoglu and Restrepo, 2020; Petrova et al., 2024), others find little to no wage scarring—or even wage gains—associated with automation and AI adoption (Barth et al., 2020; Domini et al., 2022; Bessen et al., 2025). For generative AI specifically, recent work by Humlum and Vestergaard (2025) finds no significant wage or labor displacement effects.

Overall, our study bridges two important strands of research: the literature on the optimal design of incentives in teams and the emerging literature on the organizational implications of AI adoption (e.g., Agrawal et al., 2024). Even though most firms operate in team-based structures, much of the existing research on AI abstracts away from considerations of teams.[1] Recent experimental evidence further underscores the importance of the team perspective. For example, Dell'Acqua et al. (2023a) find that AI integration can interfere with a team's ability to coordinate, and thus decrease individual effort and overall team performance. These findings highlight the need for a more nuanced understanding of human-AI interactions within teams, explicitly considering the externalities that arise when some workers are replaced (or augmented) by AI on the workers who are not. Our study addresses this gap by developing a formal model that captures both the benefits and costs of AI integration in team settings.

The rest of the paper is organized as follows. In Section 2, we set up the model, and in Section 3 we lay out the direct and indirect costs and benefits of AI adoption and describe the optimal AI adoption strategy. We offer several extensions in Section 4, and we provide a discussion of additional considerations and limitations. Finally, in Section 5, we conclude with a brief discussion of our key findings and takeaways for practice.

2 Model

A team is undertaking a project with $n \geq 3$ workers.[2] Each worker (he) $i \in \{1, \ldots, n\}$ performs a task and makes a binary effort decision $e_i \in \{0, 1\}$, where $e_i = 1$ indicates that worker i exerts effort; and $e_i = 0$ indicates he shirks. The cost of effort for worker i is $c(e_i) = ce_i$, where $c > 0$. Workers $1, \ldots, n$ perform their tasks sequentially. Each worker observes only the effort choice of his immediate predecessor.[3] Therefore, the strategy of worker $i \in \{2, \ldots, n\}$, $\sigma_i : \{0, 1\} \mapsto \{0, 1\}$, specifies his decision based on the observed decision e_{i-1} of his predecessor. Since worker 1 does not have a predecessor, his strategy is represented by $\sigma_1 \in \{0, 1\}$.

The project's success depends on the number of workers exerting effort. If k out of n workers exert effort, the project succeeds with probability p_k, where p_k strictly increases in k. Workers' efforts are assumed to be complementary: the probability of success increases at an increasing rate as more workers exert effort, i.e., the marginal change in the probability of success from adding one additional worker is higher when more workers are already contributing effort. Formally, for each $k \leq n - 2$, we have: $p_{k+2} - p_{k+1} > p_{k+1} - p_k$.

The manager of the organization (principal, *she*) can observe only the outcome of the project, not the workers' individual effort choices. Since the project's success is highly valuable to her, she aims to

[1] A few exceptions include Athey et al. (2020), Dogan and Yildirim (2022), and Dogan et al. (2024), which begin to explore how technological change affects organizational design and coordination.

[2] Throughout the paper, we will use the terms team, organization, and network interchangeably.

[3] While we adopt this sequential *chain* structure as given, Winter (2010) demonstrates that such a structure naturally emerges, among the considered alternatives, as the optimal arrangement. In Section 4.3, we will revisit this assumption.

incentivize each worker to exert effort by offering a contract $w \equiv (w_1, \ldots, w_n)$, where w_i specifies the payment worker i receives if the project succeeds. The workers are risk-neutral and seek to maximize their expected compensation net of the cost of effort.[4]

AI Replacement The principal has access to an AI technology capable of replacing human workers in performing their tasks. If AI replaces worker i, it automatically exerts effort, $e_i = 1$, at a cost c, which is incurred by the principal.[5] The AI's effort is just as effective as that of a worker in helping the project succeed. The principal decides (i) *whether* to replace a worker with AI, and if so, (ii) *which* of the n workers to replace, paying attention to the position of the worker in the production process. We allow the principal to use a mixed strategy for both decisions. Specifically, the principal's replacement strategy is:

$$x \equiv (x_1, x_2, \ldots, x_n), \quad \text{with } \bar{x} \equiv \sum_{i=1}^{n} x_i \leq 1,$$

where x_i represents the probability of replacing worker i with AI. The constraint $\bar{x} \leq 1$ allows us to treat AI as a limited resource, closely capturing constraints managers face in reality, and forces the manager to prioritize the worker to replace.[6] The principal may set $\bar{x}$ strictly less than 1 and choose not to fully utilize this capacity.

The workers observe the AI replacement strategy of the principal, but they do not observe its realized outcome, i.e., whether AI replaces a worker, as well as which worker it replaces, if it does (except when a worker himself is replaced). Each worker can observe whether his predecessor exerts effort, but cannot distinguish whether it is a worker or an AI that exerted the effort. But if he observes shirking, he knows that his predecessor is not an AI.

This setup can also be interpreted as a setting where worker i can tell whether the task completed immediately before his own $(i-1)$ will make a contribution to the overall project success or not, rather than directly observing the effort of his predecessor. Because workers and AI make equal contributions to success when they exert effort, worker i cannot tell whether a positive contribution comes from an AI or a worker. However, since AI does not shirk, if the previous task makes no contribution, he can infer that it was carried out by another worker who shirked.

Timing The sequence of events is as follows: The principal first selects a pair of replacement strategy and a compensation scheme, (x, w), which defines a specific game for workers to play. After observing (x, w), the workers make their effort choices. The project outcome is then realized. Finally, the workers

[4] Because workers have limited liability, the optimal contract grants each worker a positive payment if the project succeeds, and no payment in the event of failure.

[5] Keeping the effort cost of the AI similar to that of a worker allows us to study the nontrivial drivers of technology adoption, beyond cost reduction.

[6] We provide a discussion of relaxing this assumption in Section 4.4.

receive their compensation based on the outcome.

3 Analysis

Our objective is to characterize the optimal (x, w) pair, as defined below.

Definition 1. *Replacement strategy and compensation scheme pair (x, w) is optimal if*

(i) it induces an equilibrium in which all workers exert effort, i.e., $(e_1, \ldots, e_n) = (1, \ldots, 1)$ on the equilibrium path;

(ii) among all pairs inducing such an equilibrium, it minimizes the expected cost of compensation for the principal:

$$W_{x,w} \equiv \sum_{i=1}^{n} \left[x_i c + (1 - x_i) p_n w_i \right]. \tag{1}$$

Two factors shape the expected cost of compensation: (i) the principal bears the AI's cost, c, regardless of the project outcome when deployed, and (ii) each worker (if not replaced) is compensated only upon the success of the project, which occurs with probability p_n if all workers exert effort.

Optimal Compensation for a Fixed Replacement Strategy. We begin our analysis by fixing the replacement strategy x and characterizing the optimal compensation scheme conditional on x, which we denote by w^x. The following proposition characterizes w^x and shows that it induces a trigger strategy profile $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*)$ as an equilibrium, where each worker exerts effort as long as no deviation is observed, and shirks otherwise. Formally,

$$\sigma_1^* = 1, \quad \sigma_i^*(e_{i-1}) = \begin{cases} 1 & \text{if } e_{i-1} = 1, \\ 0 & \text{if } e_{i-1} = 0, \end{cases} \quad \text{for all } i \in \{2, \ldots, n\}.$$

Proposition 1. *Fix a replacement strategy $x = (x_1, \ldots, x_n)$. The optimal compensation scheme conditional on x, denoted by $w^x = (w_1^x, \ldots, w_n^x)$, satisfies the following for each worker $i \in \{1, \ldots, n\}$ with $x_i < 1$:*

$$w_i^x = \frac{c}{p_n - \zeta_i^x}, \tag{2}$$

where ζ_i^x is the resulting success rate when worker i shirks:

$$\zeta_i^x = p_{i-1} \sum_{k=1}^{i-1} \frac{x_k}{1 - x_i} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i} p_{n-k+i} + p_{i-1} \frac{1 - \bar{x}}{1 - x_i}. \tag{3}$$

Proposition 1 characterizes the optimal compensation for each worker, when not replaced with AI with certainty. If he is replaced with probability 1, then compensation is irrelevant. With a slight abuse of notation, we denote by W_x the principal's expected compensation cost under the *optimal* scheme w^x, obtained by substituting $w = w^x$ into the general expression of $W_{x,w}$ given in (1):

$$W_x \equiv W_{x,w^x} = \sum_{i=1}^{n} \left[x_i c + (1 - x_i) p_n w_i^x \right]. \tag{4}$$

The proof of Proposition 1 proceeds in three steps. First, we show that under w^x, the trigger strategy profile σ^* constitutes an equilibrium. Second, we show that no alternative compensation scheme with a lower expected cost can support σ^* as an equilibrium. Third, we show that inducing joint effort through another strategy profile apart from the trigger strategy necessarily incurs a higher expected compensation cost for the principal.

In the trigger strategy equilibrium induced by the compensation scheme w^x, the workers are indifferent between exerting effort and shirking upon observing the immediate predecessor exert effort. Then, due to effort complementarity, shirking becomes the optimal choice when a worker observes his immediate predecessor shirk. This generates a *domino effect*: Following a worker's deviation, all successive workers choose to shirk as well. This domino effect, serving as the strongest possible deterrent to deviations, enables the principal to fully leverage the workers' peer monitoring. Replacing a middle worker with AI disrupts the domino effect, as AI does not condition its effort on the actions of other workers. This, in turn, alters the incentives and compensation of the workers. As we shall see shortly, the principal accounts for this disruption when determining her replacement strategy.

Worker i's indifference condition on the equilibrium path, which pins down w_i^x, is given by:

$$p_n w_i^x - c = \zeta_i^x w_i^x,$$

where ζ_i^x is the resulting success rate under the trigger strategy profile from worker i's perspective, when he shirks. Therefore, the payment that worker i receives (if not replaced) upon project success is $w_i^x = \frac{c}{p_n - \zeta_i^x}$. A higher ζ_i^x leads to a higher payment w_i^x, as a higher ζ_i^x makes shirking less consequential in terms of success, and therefore more appealing for worker i, requiring the principal to offer greater compensation to offset this temptation.

Since ζ_i^x depends on worker i's position in the production sequence, so does his compensation. In the absence of AI, as is also noted by Winter (2010), worker compensation increases monotonically, with the end-most worker receiving the highest compensation. This follows from the fact that under the trigger strategy, the end-most worker's shirking results in the smallest decline in the success rate, as it does not induce subsequent shirking.

To understand how a worker's incentive to shirk depends on the AI strategy, we examine ζ_i^x and its

components closely. As shown in equation (3), ζ_i^x is a weighted sum of three success probabilities, each corresponding to a different replacement outcome when worker i deviates.

- If a predecessor of worker i is replaced, all of i's successors shirk, and the project succeeds with probability p_{i-1}.
- If a successor of worker i, say worker k, is replaced, all successors i to $k-1$ shirk, while AI at position k and all later workers exert effort, yielding success probability p_{n-k+i}.
- If no workers are replaced, all of i's successors shirk, resulting in success probability p_{i-1}.

The weights on these outcomes reflect worker i's belief about the replacement outcomes, conditional on not being replaced himself. Specifically, he assigns probability $\frac{x_k}{1-x_i}$ to worker $k \neq i$ being replaced, and probability $\frac{1-\bar{x}}{1-x_i}$ to no replacement at all.

With this understanding, we now examine how the choice of AI adoption affects ζ_i^x. Rewriting equation (3) highlights its dependence on the replacement strategy:

$$\zeta_i^x = p_{i-1} + \sum_{k=i+1}^{n} \frac{x_k}{1-x_i} \left(p_{n-k+i} - p_{i-1}\right). \tag{5}$$

As seen, ζ_i^x, the temptation to shirk for worker $i < n$, increases with (i) the likelihood of replacing worker i, and (ii) the likelihood of replacing any of i's successors. For the end-most worker, ζ_n^x is independent of the replacement strategy x.

Tradeoffs of Replacement. We analyze the tradeoffs that arise from replacing a worker with AI by examining the partial derivative of the principal's expected cost with respect to x_i:

$$\frac{\partial W_x}{\partial x_i} = -\underbrace{(p_n w_i^x - c)}_{\substack{\text{direct} \\ \text{cost saving}}} + \underbrace{(1-x_i) p_n \frac{\partial w_i^x}{\partial x_i}}_{\substack{\text{direct} \\ \text{incentive cost}}} + \underbrace{\sum_{k=1}^{i-1} (1-x_k) p_n \frac{\partial w_k^x}{\partial x_i}}_{\substack{\text{indirect} \\ \text{incentive cost}}}. \tag{6}$$

Equation (6) demonstrates that the principal needs to balance trade-offs coming from three distinct effects to make the optimal replacement decision. The first effect is *the direct cost savings*. This benefit arises because the expected compensation for worker i is $p_n w_i^x$, while the cost of an AI is c. Since $p_n w_i^x > c$, replacing a worker with AI reduces the principal's expected compensation cost. This benefit is the highest for the end-most worker.

The second effect is *the direct incentive cost*. As discussed in the paragraph following Equation (5), all else being equal, increasing the likelihood of replacing worker i amplifies his incentives to shirk, in the event he is not replaced. This is because a higher x_i implies a higher posterior belief that others are

replaced, if he is not replaced. This makes shirking less consequential and more attractive, increasing the compensation the principal needs to offer to incentivize him to work.

The third effect is *the indirect incentive cost,* imposed due to the changes in compensation paid to other workers. If the likelihood of replacing worker i increases, the compensation required to incentivize his predecessors to exert effort also increases. This is, again, due to the reduced consequences of shirking for those preceding worker i. Note that, a worker closer to the end of the work sequence will have a higher number of predecessors, and thus the number of workers whose compensation is impacted by this indirect effect will be higher as well.

The principal must carefully balance these three effects to determine the optimal replacement strategy. For example, replacing the end-most, or the highest paid worker, provides the highest direct cost saving. Yet, replacing him introduces a significant indirect incentive cost, since he has the highest number of predecessors. In contrast, replacing the front-most worker has no indirect incentive cost, as this worker has no predecessors. At the same time, replacing this worker also offers a lower direct cost saving, since his compensation is lower than that of the end-most worker. The direct incentive cost is zero for the end-most worker, but it is less straightforward to calculate it for the other two workers, as the cost varies by the AI strategy and a worker's position. In the discussion that follows, we will contrast these three effects to characterize the optimal replacement strategy.

3.1 Optimal Replacement Strategy

In this section, we characterize the optimal strategy for replacing workers with AI, denoted by x^*. As an initial step, the following proposition and the subsequent discussion highlight a key property of this strategy: randomization. We also explore the trade-offs that the principal faces while deciding her AI strategy.

Proposition 2. *The principal's optimal AI adoption strategy necessarily involves randomization:*

- *no worker is replaced with certainty, i.e.,* $x_i^* < 1$ *for all* i*; and*
- *at least one worker is replaced with positive probability, i.e.,* $x_i^* > 0$ *for some* i*.*

Proposition 2 shows that the principal adopts a randomized replacement strategy, ruling out both full replacement of any worker ($x_i = 1$ for some i) and no replacement at all ($x_i = 0$ for all i). We explain this finding by highlighting the trade-offs in the principal's replacement strategy, which encompasses two critical aspects: *whether* to replace any worker, and, if so, *which* worker(s) to replace.

To see why randomization is optimal, first consider a pure replacement strategy, where the principal replaces one worker with certainty or does not replace any workers at all. Within the set of pure strategies, the principal has two optimal choices: replacing either the front-most or the end-most worker.[7] Both

[7]This is formally stated and proved in Appendix A (Lemma 2).

options result in the same expected compensation, as they create an identical team network structure: $n-1$ sequentially connected workers and one isolated AI agent. This configuration preserves information flow among the nonreplaced workers, enabling the principal to discipline the workers by leveraging peer monitoring to the maximum extent. Replacing a middle worker, however, results in a disconnected network and a disrupted flow of information, increasing the compensation cost incurred by the principal.

Now consider a class of replacement strategies where the principal replaces the front-most and end-most workers with probabilities ρ and $1-\rho$, respectively, for some $\rho \in [0,1]$. The extreme points of this class, with $\rho = 0$ and $\rho = 1$, correspond to the two optimal pure replacement strategies. We show that all interior values of $\rho \in (0,1)$ outperform these pure strategies. Hence, the optimal replacement strategy necessarily involves randomization.

To explain why an interior value of ρ outperforms the optimal pure strategies, consider decreasing ρ, or gradually shifting the replacement probability from the front-most to the end-most worker. Since the end-most worker is compensated more than the front-most worker, this change results in higher direct cost savings for the principal. However, since the compensation of both workers is independent of ρ, the magnitude of these savings remains constant as ρ varies. This is because, in this class of strategies, the principal fully exhausts the AI capacity, and the end-most (front-most) worker knows that the front-most (end-most) worker is replaced with certainty when he himself is not replaced. As a result, ζ_1 and ζ_n (and therefore w_1 and w_n) remain fixed as ρ varies.

Yet, this shift also increases the indirect incentive cost. Specifically, as the replacement probability of the end-most worker $(1-\rho)$ increases, shirking becomes less consequential for all his predecessors (except the front-most worker), prompting the principal to raise their compensation. Importantly, this indirect cost is convex in ρ. The following observations together imply that an interior $\rho \in (0,1)$ is optimal within this class of strategies: (i) both extreme values $\rho = 1$ and $\rho = 0$ yield the same compensation cost, (ii) reducing ρ increases the direct cost savings at a constant rate, and (iii) reducing ρ increases the indirect incentive cost at an increasing rate. Therefore, all randomized replacement strategies with an interior ρ dominate the pure strategies.

Discussion of Randomization. There are a few points of discussion to bring forward related to Proposition 2. First is the optimality of randomization over the pure strategy replacement decisions, and whether the implementation of this strategy may pose challenges in real life. Since human resources may not be easily fungible, randomization can be achieved in other ways. For instance, a common application involves randomizing the assignment of AI to replace workers across projects, where a given task would be carried out by an AI some of the time and by a worker at other times. A randomization schedule may be achieved more easily if workers rotate across different divisions of the organization, and AI can replace them as they rotate out of a division. The schedule may also be carried out via rotating workers across different work shifts (e.g., day/night) or across days of work. Companies like Dmall specialize in

offering such technology solutions to rotate workers with AI across shifts. One application, for instance, replaces daytime security staff at grocery stores with AI surveillance agents during nighttime shifts.[8]

An implication of randomization is the emergence of hybrid (human-AI) teams, where the same task can be carried out by a human worker and an AI. In this outcome, workers are not fully displaced, but see a reduction in their workload, as they are occasionally substituted with AI. While, in line with earlier studies, this finding implies labor displacement effects of new technologies (e.g., Acemoglu and Restrepo, 2020; Chen et al., 2025), they also provide a more mitigated and perhaps a more optimistic perspective. Despite some scholars and pundits expressing fear of sweeping labor displacement effects from AI adoption (Runciman, 2023; OECD, 2023), our findings suggest that permanently replacing workers with AI may be suboptimal and replacing them with AI in a way that varies their participation (e.g, across different projects or work shifts) may be more profitable for an organization.

Heterogeneous Replacement Risk Based on Worker Position. In presenting the next set of results, to deliver sharper outcomes, we will focus on the case where $n = 3$. This three-worker structure captures the distinct incentives of workers within a team. Specifically:

- the *front-most* worker (worker 1) does not have any predecessors, but his contributions will be observed by his successors,
- the *middle* worker (worker 2) both observes the actions of his predecessor and is also observed by his successors, thereby acting as a connector within the team, and
- the *end-most* worker (worker 3) observes his predecessor, but is not observed by a successor, making his effort choice less consequential for the project's success.

With this structure in mind, we proceed to discuss key properties of the optimal randomization scheme outlined in Proposition 2.

Proposition 3. *In the optimal replacement strategy, the middle worker is never replaced with AI. The end-most worker is replaced with a strictly positive probability, and this probability is weakly higher than that of the front-most worker. Formally,* $x_2^* = 0$, $x_3^* > 0$, *and* $x_3^* \geq x_1^* \geq x_2^*$.

Proposition 3 highlights that workers' risk of being replaced by AI varies by their position in the production sequence.

First, the middle worker, who is essential for the connectivity of the information network, does not face the risk of AI replacement. To understand why, consider an arbitrary replacement strategy (x_1, x_2, x_3) in which the middle worker is replaced with positive probability, i.e., $x_2 > 0$. Then consider deviating from this strategy by reallocating more AI resources from the middle to the end-most worker, yielding

[8] See `http://www.dmall.com/product/5641.html` for more information.

$(x_1, 0, x_3 + x_2)$. This shift (i) increases the direct cost savings, as the end-most worker has the highest wage; (ii) reduces the direct incentive costs, as replacing the end-most worker incurs none; but (iii) raises the indirect incentive costs, as the end-most worker has more predecessors. The combined positive effects of (i) and (ii) outweigh the negative effect of (iii), making the principal better off.

Second, the end-most worker's risk of replacement is strictly positive; and while the front-most worker also faces a risk of replacement, this risk is lower than that of the end-most worker. To see why, consider deviating from a given replacement strategy where $x_1 = x_3 + \Delta$ by reallocating an additional Δ amount of resources from the front-most to the end-most worker, yielding $(x_1 - \Delta, x_2, x_3 + \Delta)$. This reallocation generates the same outcomes (i), (ii), and (iii) discussed in the previous paragraph, and the combined positive effects of (i) and (ii) once again outweigh the negative effect of (iii). Thus, any strategy where the front-most worker is replaced at a greater rate than the end-most worker cannot be an optimal AI replacement strategy. This, together with the optimality of randomization (Proposition 2), implies that the risk of replacement the end-most worker faces is always strictly positive, i.e., $x_3^* > 0$.

From a practical perspective, these findings imply that managers wishing to integrate AI optimally into their existing organizational structures need to take into account factors beyond considerations such as costs and technological feasibility of AI. It is essential that managers consider how the integration of AI will disrupt the information flow between workers in a team and follow an AI strategy that preserves the information flow. In our setting, this is equivalent to not replacing the middle agent. Moreover, AI adoption strategy should go beyond the naive replacement strategies, i.e., those that focus on replacing the high-compensation workers, and also focus on reducing the negative externalities of replacement on other workers. This consideration, in our setting, corresponds to the possibility of replacing the front-most worker, rather than focusing on the end-most worker alone.

Utilization of AI Capacity. So far, we have described some key aspects of the optimal AI replacement strategy in an organization, assuming the principal has sufficient resources that she can optimally allocate. An important question is whether the principal would always choose to exhaust these resources available to her.

If AI adoption decisions were purely driven by direct cost savings, a principal should choose to exhaust AI resources. But in Proposition 4, we will find a seemingly counterintuitive result: the principal may choose not to fully utilize the AI capacity available to her. Put differently, replacing workers with technology is *not* always the most preferred, or the most profitable, managerial strategy, and underutilization of AI resources is possible.

Proposition 4. *The principal chooses to underutilize AI capacity, setting $\bar{x}^* < 1$, if and only if the condition $p_1^2 - p_3\, p_0 > 0$ is satisfied.*

Proposition 4 establishes that the principal does not always fully utilize the available AI capacity. It

provides a necessary and sufficient condition for when underutilization is optimal. Given that we have already established the absence of middle replacement (i.e., $x_2 = 0$), the key implication of Proposition 4 is that, despite having the option to increase the replacement probability of the front-most and the end-most workers, the principal may choose not to do so.

We explain the mechanisms driving the underutilization result by focusing on the replacement likelihood of the front-most worker. A marginal decline in the front-most worker's replacement likelihood results in the following two effects: (i) a direct cost saving of $c\frac{\zeta_1^x}{p_3-\zeta_1^x}$, and (ii) a direct incentive cost of $c\frac{p_3(\zeta_1^x-p_0)}{(p_3-\zeta_1^x)^2}$. [9]

Under full utilization, the front-most worker knows that, if he is not replaced, the end-most worker will be replaced with certainty, implying a success rate $\zeta_1^x = p_1$ if he shirks. Plugging in $\zeta_1^x = p_1$ and comparing these two expressions, when the production function satisfies the condition $p_1^2 > p_0 p_3$. The direct incentive cost effect dominates the former effect, therefore a marginal reduction in replacement of the front-most worker (x_1) benefits the principal. This implies that the principal prefers not to utilize the full AI capacity.

When $p_1^2 \leq p_0 p_3$, underutilization (which results in $\zeta_1^x < p_1$) cannot be optimal as the direct cost savings effect dominates the direct incentive cost effect. In this case, increasing the probability of the front-most worker in a direction to move away from underutilization benefits the principal.

An important implication of whether the principal fully utilizes or underutilizes the available AI capacity is how utilization shapes the beliefs of workers about the presence and position of the AI agent in the production sequence. Moving from full to underutilization alters the uncertainty workers face, which can be used as a strategic tool that benefits the principal. With full utilization of AI resources, only the middle worker is facing uncertainty about which worker is replaced. The front-most (the end-most) worker knows that it must be the end-most (the front-most) worker replaced, if he is not replaced. Underutilization introduces an additional layer of uncertainty. All workers are now unsure whether a replacement has taken place.

To see how this uncertainty may benefit the principal, consider moving from full to under utilization by reducing x_1, while keeping $x_2 = 0$ and x_3 constant. This shift results in a reduction in w_1 while w_2, w_3 remain unchanged, which may benefit the principal. At the same time, since x_1 is reduced, now the principal needs to pay the front-most worker more frequently. When the former effect outweighs the latter, underutilization, and the resulting environment of uncertainty, can make the principal better off. A similar strategic use of uncertainty by a principal in organizations has been the focus of several theoretical studies recently (e.g., Halac et al., 2021; Halac, 2025).

Impact of AI adoption on Wages by Worker's Team Position. Having demonstrated that the optimal AI adoption strategy in a team setting involves randomization, we would like to next investigate

[9] Note that the indirect incentive cost is zero for the front-most worker.

how, as the principal adopts AI, the wages and the expected payoffs of the workers change, and do so conditional on their position in the production sequence. Let w_i^0 be the optimal compensation of worker i in the absence of AI adoption, and $w_i^{x^*}$ be the optimal compensation following optimal AI replacement. The following proposition shows that, while wages of some workers may increase, the intra-team wage hierarchy of all-human teams is maintained under optimal AI adoption.

Proposition 5. *Optimal AI adoption preserves the order of the wages based on the position of the workers ($w_3^{x^*} > w_2^{x^*} > w_1^{x^*}$). Moreover, after optimal AI adoption, worker 1 and worker 2's wages increase ($w_1^{x^*} > w_1^0, w_2^{x^*} > w_2^0$) and worker 3's wage remains identical ($w_3^{x^*} = w_3^0$).*

Proposition 5 provides two key insights. First, optimal AI adoption increases the wages of the front-most and the middle workers, but not the wage of the end-most worker, for whom the wage remains unchanged. The wage increase for the first two workers is due to the positive replacement probability of their successor workers, which makes shirking more appealing for them. Since the end-most worker does not have a successor, the likelihood of success following his shirking does not depend on the AI adoption strategy; thus, his wage remains unchanged.

A second insight concerns the potential impact of AI replacement on the intra-team wage hierarchy. Proposition 1 and the findings from Winter (2010) imply that with all-human teams, optimal wages increase monotonically in the worker's position in the production sequence. A suboptimal adoption of AI can, in principle, disrupt this hierarchy–as we formally establish in Appendix A.3 (Lemma 3). Proposition 5 shows that under the optimal AI strategy, the wage hierarchy of all-human teams is preserved, despite the upward pressure on the wages.

Next, we investigate how the intra-team wage gap (i.e., the dispersion between the highest and the lowest wage) changes following AI adoption. For a given replacement strategy x, we denote this wage gap by Gap^x:

$$Gap^x = \max_i\{w_i^x\} - \min_i\{w_i^x\}.$$

Note that, in a team with three workers, $\max_i\{w_i^x\} = w_3^{x^*}$, and $\min_i\{w_i^x\} = w_1^{x^*}$, therefore $Gap^{x^*} = w_3^{x^*} - w_1^{x^*}$. A comparison of the wages indicates that following optimal AI adoption, the intra-team wage gap declines. Corollary 1 states this finding formally.

Corollary 1. *Intra-team wage gap decreases following optimal AI adoption: $Gap^{x^*} > Gap^0$.*

Our finding that highlights the decline in the intra-team wage gap following technology adoption can be contrasted with those from earlier studies. While several studies point to a widening in pay inequality as a result of technology, depending on the skill level of workers (Acemoglu and Loebbing, 2024) or the job of the worker within a firm (Barth et al., 2020), the extent to which these workers' tasks are exposed

to technology may be vastly different. Focusing on a narrow context where the workers are, aside from their position in the production sequence, identical and make the same contributions to the success of their firms, we show a decline in the intra-team wage gap. While this finding may seem contradictory, it is not; as the contradiction can be explained by how competitive market forces influence workers of different skill and education levels. In our setting, we are able to abstract away from these confounds and demonstrate a declining wage gap that is directly attributable to how AI adoption alters the incentives in a team.

The findings thus far may paint an optimistic picture about the effects of new technologies on organizations, with only partial job losses and wages weakly increasing. It is important to consider the potential effects on workers' overall earnings, particularly if workers face income losses during their "off" time. To this end, Proposition 6 summarizes the change in workers' payoffs, defined as $p_3 w_i^x - c$ for worker i, comparing them before and after the optimal deployment of AI.

Proposition 6. *Following the optimal AI strategy, the middle worker's payoff increases, the end-most worker's payoff decreases, and the front-most worker's payoff may increase or decrease relative to the payoffs in the absence of AI.*

Given the probabilistic nature of work following optimal AI adoption, it is useful to think about the expected payoffs, in addition to the wages earned when workers are not displaced. Proposition 6 completes this picture and underscores that even in homogeneous teams, AI adoption creates winners and losers. Depending on their position, some workers are better off and some are worse off. Recall that the front-most and the middle workers' wages increase, whereas that of the end-most worker remains identical to the wages before AI adoption. Moreover, the front-most and the end-most workers experience partial replacement, while the middle worker does not. Combining these effects, the end-most worker, who is the highest earner, loses payoff. The middle worker, who is essential for maintaining the information flow in the network and thus is not replaced, faces higher payoffs. The front-most worker may experience an increase or decrease in his payoff depending on his extent of replacement and wage increase. Following these changes, the payoffs may no longer follow the hierarchy that the wages follow; and we discuss these changes in more detail in the context of a specific production function in Appendix C.

4 Extensions

4.1 Team Synergies and Task Complementarity

A key aspect of teamwork is the idea that the tasks, or the effort carried out by individuals in the team, build on each other to result in the success of the project. Our main model assumes such complementarity, but how does the degree of task complementarity alter the likelihood of AI replacement and wage outcomes

of workers? In this subsection, we will explore these questions. To facilitate the analysis, we will adopt a production function that incorporates a parameter to capture the degree of complementarity of tasks, the *O-ring* production function, as known and well cited in the literature (Kremer, 1993; Winter, 2004). Let $p_k = \alpha^{n-k}$, where $\alpha \in (0,1)$ calibrates the degree of complementarity.[10] The smaller the α, the more complementary the efforts are. Put differently, α indicates how consequential the effort of a worker is to success. Proposition 7 provides the optimal replacement strategy and wages with this production function.

Proposition 7. *When $p_k = \alpha^{n-k}$, the optimal replacement strategy is as follows:*

$$x_1^* = \frac{\sqrt{1+\alpha}-1}{\alpha}, \quad x_2^* = 0, \quad x_3^* = 1 - x_1^*.$$

As the degree of complementarity increases (α becomes smaller), the likelihood of the front-most (end-most) worker being replaced increases (decreases). Moreover, the workers' compensations satisfy

$$w_1^{x^*} = \frac{c}{1-\alpha^2}, \quad w_2^{x^*} = \frac{c}{(1-\alpha)\sqrt{1+\alpha}}, \quad w_3^{x^*} = \frac{c}{1-\alpha}.$$

Proposition 7, while characterizing the optimal replacement strategy under the O-ring production function, also affirms our earlier results. Specifically, it confirms that (i) the optimal strategy involves randomization (Proposition 2), (ii) there is no middle replacement (Proposition 3). At the same time, the principal chooses to always fully utilize the AI capacity because, under this functional form, the necessary and sufficient condition for full utilization (stated in Proposition 4) is satisfied; $p_1^2 \leq p_0\, p_3$ is satisfied as $p_1^2 = \alpha^4$, and $p_0 p_3 = \alpha^3$.

The proposition indicates that as worker efforts become more complementary, i.e., when α gets smaller, the likelihood of replacing the front-most worker increases and that of the end-most worker decreases. In fact, in the extreme case when $\alpha \to 0$, both x_1 and x_3 converge to 0.5. This is intuitive, because when team efforts are highly complementary, a single worker's shirking is enough to compromise the success of the project. In consequence, the extent to which the domino effect discussed in Section 3 deters shirking is limited, and it does not play a significant role in disciplining the workers. When the domino effect is less potent, the wages of workers vary less by position, thus the direct cost savings are similar whether the principal replaces the front-most or the end-most worker. At the same time, the direct and indirect incentive costs are negligible, once again implying that replacing these two workers will yield similar outcomes for the principal. Taken together, higher task complementarity results in more homogeneous

[10] One may interpret this functional form as a task structure, where each worker is responsible for a task that succeeds with probability 1 if the worker exerts effort or α if he shirks. The project succeeds only if all tasks succeed.

forms of replacement, moving both the front-most and the end-most worker's replacement probabilities closer to 0.5.

When we turn to wages, we see that as task complementarity increases (as α gets smaller), the wages of all workers decline. This is intuitive, since a higher degree of task complementarity disciplines workers, reducing the need for the principal to incentivize them through higher wages.

Recall from Corollary 1 that, optimal AI adoption reduces the intra-team wage gap. Thus, a natural question to ask is, how does this inequality reducing effect of AI vary across teams with high- and low-complementarity tasks? Under the O-ring production function, in the absence of AI adoption, the wage gap is $Gap^0 = \frac{c\alpha(1+\alpha)}{1-\alpha^3}$, and that under the optimal AI strategy is $Gap^{x^*} = \frac{\alpha c}{1-\alpha^2}$. Then

$$\frac{Gap^{x^*}}{Gap^0} = 1 - \frac{\alpha}{(1+\alpha)^2}.$$

It is easy to see that the ratio is always lower than 1, indicating a reduction in intra-team wage inequality with the adoption of AI. At the same time, as task complementarity increases (as $\alpha \to 0$), the potency of AI in reducing this gap is lower. Put differently, AI makes a bigger difference for wage inequality for teams where tasks are highly independent of each other, and the shirking of a single team member is less consequential to the project's success.

In Appendix C, we also extend the analysis to compare payoffs between workers, in addition to comparing payoffs of workers before and after AI adoption. The analysis provides a robustness check and shows that the payoffs change in identical directions to those discussed in Proposition 6.

4.2 Task-Based AI Substitution

Our analysis thus far focused on a scheme in which a worker is replaced by an AI for his entire task. However, in real-world applications, AI systems often complement human labor by taking over only a subset of tasks, resulting in substitution at the task level rather than complete worker replacement. To capture this possibility, we extend the model to allow AI to substitute for a fraction of a worker's tasks. We show that, in equilibrium, the principal prefers full worker-level replacement over partial task-level replacement.

Let x_i denote the fraction of worker i's tasks performed by the AI, thereby proportionally reducing his effort cost to $(1 - x_i)c$. For consistency with our main analysis, we assume an AI capacity of 1:

$$\bar{x} = \sum_{i=1}^{n} x_i \leq 1.$$

With this modification, each position in the network is occupied by a worker-AI pair. As in the main model, the success of the project depends on the number of positions *effectively* exerting effort. If worker i exerts effort, the entire pair is considered to exert effort. If the worker shirks, only the AI, which performed

a fraction x_i of tasks, remains functional. Thus, the pair contributes to effective effort with probability x_i. When k worker-AI pairs are effectively exerting effort, the project succeeds with probability p_k. This adjustment allows the success probability to reflect partial AI substitution within each position.

We also modify the peer monitoring structure. A shirking worker is detected by his successor only if the worker-AI pair appears to be shirking, which occurs with probability $1 - x_i$. Because the AI handles a fraction x_i of worker i's tasks, this partially conceals the worker's shirking, lowering the probability of detection.

Finally, we reconfigure the compensation structure. Under this task-based AI adoption scheme, compensation is no longer tied to a worker's entire role, but is instead determined per unit of task performed. To make this distinction clear, we now use w_i to denote the compensation per task carried out by worker i. If a fraction x_i of his tasks is handled by AI, his total compensation upon successful project completion becomes $(1 - x_i)w_i$. For any pair (w, x) that induces joint effort among workers, the principal's expected compensation cost remains the same as in Equation (1).

Worker Incentives and Optimal Compensation. The following result, a counterpart to Proposition 1, characterizes the optimal compensation scheme for a given task-based AI adoption strategy x.

Proposition 8. *Fix a task-based AI adoption strategy $x = (x_1, \ldots, x_n)$. The optimal compensation scheme conditional on x, denoted by $w^x = (w_1^x, \ldots, w_n^x)$, satisfies:*

$$w_i^x = \frac{c}{p_n - \zeta_i^x}, \text{ for each } i \in \{1, \ldots, n\} \text{ with } x_i < 1,$$

where

$$\zeta_i^x = x_i p_n + \sum_{k=1}^{n-i} \left[p_{n-k}\, x_{i+k} \prod_{j=i}^{i+k-1} (1 - x_j) \right] + p_{i-1} \prod_{j=i}^{n} (1 - x_j)\,.$$

Proposition 8 characterizes the optimal compensation for workers whose tasks are not entirely performed by the AI, as otherwise compensation is irrelevant. As in the main model, the optimal compensation scheme induces effort by supporting a trigger strategy profile as an equilibrium, and ensures that workers remain indifferent on the equilibrium path. The compensation of worker i decreases with $p_n - \zeta_i^x$, the difference in the probability of success when he exerts effort versus when he shirks. The explicit value of this difference is:

$$\begin{aligned} p_n - \zeta_i^x &= (1 - x_i)p_n - \sum_{k=1}^{n-i} \left[p_{n-k}\, x_{i+k} \prod_{j=i}^{i+k-1} (1 - x_j) \right] - p_{i-1} \prod_{j=i}^{n} (1 - x_j) \\ &= (1 - x_i) \left(p_n - \sum_{k=1}^{n-i} \left[p_{n-k}\, x_{i+k} \prod_{j=i+1}^{i+k-1} (1 - x_j) \right] - p_{i-1} \prod_{j=i+1}^{n} (1 - x_j) \right) \end{aligned}$$

The key insight here is that worker i's per-task compensation decreases proportionally with $1 - x_i$, so

his total compensation, $(1-x_i)w_i^x$, remains unchanged regardless of how many tasks he retains. This is because AI replacement affects both production and monitoring in parallel: while it lowers the worker's effort cost by offloading tasks, it also reduces the detectability of shirking. As a result, unless a worker's entire task is replaced, partial replacement—where AI handles only a fraction of his tasks—offers no clear advantage to the principal. This intuition motivates the formal analysis that follows.

Optimal task-based AI substitution. For conciseness, we provide expressions for the wages for a three-worker sequential production team, or the front-most, middle, and end-most workers, respectively:

$$\begin{aligned} w_1^x &= \frac{c}{(1-x_1)\left(p_3 - [x_2p_2 + (1-x_2)x_3p_1 + (1-x_2)(1-x_3)p_0]\right)}, \\ w_2^x &= \frac{c}{(1-x_2)\left(p_3 - [x_3p_2 + (1-x_3)p_1]\right)}, \\ w_3^x &= \frac{c}{(1-x_3)\left(p_3 - p_2\right)}. \end{aligned}$$

Then, by focusing on a replacement strategy where no worker is fully replaced ($x_i < 1$ for each i) we can formulate the principal's problem of minimizing the expected compensation cost as:

$$\begin{aligned} \min_{\substack{x_1,x_2,x_3\in[0,1) \\ x_1+x_2+x_3\leq 1}} \quad & (x_1+x_2+x_3)\,c + \frac{p_3c}{p_3 - [x_2p_2 + (1-x_2)x_3p_1 + (1-x_2)(1-x_3)p_0]} \\ & + \frac{p_3c}{p_3 - [x_3p_2 + (1-x_3)p_1]} + \frac{p_3c}{p_3 - p_2}. \end{aligned}$$

It is clear that the solution to this restricted problem ($x_i < 1$ for each i) yields $(x_1, x_2, x_3) = (0, 0, 0)$, as x_1, x_2, and x_3 all have strictly positive partial derivatives. Then, to determine the globally optimal strategy, we compare the solution $(x_1, x_2, x_3) = (0, 0, 0)$ with the strategies that involve full replacement of a worker: $(x_1, x_2, x_3) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. This means that the optimal solution in this task-based setting reduces to selecting the best full-replacement strategy, which aligns with the optimal pure-replacement strategy from the main model. As established earlier, this optimal strategy involves fully replacing either the front-most or the end-most worker. Proposition 9 formalizes this result.

Proposition 9. *The optimal task-based AI adoption strategy involves fully replacing either the entire task of the front-most or the entire task of the end-most worker.*

Interestingly, while this alternative model allows the AI to handle a fraction of a worker's tasks, this possibility does not materialize in equilibrium and full replacement of workers' tasks is preferred by the principal. This outcome is closer to the observations from reality, where, workers' tasks are displaced in their entirety, which is equivalent to replacing a worker, connecting our findings to the lay observations from reality.

Finally, we turn our attention to the utilization of AI capacity, and question what is different in a task-based replacement relative to the benchmark model. To this end, we inquire a scenario where $n \geq \mathcal{A} > 0$ denotes the total AI capacity available to the principal, and the replacement strategy $x = (x_1, \ldots, x_n)$ satisfies $\sum_{i=1}^{n} x_i \leq \mathcal{A}$. Similar to what we noted for the main model, a fractional task-based replacement weakens peer monitoring and raises compensation costs. Consequently, with task-based substitution, the principal prefers to replace the entire task of some workers, while letting the other workers carry out their tasks in their entirety, but does not facilitate fractional AI allocation. Corollary 2 provides the solution for a general AI capacity, following the same approach as in Proposition 9.

Corollary 2. *The optimal AI adoption strategy allows for no fractional task allocation and replaces the entire tasks of $\lfloor \mathcal{A} \rfloor$ workers. The replacement is done in a way that ensures the remaining workers, whose tasks are not replaced, are consecutive to each other.*

An immediate implication of the corollary is the possibility of underutilization of the AI capacity. For example, if $\mathcal{A} < 1$, the optimal strategy is to refrain from adopting AI altogether. This finding bolsters the idea that even under a limited AI capacity, the principal may, under some circumstances, choose not to exhaust it.

4.3 Network Structure

In the main part of the paper, we focus on an organization with a chain network structure, as its optimality is proven in Winter (2010). The chain structure well represents organizations where tasks are carried out sequentially among team members. Naturally, for reasons we abstract away from in this paper, teams may be organized in other structures than the chain structure. Another commonly observed structure is the *star* network, where some team members first simultaneously carry out their tasks, followed by a central, possibly higher level, executive carrying out his own. In this section, we will extend our framework to the star network and summarize the insights identical to and different from a chain environment.

Consider a team where $n-1$ *peripheral* workers (denoted by $1, \cdots, n-1$) are observed by a *central* worker, who may be a manager (denoted by n). Proposition 10 demonstrates the robustness of two of our earlier insights: (i) a team member who is essential to maintaining information flow between team members is less likely to be replaced, and (ii) a randomized replacement strategy may still be optimal.

Proposition 10. *In a star structure, all AI replacement strategies satisfying $\sum_{i=1}^{n-1} x_i = 1$ and $x_n = 0$ are optimal, if the condition $p_{n-1}^2 - p_n p_{n-2} \leq 0$ is satisfied.*

The proposition suggests that, the manager—the team member who is essential for maintaining the information flow—is never replaced, and the peripheral workers face a higher risk of being replaced with AI. Since these peripheral workers are identical, any AI replacement strategy, where the probability of

replacement sums up to the AI capacity, is optimal when the sufficient condition $p_{n-1}^2 - p_n p_{n-2} \leq 0$ is met. This implies that the key insights we derive in our benchmark model, while shaped by the characteristics of the network, are not specific to the chain network assumption, and carry forward to a star organizational structure as well.

4.4 Additional Considerations and Discussion

Our study is built on a series of key assumptions. While these assumptions capture various realistic aspects of AI, they are not without loss of generalizability. We provide a qualitative discussion on a set of possible additional questions in this section.

Heterogeneity of Team Members. In the main model, we purposefully abstract away from creating asymmetries among workers other than their position in the network. We also assume that each task contributes identically to the success of the project. We do so to demonstrate that informational position alone is sufficient to generate differential risks of AI replacement and wage outcomes. Team members may, however, vary in other ways than their position, for instance in their skills, or the cost of effort. Moreover, the contributions of tasks to the project success may vary by their position. Without building a full-fledged model, we can argue that the replacement probability of the workers may look different from what our propositions suggest, and the middle worker may face a non-zero probability of replacement if he is more costly or lower-skilled than his peers in the team.

Imperfect observability of AI utilization. In the main model, while the workers directly observe the realized outcome of the AI strategy, the replacement itself is covert. This setting may naturally fit to cases where sequential tasks are carried out by workers over a period of time or in physically distant settings. At the same time, in other environments, partial or full observability of worker replacement may be feasible.

Consider, for instance, a setting where the timeline of the game is altered to facilitate observability. Specifically, the principal first chooses a (potentially mixed) AI replacement strategy, and then the realized replacement is publicly observed. Second, then the principal chooses a compensation strategy. Finally, workers choose their efforts and the outcome of the project is realized.

In this setting, the principal's payoff associated with any replacement strategy x will be linear in pure strategies. Moreover, we know that there are two optimal pure strategies: replacing the front-most and the end-most worker. Therefore, any randomization between replacing these two workers will be optimal. Since payoffs from randomization are equivalent to those from pure replacement strategies, the principal does not have a strict preference for randomization. These observations together suggest that imperfect observability benefits the principal.

Another important observation here is that, in this alternate perfect observability setting, strategic

uncertainty or shaping the worker beliefs about replacement is no longer feasible. Thus, underutilization of AI resources is no longer optimal, and the principal utilizes the full AI capacity.

5 Conclusion

Owing to the rapid growth of robotization and AI, organizations are going through a profound transformation about what work is and how work is done. New AI technologies have transformed how employees interact with each other as well, as their tasks are partially or fully carried out by the AI agents. Given the scale and scope of this transformation, it is essential to ask how work and groups of employees should be re-optimized around these new technologies.

There is a growing literature on how machines alter worker and business performance (Otis et al., 2024) and human decision-making (De Véricourt and Gurkan, 2023; Boyacı et al., 2024; Li et al., 2024; Burtch et al., 2025). Much of the existing literature focuses on individual-level performance effects of the workers who adopt AI (e.g., Dell'Acqua et al., 2023a) or the labor displacement and wage changes due to AI (e.g., Chen et al., 2025; Humlum and Vestergaard, 2025). We contribute to these understudied areas by focusing on the effects of AI agents on worker productivity from a team perspective, explicitly modeling how adoption of AI alters the motivation of the workers and the incentives set by the principal. We demonstrate how a principal integrates AI in a team previously optimized for human employees, and characterize the replacement risk and resulting wage changes for each team in the production sequence based on their position.

First, our analysis reveals that optimal AI adoption involves stochastic rather than deterministic labor force replacement. AI resources are best utilized to complement the workers' roles, possibly in a way that alters or reduces the workload, while maintaining the critical benefits of human worker input. This finding aligns with recent theoretical work emphasizing the importance of human-AI complementarity in maximizing organizational productivity (Brynjolfsson and Mitchell, 2017).

Second, we demonstrate that the likelihood of AI replacement varies by a worker's position within the team's information network. To maintain essential information flows and leverage peer monitoring effectively, it is optimal to minimize the replacement risk for the middle worker, while setting positive replacement probabilities on the front-most and end-most team members. The organizational position relates to the recent literature emphasizing the critical role of information and communication in teams for organizational efficiency (Garicano, 2000; Angelucci, 2017; Matouschek et al., 2025).

Third, we investigate the wage implications of optimal AI integration within production teams. In his analysis of the AI effects on wages and pay inequality, Acemoglu (2025) states that while theoretically it is possible for AI to reduce wages and exacerbate inequality, there is little evidence in practice for these predictions. Our findings provide theoretical support. We suggest that the strategic adoption of AI does not necessitate wage reductions; instead, managers should anticipate comparable or higher wages post AI

adoption relative to the wages before AI adoption. Importantly, while existing wage hierarchies remain unchanged following AI adoption, the intra-team wage gap is significantly reduced as lower-paid workers receive wage increases.

Collectively, these insights challenge prevailing assumptions and provide practical managerial guidance for navigating AI integration. Table 1 summarizes these key managerial prescriptions.

Table 1: Summary of Key Managerial Insights

Question	Findings from our setting
How should a manager optimally integrate AI to a human team?	AI deployment to replace workers should be randomized, e.g., across different projects or work shifts.
Which workers in the team are more likely to be replaced with AI?	The front-most and the end-most workers face higher replacement risk. The middle worker should not be replaced.
How does AI adoption affect worker wages?	Wages increase for all except the end-most worker, whose wage remains unchanged.
How does the optimal deployment of AI impact intra-team wage inequality?	Intra-team wage gap declines following optimal AI adoption.
When is AI's wage inequality reducing effect more potent?	AI's wage inequality reducing effect is more potent in teams with lower complementarity.
Is it better for AI to carry out a fraction of a worker's tasks, rather than fully substituting him?	No, the principal prefers to replace a worker rather than a fraction of tasks.

While AI stands to offer productivity gains, our findings highlight that its integration of AI to human systems must be carefully designed to maintain the information flows between the team members. By doing so, organizations can benefit from replacing workers with AI while minimizing the negative externalities on the non-replaced members of the team.

References

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy*, 40(121):13–58.

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340.

Acemoglu, D. and Loebbing, J. (2024). Automation and polarization. *MIT Working Paper*.

Acemoglu, D. and Restrepo, P. (2018). Modeling automation. *AEA Papers and Proceedings*, 108:48–53.

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Acemoglu, D. and Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6):2188–2244.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2024). Artificial intelligence adoption and system-wide change. *Journal of Economics & Management Strategy*, 33(2):327–337.

Angelucci, C. (2017). Motivating agents to acquire information. *Working paper, Available at SSRN 2905367*.

Athey, S. C., Bryan, K. A., and Gans, J. S. (2020). The allocation of decision authority to human and artificial intelligence. *AEA Papers and Proceedings*, 110:80–84.

Au, P. H. and Chen, B. R. (2021). Matching with peer monitoring. *Journal of Economic Theory*, 192:105172.

Barth, E., Roed, M., Schone, P., and Umblijs, J. (2020). How robots change within-firm wage inequality. *No. 13605. IZA Discussion Papers*.

Bessen, J., Goos, M., Salomons, A., and Van den Berge, W. (2025). What happens to workers at firms that automate? *Review of Economics and Statistics*, 107(1):125–141.

Boyacı, T., Canyakmaz, C., and De Véricourt, F. (2024). Human and machine: The impact of machine input on decision making under cognitive limitations. *Management Science*, 70(2):1258–1275.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140(2):889–944.

Brynjolfsson, E. and Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370):1530–1534.

Brynjolfsson, E., Mitchell, T., and Rock, D. (2022). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 112:43–47.

Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A., and Subramaniam, A. (2018). Skill shift: Automation and the future of the workforce. Technical report, McKinsey Global Institute.

Burtch, G., Greenwood, B. N., and Watson, J. (2025). The effect of gunshot detection technologies on policing practices: An empirical examination of the chicago police department. *Available at SSRN 5223847*.

Che, Y.-K. and Yoo, S.-W. (2001). Optimal incentives for teams. *American Economic Review*, 91(3):525–541.

Chen, W. X., Srinivasan, S., and Zakerinia, S. (2025). Displacement or complementarity? The labor market impact of generative AI. Technical report, Harvard Business School Working Paper.

De Véricourt, F. and Gurkan, H. (2023). Is your machine better than you? you may never know. *Management Science*.

Dell'Acqua, F., Kogut, B., and Perkowski, P. (2023a). Super Mario meets AI: Experimental effects of automation and skills on team performance and coordination. *Review of Economics and Statistics*, pages 1–47.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023b). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Dogan, M., Jacquillat, A., and Yildirim, P. (2024). Strategic automation and decision-making authority. *Journal of Economics & Management Strategy*, 33(1):203–246.

Dogan, M. and Yildirim, P. (2022). Managing automation in teams. *Journal of Economics & Management Strategy*, 31(1):146–170.

Domini, G., Grazzi, M., Moschella, D., and Treibich, T. (2022). For whom the bell tolls: The firm-level effects of automation on wage and gender inequality. *Research Policy*, 51(7):104533.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108(5):874–904.

Gibbons, R. and Roberts, J. (2013). Economic theories of incentives in organizations. In Gibbons, R. and Roberts, J., editors, *The Handbook of Organizational Economics*, pages 56–99. Princeton University Press, Princeton, New Jersey, USA.

Glynn, K., Kaplan, B., and Baker, R. (2024). No one said it was easy: Pain points with AI implementations. https://deloitte.wsj.com/cfo/no-one-said-it-was-easy-pain-points-with-ai-implementations-9b5818f1. Accessed: 2025-05-17.

Halac, M. (2025). Contracting for coordination. *Journal of the European Economic Association*. Advance online publication.

Halac, M., Lipnowski, E., and Rappoport, D. (2021). Rank uncertainty in organizations. *American Economic Review*, 111(3):757–786.

Humlum, A. and Vestergaard, E. (2025). Large language models, small labor market effects. Working Paper 33777, National Bureau of Economic Research.

Jiang, W., Park, J., Xiao, R. J., and Zhang, S. (2025). AI and the extended workday productivity, contracting efficiency, and distribution of rents. Working Paper.

Kremer, M. (1993). The O-ring theory of economic development. *The Quarterly Journal of Economics*, 108(3):551–575.

Li, H., Li, J., Luo, Y., and Zhang, X. (2024). Ai persuasion, bayesian attribution, and career concerns of doctors. *Available at SSRN 4973720*.

Lu, Z. and Song, Y. (2025). Network-based peer monitoring design. *Journal of Economic Theory*, 224:105969.

Matouschek, N., Powell, M., and Reich, B. (2025). Organizing modular production. *Journal of Political Economy*, 133(3):986–1046.

OECD (2023). *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. OECD Publishing, Paris.

Otis, N., Clarke, R., Delecourt, S., and Holtz, D. (2024). The uneven impact of generative AI on entrepreneurial performance. In *Academy of Management Proceedings*, volume 2024, page 21128. Academy of Management Valhalla, NY 10595.

Petrova, M., Schubert, G., Taska, B., and Yildirim, P. (2024). Automation, career values, and political preferences. *Available at SSRN 4728005*.

Runciman, D. (2023). The end of work: Which jobs will survive the AI revolution? *The Guardian*. Accessed: 2025-05-17.

Singla, A., Sukharevsky, A., Yee, L., Chui, M., and Hall, B. (2025). The state of AI: How organizations are rewiring to capture value. `https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai`. Accessed: 2025-05-17.

Smith, R. A. (2025). AI is starting to threaten white-collar jobs—few industries are immune. *The Wall Street Journal*. Accessed: 2025-01-09.

Sun, P. and Tian, F. (2018). Optimal contract to induce continued effort. *Management Science*, 64(9):4193–4217.

Tian, F., Zhang, F., Sun, P., and Duenyas, I. (2023). Dynamic moral hazard with adverse selection–probation, sign-on bonus and delayed payment. *Working paper, Duke University, Durham, NC*.

Upton, H. (2024). Implicit incentives and delegation in teams. *Management Science*, 70(7):4722–4741.

Villas-Boas, J. M. (2020). Repeated interaction in teams: Tenure and performance. *Management Science*, 66(3):1496–1507.

Winter, E. (2004). Incentives and discrimination. *American Economic Review*, 94(3):764–773.

Winter, E. (2010). Transparency and incentives among peers. *The RAND Journal of Economics*, 41(3):504–523.

Yiu, S., Seamans, R., Raj, M., and Liu, T. (2025). Strategic responses to technological change: Evidence from online labor markets. *The Wharton School Research Paper*.

Zhou, J. (2016). Economics of leadership and hierarchy. *Games and Economic Behavior*, 95:88–106.

Appendix

Appendix A Supporting Statements

This appendix section presents supporting statements omitted from the main text for brevity, along with their corresponding proofs.

A.1 Existence

We first establish the existence of the optimal AI replacement.

Lemma 1. *The optimal replacement strategy exists.*

Proof of Lemma 1.
We first prove that the principal's payoff is continuous with respect to x. If $x_i < 1$ for all i, then

$$\frac{W_x}{c} = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{(1-x_i)\,p_n}{p_n - \zeta_i^x}.$$

Let $\vec{x}_j$ be a vector whose jth element is 1 and all other elements are zeros, then

$$\lim_{x\to\vec{x}_j} \frac{W_x}{c} = 1 + \sum_{i\neq j} \frac{p_n}{p_n - \zeta_i^{\vec{x}_j}} + \lim_{x\to\vec{x}_j} \frac{(1-x_j)\,p_n}{p_n - \zeta_j^x}.$$

As can be seen in equation (3) when $x_j = 1$, the zero denominator makes ζ_j^x not well-defined. (3) implies

$$p_{j-1} \leq \zeta_j^x \leq p_{n-1}$$

and thus

$$\frac{(1-x_j)\,p_n}{p_n - p_{j-1}} \leq \frac{(1-x_j)\,p_n}{p_n - \zeta_j^x} \leq \frac{(1-x_j)\,p_n}{p_n - p_{n-1}}.$$

Note that

$$\lim_{x\to\vec{x}_j} \frac{(1-x_j)\,p_n}{p_n - p_{j-1}} = 0$$
$$\lim_{x\to\vec{x}_j} \frac{(1-x_j)\,p_n}{p_n - p_{n-1}} = 0.$$

By squeeze theorem,

$$\lim_{x\to\vec{x}_j} \frac{(1-x_j)\,p_n}{p_n - \zeta_j^x} = 0.$$

So

$$\lim_{x\to\vec{x}_j} \frac{W_x}{c} = 1 + \sum_{i\neq j} \frac{p_n}{p_n - \zeta_i^{\vec{x}_j}},$$

which is indeed the principal's payment when $x = \vec{x}_j$.

Moreover, note that the set of feasible replacement strategies,

$$\left\{(x_1, \cdots, x_n) \,|x_i \geq 0, \forall i = 1, \cdots, n, \sum_{i=1}^{n} x_i \leq 1\right\},$$

is a compact set. Therefore minimum of a continuous function on a compact set exists. □

A.2 The Optimal Pure Replacement Strategy

The next statement characterizes the optimal pure replacement strategy.

Lemma 2. *The optimal pure replacement strategies involve replacing either the front-most or the end-most worker. That is,* $(1, 0, \ldots, 0)$ *and* $(0, \ldots, 0, 1)$ *are the optimal pure replacement strategies.*

Proof of Lemma 2.
With slight abuse of notation, denote the principal's expected compensation cost under the pure replacement strategy in which worker $i \in \{1, \cdots, n\}$ is replaced with probability 1 by W_i and that under the pure replacement strategy in which no worker is replaced by $W_\emptyset$. We will show that:

$$W_1 = W_n = min\{W_\emptyset, W_1, W_2, \cdots, W_n\}.$$

We have:

$$\begin{aligned}
\frac{W_\emptyset}{c} &= \sum_{i=1}^{n} \frac{p_n}{p_n - p_{i-1}} \\
\frac{W_i}{c} &= 1 + \sum_{k=1}^{i-1} \frac{p_n}{p_n - p_{n+k-i}} + \sum_{k=i+1}^{n} \frac{p_n}{p_n - p_{k-1}}, \text{ for each } i \in \{1, \cdots, n\}.
\end{aligned}$$

Then:

$$\frac{W_{i+1} - W_i}{c} = \frac{p_n}{p_n - p_{n-i}} - \frac{p_n}{p_n - p_i}, \text{ for each } i \in \{1, \cdots, n-1\}.$$

Therefore:

$$W_{i+1} \begin{cases} > W_i, & \text{if } i < n/2, \\ = W_i, & \text{if } i = n/2, \\ < W_i, & \text{if } i > n/2, \end{cases} \text{ for each } i \in \{1, \cdots, n-1\}.$$

Note that

$$\begin{aligned}
\frac{W_{n+1-i}}{c} &= 1 + \sum_{k=1}^{n+1-i-1} \frac{p_n}{p_n - p_{n+k-(n+1-i)}} + \sum_{k=n+1-i+1}^{n} \frac{p_n}{p_n - p_{k-1}} \\
&= 1 + \sum_{k=i+1}^{n} \frac{p_n}{p_n - p_{k-1}} + \sum_{k=1}^{i-1} \frac{p_n}{p_n - p_{n+k-i}}. \\
&= \frac{W_i}{c}
\end{aligned}$$

This implies that $W_1 = W_n$. Moreover,

$$\frac{W_1}{c} = 1 + \sum_{i=2}^{n} \frac{p_n}{p_n - p_{i-1}} < \frac{p_n}{p_n - p_0} + \sum_{i=2}^{n} \frac{p_n}{p_n - p_{i-1}} = \frac{W_\emptyset}{c}.$$

This concludes the proof. □

A.3 Possibility of Altering Wage Hierarchy

The next result shows that (suboptimal) AI replacement can alter the wage hierarchy within the production team compared to the case where no replacement occurs.

Lemma 3. *For any $i < j < n$, there exists a replacement strategy x such that $w_i^x > w_j^x$.*

Proof of Lemma 3.
Consider the replacement strategy with $x_i + x_{i+1} = 1$ and $x_i \in (0, 1)$. Then from equation (5) we have $\zeta_i^x = p_{n-1}$ and $\zeta_j^x = p_{j-1}$, which implies $w_i^x > w_j^x$. □

Appendix B Proofs of Results

This appendix section provides the proofs of the results stated in Section 3 and Section 4.

B.1 Proofs of Section 3

Proof of Proposition 1.
The proof proceeds in three steps. We show the following in sequence:

1. Trigger strategy profile σ^* constitutes an equilibrium under w^x.
2. No alternative compensation scheme with a lower expected cost can support σ^* as an equilibrium.
3. Supporting any other strategy profile, apart from the trigger strategy profile, that results in joint effort, as an equilibrium necessarily incurs a higher expected compensation cost for the principal.

Step 1. We will prove that, under the compensation scheme w^x, the trigger strategy is a best response for each worker $i \in \{1, \ldots, n\}$ given that all other workers follow the trigger strategy.

If worker i observes $e_{i-1} = 1$, he believes that nobody has deviated. In this case, choosing effort is a best response for worker i if and only if:

$$p_n\, w_i - c \geq \zeta_i^x\, w_i. \tag{7}$$

The left-hand side is the expected payoff for worker i when he exerts effort, which leads all subsequent workers to exert effort, resulting in a p_n probability of success. The right-hand side is his expected payoff

when he shirks, where ζ_i^x is the probability of success from his perspective given that the other workers follow the trigger strategy profile.

Note that w_i^x is defined so that condition (7), a necessary condition for the trigger strategy to comprise an equilibrium, holds with equality: $w_i^x = \frac{c}{p_n - \zeta_i}$ (equation (2)). Therefore worker i has no incentive to deviate from the trigger strategy after observing $e_{i-1} = 1$.

If worker i observes $e_{i-1} = 0$, we are off the equilibrium path. Worker i then believes that all the predecessors of worker $i-1$ exerted effort and worker $i-1$ is the first to shirk.[11] In this case, shirking ($e_i = 0$) is a best response for worker i if and only if:

$$p_{n-1} w_i^x - c \leq \hat{\zeta}_i^x \, w_i^x. \tag{8}$$

The left-hand side is the expected payoff of worker i when he exerts effort. By exerting effort, worker i ensures that all his successors will also exert effort, resulting in a probability of success of p_{n-1}, as all workers other than $i-1$ have exerted effort. The right-hand side denotes the expected payoff for worker i when he shirks, where $\hat{\zeta}_i^x$ is the probability of success when worker i shirks, given that he has observed $e_{i-1} = 0$ and believes that all workers preceding $i-1$ have exerted effort, and all his successors will adhere to trigger strategies. That is:

$$\begin{aligned}
\hat{\zeta}_i^x &= p_{i-2} \sum_{k=1}^{i-2} \frac{x_k}{1 - x_i - x_{i-1}} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i - x_{i-1}} p_{n-1-k+i} + p_{i-2} \frac{1 - \bar{x}}{1 - x_i - x_{i-1}} \\
&= p_{i-2} \frac{1 - \sum_{k=i-1}^{n} x_k}{1 - x_i - x_{i-1}} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i - x_{i-1}} p_{n-1-k+i}
\end{aligned}$$

More precisely, after observing $e_{i-1} = 0$, worker i knows that neither he nor worker $i-1$ has been replaced. Therefore, the probability of worker k being replaced (for $k \neq i, i-1$) is given by $\frac{x_k}{1 - x_i - x_{i-1}}$, while the probability of no replacement from worker i's perspective is $\frac{1 - \bar{x}}{1 - x_i - x_{i-1}}$. When the replaced worker is a predecessor of worker i (and $i-1$) (i.e., when $k < i-1$), or when there is no AI replacement, all the successors of worker i will shirk upon observing that worker i shirks, resulting in a success probability of p_{i-2}. When the replaced worker is a successor of worker i (i.e., when $k > i$), all the workers between i and k will shirk, and all the successors of k will exert effort, resulting in a success probability of $p_{n-1-k+i}$.

We already know that $w_i^x(p_n - \zeta_i^x) = c$. Therefore showing that $p_n - \zeta_i^x \geq p_{n-1} - \hat{\zeta}_i^x$ would suffice to show that condition (8) holds, and hence each worker finds it optimal to shirk upon observing that his immediate predecessor shirks. Therefore, we would like to show that

$$p_n - p_{n-1} \geq \zeta_i^x - \hat{\zeta}_i^x$$

[11] All subsequent arguments would analogously hold even if worker i has different out-of-path beliefs regarding the behavior of his predecessors.

Note that

$$\zeta_i^x - \hat{\zeta}_i^x = \underbrace{\left(p_{i-1} \frac{1 - \sum_{k=i}^{n} x_k}{1 - x_i} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i} p_{n-k+i} \right)}_{\zeta_i^x} - \underbrace{\left(p_{i-2} \frac{1 - \sum_{k=i-1}^{n} x_k}{1 - x_i - x_{i-1}} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i - x_{i-1}} p_{n-1-k+i} \right)}_{\hat{\zeta}_i^x}$$

After some algebra, we get:

$$\zeta_i^x - \hat{\zeta}_i^x = (p_{i-1} - p_{i-2}) \frac{1 - \sum_{k=i-1}^{n} x_k}{1 - x_i - x_{i-1}} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i} (p_{n-k+i} - p_{n-1-k+i}) + p_{i-1} \frac{x_{i-1} \sum_{k=i+1}^{n} x_k}{(1 - x_i)(1 - x_i - x_{i-1})} - \sum_{k=i+1}^{n} \frac{x_{i-1} x_k}{(1 - x_i)(1 - x_i - x_{i-1})} p_{n-1-k+i}.$$

Simplifying further, we get:

$$\zeta_i^x - \hat{\zeta}_i = (p_{i-1} - p_{i-2}) \frac{1 - \sum_{k=i-1}^{n} x_k}{1 - x_i - x_{i-1}} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i} (p_{n-k+i} - p_{n-1-k+i}) + \sum_{k=i+1}^{n} \frac{x_{i-1} x_k}{(1 - x_i)(1 - x_i - x_{i-1})} (p_{i-1} - p_{n-1-k+i}).$$

However, the term on the second line of this equality is negative as $p_{i-1} \leq p_{n-1-k+i}$ for each $k \geq i+1$. Then we can write

$$\zeta_i^x - \hat{\zeta}_i^x \leq (p_{i-1} - p_{i-2}) \frac{1 - \sum_{k=i-1}^{n} x_k}{1 - x_i - x_{i-1}} + \sum_{k=i+1}^{n} \frac{x_k}{1 - x_i} (p_{n-k+i} - p_{n-1-k+i}).$$

The right hand side of this inequality is a weighted average of incremental increase in the probabilities, and the sum of the weights is less than 1. Then by using the fact that $p_n - p_{n-1}$ is the largest incremental increase in the probability, we get:

$$\zeta_i^x - \hat{\zeta}_i^x \leq p_n - p_{n-1}.$$

Therefore, condition (8) is satisfied, and worker i finds it optimal to shirk upon observing $e_{i-1} = 0$.

Therefore, the trigger strategy profile comprises an equilibrium under compensation scheme w^x.

Step 2. Moreover, we know that w^x is the least costly compensation scheme to support the trigger strategy profile as an equilibrium. This is because w^x satisfies condition (7), a necessary condition to support the trigger strategy as an equilibrium, with equality. Any compensation scheme with a lower expected cost to the principal would violate this condition.

Step 3. Finally, in any alternative equilibrium strategy profile with joint effort as the equilibrium outcome, the probability of success following worker i's deviation (from his perspective) is (weakly) higher than that under the trigger strategy profile, ζ_i^x. This imposes a (weakly) stronger version of the above constraint and thus requires a (weakly) higher payment to each worker. Consequently, w^x must also be the optimal compensation scheme conditional on replacement strategy x. □

Proof of Proposition 2.

Consider a replacement strategy x of the form:

$$x = (\rho, 0, \ldots, 0, 1-\rho), \quad \text{for } \rho \in [0,1].$$

The two extreme points within this class, which are obtained by setting $\rho = 0$ and $\rho = 1$, correspond to the optimal pure replacement strategies. We will show that choosing any interior value $\rho \in (0,1)$ outperforms these pure strategies. Therefore, the optimal replacement strategy necessarily involves randomization.

With a slight abuse of notation, for $x = (\rho, 0, \ldots, 0, 1-\rho)$, we denote ζ_i^x as ζ_i^ρ, w_i^x as w_i^ρ, and W_x as W_ρ. It is clear that:

$$\begin{aligned}
&\zeta_1^\rho = p_1, \\
&\zeta_n^\rho = p_{n-1}, \\
&\zeta_i^\rho = (1-\rho)\, p_i + \rho\, p_{i-1}, \ \forall i \in \{2, \ldots, n-1\}.
\end{aligned}$$

Moreover,

$$\frac{W_\rho}{c} = \sum_{i=2}^{n-1} \frac{p_n}{p_n - (1-\rho)\, p_i - \rho p_{i-1}} + \frac{p_n}{p_n - p_1}(1-\rho) + \rho \frac{p_n}{p_n - p_{n-1}} + 1.$$

Note this holds even when $\rho = 0$ or $\rho = 1$. The second order derivative of the W_ρ with respect to ρ satisfies:

$$\frac{\partial^2 W_\rho}{\partial \rho^2} \frac{1}{c} = 2 p_n \left(\sum_{i=2}^{n-1} \frac{(p_i - p_{i-1})^2}{(p_n - (1-\rho)\, p_i - \rho p_{i-1})^3} \right).$$

With $n \geq 3$, the summation contains at least one strictly positive term, as $(p_i - p_{i-1})^2 > 0$ and $(p_n - (1-\rho)\, p_i - \rho p_{i-1})^3 > 0$. Therefore:

$$\frac{W_\rho}{\partial \rho^2} > 0,$$

implying that W_ρ is strictly convex in ρ. But we know that $\rho = 0$ and $\rho = 1$ corresponds to the optimal pure replacement strategies. That is, $W_\rho\big|_{\rho=0} = W_\rho\big|_{\rho=1}$. Then, from the convexity of W_ρ, there must exist a $\rho \in (0,1)$ minimizing W_ρ. This implies that, a replacement strategy with some interior ρ is better than the optimal pure replacement strategies. Therefore, the optimal scheme must involve randomization. □

Proof of Proposition 3.

Step 1. First we prove a stronger statement that implies $x_2^* = 0$ when $n = 3$.

Claim. *In the optimal replacement strategy,* $x_{n-1}^* = 0$.

Proof of the Claim. Let $x = (x_1, \cdots, x_n)$ be an optimal replacement strategy, and suppose that $x_{n-1} > 0$ to get a contradiction. Then we must also have $x_{n-1} < 1$ as the optimal replacement strategy involves randomization (Proposition 2).

Recall that the expected compensation cost for the principal satisfies:

$$\frac{W_x}{c} = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{(1-x_i)^2 p_n}{R_i},$$

where

$$R_i = (1-x_i)(p_n - p_{i-1}) - \sum_{k=i+1}^{n} x_k (p_{n+i-k} - p_{i-1}).$$

Now, given this optimal replacement strategy with $x_{n-1} > 0$, consider a deviation in which the probability of replacement for the $(n-1)$th worker is transferred to the nth worker. That is, the new replacement strategy x' obtained by this deviation from x is:

$$x' = (x_1, \cdots, x_{n-2}, 0, x_{n-1} + x_n).$$

Case 1: First consider the case where $x_{n-1} + x_n < 1$. The new expected cost is:

$$\frac{W_{x'}}{c} = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n-2} \frac{(1-x_i)^2 p_n}{R_i'} + \frac{p_n}{R_{n-1}'} + \frac{(1-x_{n-1}-x_n)^2 p_n}{R_n'},$$

where

$$\begin{aligned}
R_i' &= R_i + x_{n-1}(p_{i+1} - p_i), \text{ if } i < n-1, \\
R_{n-1}' &= R_{n-1} + x_{n-1}(p_n - p_{n-1}), \\
R_n &= (1-x_n)(p_n - p_{n-1}), \\
R_n' &= (1-x_{n-1}-x_n)(p_n - p_{n-1}).
\end{aligned}$$

We would like to show that $W_{x'} < W_x$ holds. Given $R_i < R'_i$ if $i < n-1$, it suffices to show that:

$$\frac{(1-x_n)^2}{R_n} - \frac{(1-x_{n-1}-x_n)^2}{R'_n} + \frac{(1-x_{n-1})^2}{R_{n-1}} - \frac{1}{R'_{n-1}} \geq 0. \tag{9}$$

The first two terms of (9) can be rewritten as:

$$\begin{aligned}\frac{(1-x_n)^2}{R_n} - \frac{(1-x_{n-1}-x_n)^2}{R'_n} =& \frac{(1-x_n)^2}{(1-x_n)(p_n-p_{n-1})} - \frac{(1-x_{n-1}-x_n)^2}{(1-x_{n-1}-x_n)(p_n-p_{n-1})} \\ =& \frac{1-x_n}{p_n-p_{n-1}} - \frac{1-x_{n-1}-x_n}{p_n-p_{n-1}} \\ =& \frac{x_{n-1}}{p_n-p_{n-1}}\end{aligned} \tag{10}$$

Moreover, by using the equality $R_{n-1} = R'_{n-1} - x_{n-1}(p_n - p_{n-1})$, we can rewrite the third and fourth terms of (9) as:

$$\begin{aligned}\frac{(1-x_{n-1})^2}{R_{n-1}} - \frac{1}{R'_{n-1}} &= \frac{R'_{n-1}(1-x_{n-1})^2 - R'_{n-1} + x_{n-1}(p_n-p_{n-1})}{R_{n-1}R'_{n-1}} \\ &= \frac{x_{n-1}\left[(p_n-p_{n-1}) - R'_{n-1}(2-x_{n-1})\right]}{R_{n-1}R'_{n-1}}.\end{aligned}$$

Therefore, inequality (9) holds if and only if

$$\frac{1}{p_n-p_{n-1}} + \frac{(p_n-p_{n-1}) - R'_{n-1}(2-x_{n-1})}{R_{n-1}R'_{n-1}} \geq 0.$$

This inequality holds if and only if

$$(p_n-p_{n-1})^2 - R'_{n-1}(2-x_{n-1})(p_n-p_{n-1}) + R_{n-1}R'_{n-1} \geq 0,$$

which further simplifies to

$$(p_n-p_{n-1})^2 - R'_{n-1}(2-x_{n-1})(p_n-p_{n-1}) + \left[R'_{n-1} - x_{n-1}(p_n-p_{n-1})\right]R'_{n-1} \geq 0.$$

This can be rewritten as

$$(p_n-p_{n-1})^2 - 2(p_n-p_{n-1})R'_{n-1} + (R'_{n-1})^2 \geq 0,$$

which is equivalent to

$$\left(p_n - p_{n-1} - R'_{n-1}\right)^2 \geq 0.$$

Since the square of a real number is always non-negative, the inequality holds.

This implies that the deviation is profitable for the principal, which contradicts the optimality of the allocation $x = (x_1, \ldots, x_{n-1}, x_n)$. Therefore, the optimal replacement strategy x^* has to satisfy $x^*_{n-1} = 0$.

Case 2: Next consider the case where $x_{n-1} + x_n = 1$. The first two terms in (9) becomes $\frac{(1-x_n)^2}{R_n}$ in this case. Moreover, note that:

$$\frac{(1-x_n)^2}{R_n} = \frac{(1-x_n)^2}{(1-x_n)(p_n - p_{n-1})} = \frac{1-x_n}{p_n - p_{n-1}} = \frac{x_{n-1}}{p_n - p_{n-1}}$$

which coincides with (10). Therefore, the optimal replacement strategy x^* has to satisfy $x^*_{n-1} = 0$. This completes the proof of the claim.

Step 2. Next, we prove that the optimal replacement strategy x^* satisfies $x^*_3 > 0$. Suppose, for contradiction, that a replacement strategy $x = (x_1, x_2, x_3)$ with $x_3 = 0$ is optimal. Note that, the partial derivative of the principal's expected cost with respect to x_1 (as long as $x_1 < 1$) is:

$$\begin{aligned}
\frac{1}{c}\frac{\partial W_x}{\partial x_1} &= 1 - p_3\frac{w_1}{c} + p_3(1-x_1)\frac{1}{c}\frac{\partial w_1}{\partial x_1} \\
&= 1 - \frac{p_3}{p_3 - \zeta_1} + p_3(1-x_1)\frac{1}{(p_3-\zeta_1)^2}\frac{\partial \zeta_1}{\partial x_1} \\
&= -\frac{\zeta_1}{p_3 - \zeta_1} + p_3\frac{1}{(p_3-\zeta_1)^2}\frac{\sum_{k=2}^{3} x_k(p_{4-k} - p_0)}{1-x_1} \\
&= -\frac{\zeta_1}{p_3 - \zeta_1} + \frac{p_3(\zeta_1 - p_0)}{(p_3-\zeta_1)^2} \\
&= \frac{\zeta_1^2 - p_3 p_0}{(p_3-\zeta_1)^2}. \qquad (11)
\end{aligned}$$

From the previous step, we know that $x_2 = 0$. Then $x_2 = x_3 = 0$ implies that $\zeta_1 = p_0$. By substituting this into (11), we get $\frac{\partial W_x}{\partial x_1} \leq 0$. This, in turn, implies that $x_1, x_2, x_3) = (1, 0, 0)$ is an optimal replacement strategy, which contradicts Proposition 2. Therefore, the optimal replacement strategy x^* satisfies $x^*_3 > 0$.

Step 3. Now we prove that the optimal replacement strategy x^* satisfies $x^*_3 \geq x^*_1$. Suppose not for contradiction. Then, a replacement strategy $x = (x_1, 0, x_3)$ with $x_1 > x_3$ is optimal. This in turn requires $x_1 \in (0, 1)$.

The expected compensation cost of the principal under the scheme $(x_1, 0, x_3)$ is:

$$c\left(x_1 + x_3 + p_3\left[\frac{(1-x_1)^2}{(1-x_1)(p_3-p_0) - x_3(p_1-p_0)} + \frac{1}{(p_3-p_1) - x_3(p_2-p_1)} + \frac{1-x_3}{p_3-p_2}\right]\right).$$

The expected compensation cost of the principal under the scheme $(x_3, 0, x_1)$ is:

$$c\left(x_1 + x_3 + p_3\left[\frac{(1-x_3)^2}{(1-x_3)(p_3-p_0) - x_1(p_1-p_0)} + \frac{1}{(p_3-p_1) - x_1(p_2-p_1)} + \frac{1-x_1}{p_3-p_2}\right]\right).$$

In the subsequent analysis, we will show that the precedent is strictly larger than the latter, i.e., we will show that:

$$\left[\frac{(1-x_1)^2}{(1-x_1)(p_3-p_0) - x_3(p_1-p_0)} + \frac{1}{(p_3-p_1) - x_3(p_2-p_1)} + \frac{1-x_3}{p_3-p_2}\right]$$

$$-\left[\frac{(1-x_3)^2}{(1-x_3)(p_3-p_0)-x_1(p_1-p_0)}+\frac{1}{(p_3-p_1)-x_1(p_2-p_1)}+\frac{1-x_1}{p_3-p_2}\right]>0. \tag{12}$$

This, in turn, implies that the principal is better off under the allocation $(x_3, 0, x_1)$, which contradicts the optimality of $(x_1, 0, x_3)$.

To this end, note first that:

$$\frac{1-x_3}{p_3-p_2}-\frac{1-x_1}{p_3-p_2}=\frac{x_1-x_3}{p_3-p_2}.$$

Moreover, we have:

$$\frac{1}{(p_3-p_1)-x_3(p_2-p_1)}-\frac{1}{(p_3-p_1)-x_1(p_2-p_1)}=\frac{-(p_2-p_1)(x_1-x_3)}{[(p_3-p_1)-x_3(p_2-p_1)][(p_3-p_1)-x_1(p_2-p_1)]}.$$

Claim. *The following inequality holds:*

$$\frac{-(p_2-p_1)(x_1-x_3)}{[(p_3-p_1)-x_3(p_2-p_1)][(p_3-p_1)-x_1(p_2-p_1)]}>\frac{-(p_2-p_1)(x_1-x_3)}{(p_3-p_1)(p_3-p_2)}.$$

Proof of the Claim. The inequality holds if and only if:

$$[(p_3-p_1)-x_3(p_2-p_1)][(p_3-p_1)-x_1(p_2-p_1)]>(p_3-p_1)(p_3-p_2).$$

This is equivalent to:

$$(p_3-p_1)^2-x_1(p_2-p_1)(p_3-p_1)-x_3(p_2-p_1)(p_3-p_1)+x_1x_3(p_2-p_1)^2-(p_3-p_1)(p_3-p_2)>0,$$

which can be rewritten as:

$$(p_3-p_1)(p_2-p_1)-x_1(p_2-p_1)(p_3-p_1)-x_3(p_2-p_1)(p_3-p_1)+x_1x_3(p_2-p_1)^2>0.$$

Rearranging further, we obtain:

$$(p_3-p_1)(1-x_1-x_3)+x_1x_3(p_2-p_1)>0,$$

which is correct because $1-x_1-x_3\geq 0$, $x_1>0$, and $x_3>0$. □

Getting back to the inequality (12), note that:

$$\frac{x_1-x_3}{p_3-p_2}-\frac{(p_2-p_1)(x_1-x_3)}{(p_3-p_1)(p_3-p_2)}=\frac{x_1-x_3}{p_3-p_1}.$$

Define:

$$\frac{(1-x_1)^2}{(1-x_1)(p_3-p_0)-x_3(p_1-p_0)}-\frac{(1-x_3)^2}{(1-x_3)(p_3-p_0)-x_1(p_1-p_0)}=\frac{\Omega}{\Theta},$$

where

$$\Theta = [(1-x_1)(p_3-p_0) - x_3(p_1-p_0)]\,[(1-x_3)(p_3-p_0) - x_1(p_1-p_0)]$$

$$\begin{aligned}\Omega &= (p_3-p_0)\left[(1-x_1)^2(1-x_3) - (1-x_3)^2(1-x_1)\right] + (p_1-p_0)\left[(1-x_3)^2 x_3 - (1-x_1)^2 x_1\right] \\ &= -(x_1-x_3)\left[(p_3-p_0)(1-x_1)(1-x_3) + (p_1-p_0)\left(x_1^2 + x_1x_3 + x_3^2 - 2x_1 - 2x_3 + 1\right)\right].\end{aligned}$$

Also denote

$$\Lambda \equiv -\frac{\Omega}{x_1 - x_3}.$$

To show that inequality (12) holds, it suffices to show:

$$\frac{x_1 - x_3}{p_3 - p_1} - \frac{(x_1 - x_3)\,\Lambda}{\Theta} \geq 0.$$

This is equivalent to:

$$\Theta - (p_3 - p_1)\,\Lambda \geq 0.$$

By rearranging it, we obtain:

$$\begin{aligned}&\underbrace{(1-x_1)(1-x_3)(p_3-p_0)^2}_{S_1} - \underbrace{(p_3-p_1)(p_3-p_0)(1-x_1)(1-x_3)}_{S_2} \\ &- \underbrace{(1-x_1)x_1(p_3-p_0)(p_1-p_0)}_{S_3} - \underbrace{x_3(1-x_3)(p_1-p_0)(p_3-p_0)}_{S_4} \\ &- \underbrace{(p_3-p_1)(p_1-p_0)\left(x_1^2 + x_1x_3 + x_3^2 - 2x_1 - 2x_3 + 1\right)}_{S_5} + \underbrace{x_1x_3(p_1-p_0)^2}_{S_6} \geq 0.\end{aligned}$$

However, note that:

$$S_1 - S_2 = (p_1-p_0)(p_3-p_0)(1-x_1)(1-x_3).$$

Thus, we obtain:

$$\underbrace{(p_1-p_0)(p_3-p_0)(1-x_1)(1-x_3)}_{S_1-S_2} - S_3 - S_4 = (p_3-p_0)(p_1-p_0)\left(x_1^2 + x_1x_3 + x_3^2 - 2x_1 - 2x_3 + 1\right).$$

Therefore, we further derive:

$$\underbrace{(p_3-p_0)(p_1-p_0)\left(x_1^2 + x_1x_3 + x_3^2 - 2x_1 - 2x_3 + 1\right)}_{S_1-S_2-S_3-S_4} - S_5 + S_6 = (1-x_1-x_3)^2(p_1-p_0)^2 \geq 0.$$

Thus, inequality (12) holds, which contradicts with the optimality $(x_3, 0, x_1)$. Therefore, the optimal replacement strategy x^* satisfies $x_3^* \geq x_1^*$. □

Proof of Proposition 4.

We prove Proposition 4 in two steps:

1. If $p_1^2 - p_3 p_0 \leq 0$, the principal fully utilizes the AI capacity, i.e., $\bar{x}^* = 1$.
2. If $p_1^2 - p_3 p_0 > 0$, the principal does not fully utilize the AI capacity, i.e., $\bar{x}^* < 1$.

Step 1. Assume that $p_1^2 - p_3 p_0 \leq 0$. Suppose, for contradiction, that the principal does not fully utilize the AI capacity under the optimal replacement strategy, i.e., a strategy x with $\bar{x} < 1$ is optimal. Note that

$$\zeta_1 = p_0 + \sum_{k=2}^{3} \frac{x_k}{1 - x_1} (p_{4-k} - p_0).$$

As we know from Proposition 3 that $x_2 = 0$, we obtain:

$$\zeta_1 = p_0 + \frac{x_3}{1 - x_1} (p_1 - p_0).$$

Since $\bar{x} < 1$, we must have $x_3 < 1 - x_1$, so $\zeta_1 < p_1$. Therefore, we must have:

$$\zeta_1^2 - p_3 p_0 < p_1^2 - p_3 p_0 \leq 0.$$

Then, from equation (11), we know that:

$$\frac{1}{c} \frac{\partial W_x}{\partial x_1} = \frac{\zeta_1^2 - p_3 p_0}{(p_3 - \zeta_1)^2} \quad \Longrightarrow \quad \frac{\partial W_x}{\partial x_1} < 0.$$

Therefore, increasing x_1 slightly, which is feasible since $x_1 + x_3 < 1$, leads to an improvement in terms of the expected cost of compensation. This contradicts the optimality of x. Hence, under the optimal replacement strategy x^*, the principal fully utilizes the AI capacity, i.e., $\bar{x}^* = 1$.

Step 2. Assume $p_1^2 - p_3 p_0 > 0$. Suppose, for contradiction, that the principal fully utilizes the AI capacity under the optimal replacement strategy, i.e., a replacement strategy x with $\bar{x} = 1$ is optimal. Then we must have $x_1 + x_3 = 1$, since Proposition 3 shows that $x_2 = 0$. Note that, $x_3 = 1 - x_1$ implies:

$$\zeta_1 = p_0 + \frac{x_3}{1 - x_1} (p_1 - p_0) = p_1.$$

By substituting $\zeta_1 = p_1$, we obtain:

$$\zeta_1^2 - p_3 p_0 = p_1^2 - p_3 p_0 > 0.$$

Then, from equation (11), we know that:

$$\frac{1}{c} \frac{\partial W_x}{\partial x_1} = \frac{\zeta_1^2 - p_3 p_0}{(p_3 - \zeta_1)^2} \quad \Longrightarrow \quad \frac{\partial W_x}{\partial x_1} > 0.$$

This implies that decreasing x_1 slightly, which is feasible as $x_1 > 0$, leads to an improvement in terms

of the expected compensation cost. This contradicts the optimality of x. Hence, under the optimal replacement strategy x^*, the principal under-utilizes the AI capacity, i.e., $\bar{x}^* < 1$, in this case. □

Proof of Proposition 5.

Recall that:

$$\begin{aligned}
\zeta_1^x &= p_2 \frac{x_2}{1-x_1} + p_1 \frac{x_3}{1-x_1} + p_0 \frac{1-x_1-x_2-x_3}{1-x_1}, \\
\zeta_2^x &= p_2 \frac{x_3}{1-x_2} + p_1 \frac{1-x_2-x_3}{1-x_2}, \\
\zeta_3^x &= p_2.
\end{aligned}$$

1. From earlier results, we know that $x_3^* \in (0,1)$, and $x_2^* = 0$. First, $x_2^* = 0$ implies $\zeta_1^{x^*} \leq p_1$. Next, $x_3^* \in (0,1)$ implies $\zeta_2^{x^*} \in (p_1, p_2)$. In consequence, we have $\zeta_3^{x^*} > \zeta_2^{x^*} > \zeta_1^{x^*}$, which implies that the wages are increasing in the position after the optimal automation: $w_3^{x^*} > w_2^{x^*} > w_1^{x^*}$.

2. As $x_3^* > 0$, we have $\zeta_1^{x^*} > p_0 = \zeta_1^0$. This implies that the wage of the front-end worker increases after the automation: $w_1^{x^*} > w_1^0$.

3. $x_3^* > 0$ also implies that $\zeta_2^{x^*} > p_1 = \zeta_2^0$. Therefore, the wage of the middle worker also increases after the automation: $w_2^{x^*} > w_2^0$.

4. Finally, note that, regardless of the replacement strategy, we have: $\zeta_3^{x^*} = p_2$. Therefore, the wage of the end-most worker remains the same after the automation: $w_3^{x^*} = w_3^0$.

□

Proof of Proposition 6.

From Proposition 5, we know that after AI replacement, the wages of the front-most and middle workers increase, while the wage of the end-most worker remains unchanged. Additionally, the end-most worker faces a risk of replacement, whereas the middle worker does not. Consequently, the middle worker's payoff increases, while the end-most worker's payoff decreases relative to their payoffs prior to AI replacement. The front-most worker's payoff depends on his replacement probability—while his wage increases, his payoff may rise or fall. □

B.2 Proofs of Section 4.1

Proof of Proposition 7.

Step 1. First, we will derive the closed-form solution for the optimal replacement strategy. Note that this functional form satisfies the necessary and sufficient condition for full utilization given in Proposition 4, ensuring $\bar{x}^* = 1$. Moreover, Proposition 3 establishes that $x_2^* = 0$. Therefore, we have $x_1^* + x_3^* = 1$.

As the optimal replacement strategy must involve randomization (Proposition 2), we must have $x_1^*, x_3^* \in (0,1)$. Consider an arbitrary such strategy given by $x = (x_1, x_2, x_3) = (\rho, 0, 1-\rho)$ for some $\rho \in (0,1)$. Denoting, with a slight abuse of notation, the principal's expected compensation cost under this strategy by W_ρ, we obtain:

$$\frac{W_\rho}{c} = 1 + \frac{p_3}{p_3 - p_1}(1-\rho) + \frac{p_3}{p_3 - (1-\rho)p_2 - \rho p_1} + \rho\frac{p_3}{p_3 - p_2}.$$

Under the optimal replacement strategy, ρ must satisfy:

$$\frac{1}{c}\frac{\partial W_\rho}{\partial \rho} = -\frac{p_3(p_2 - p_1)}{\left(p_3 - (1-\rho)p_2 - \rho p_1\right)^2} - \frac{p_3}{p_3 - p_1} + \frac{p_3}{p_3 - p_2} = 0.$$

Rearranging, this simplifies to:

$$\frac{p_2 - p_1}{\left(p_3 - (1-\rho)p_2 - \rho p_1\right)^2} = \frac{1}{p_3 - p_2} - \frac{1}{p_3 - p_1} = \frac{p_2 - p_1}{(p_3 - p_2)(p_3 - p_1)}.$$

Solving for ρ yields:

$$p_3 - (1-\rho)p_2 - \rho p_1 = \sqrt{(p_3 - p_2)(p_3 - p_1)}.$$

From which we obtain:

$$\rho = \frac{\sqrt{p_3 - p_2}\left(\sqrt{p_3 - p_1} - \sqrt{p_3 - p_2}\right)}{p_2 - p_1} = \frac{\sqrt{1-\alpha}\left(\sqrt{1-\alpha^2} - \sqrt{1-\alpha}\right)}{\alpha - \alpha^2} = \frac{\sqrt{1+\alpha} - 1}{\alpha}.$$

This provides the closed-form solution for the optimal replacement strategy as stated.

Step 2. Next, we will show the comparative statics results. From the previous step, for a given α, the optimal replacement probability of the front-most worker satisfies:

$$x_1^* = \frac{\sqrt{1+\alpha} - 1}{\alpha}.$$

Taking the derivative of x_1^* with respect to α, we obtain:

$$\frac{\partial x_1^*}{\partial \alpha} = \frac{\frac{\alpha}{2\sqrt{1+\alpha}} - \sqrt{1+\alpha} + 1}{\alpha^2}.$$

The numerator of this expression is a decreasing function of α as:

$$\frac{\partial\left(\frac{\alpha}{2\sqrt{1+\alpha}} - \sqrt{1+\alpha} + 1\right)}{\partial \alpha} = \frac{\sqrt{1+\alpha} - \frac{\alpha}{2}\frac{1}{\sqrt{1+\alpha}}}{1+\alpha} - \frac{1}{\sqrt{1+\alpha}} = -\frac{\alpha}{2(1+\alpha)^{3/2}} < 0.$$

Moreover, the numerator evaluates to zero at $\alpha = 0$, which implies that it remains strictly negative for all $\alpha \in (0,1)$. Thus, we obtain:

$$\frac{\partial x_1^*}{\partial \alpha} < 0,$$

proving that as the degree of complementarity increases (α becomes smaller), the likelihood of the front-most (end-most) worker being replaced increases (decreases). □

B.3 Proofs of Section 4.2

Proof of Proposition 8.

By applying analogous arguments from the proof of Proposition 1, we can establish that under a given replacement strategy x: (i) the optimal compensation scheme sustains effort by inducing a trigger strategy equilibrium, and (ii) workers remain indifferent between exerting effort and shirking along the equilibrium path. For conciseness, we do not replicate these arguments here. Denote by ζ_i^x the probability of project success when worker i shirks, assuming all other workers follow the trigger strategy profile. Worker i's incentive constraint then becomes:

$$p_n(1-x_i)w_i - (1-x_i)c \geq \zeta_i^x w_i.$$

Under the optimal compensation scheme, this constraint binds, which yields:

$$w_i^x = \frac{c}{p_n - \zeta_i^x}.$$

To understand the expression ζ_i^x in the proposition statement, consider what happens when worker i chooses to shirk. Only a fraction $1-x_i$ of his tasks are unperformed. The overall project success rate then depends on how this shirking decision affects downstream behavior through peer monitoring.

- With probability x_i, the AI completes worker i's tasks, so the worker-AI unit contributes to the project. In this case, the project succeeds with probability p_n.
- With probability $(1-x_i)x_{i+1}$, worker i's shirking is noticed by worker $i+1$, who responds by shirking. However, the AI assigned to worker $i+1$ completes his share of tasks, so the unit at $i+1$ still contributes to the project. The project thus succeeds with probability p_{n-1}.
- With probability $(1-x_i)(1-x_{i+1})x_{i+2}$, shirking cascades further. Worker $i+1$ shirks, and $1-x_{i+1}$ fraction of the tasks performed by AI doesn't fully compensate, and worker $i+2$ observes this and also shirks. Yet, the AI at $i+2$ performs x_{i+2} of the tasks, allowing partial contribution. The project succeeds with probability p_{n-2}.
- This process continues recursively, with each successive worker potentially shirking in response, depending on how much of the previous worker's task is conducted by AI. Eventually, the entire downstream chain might shirk, in which case no further contribution comes from any of them.
- In the worst case, where all successors also shirk and none of their AI compensates, the project succeeds only with probability p_{i-1}, which reflects the effort of workers before position i.

Summing over all these possible outcomes gives us:

$$\zeta_i^x = x_i p_n + \sum_{k=1}^{n-i} \left[p_{n-k}\, x_{i+k} \prod_{j=i}^{i+k-1} (1 - x_j) \right] + p_{i-1} \prod_{j=i}^{n} (1 - x_j)\,.$$

This expression represents the expected project success probability when worker i shirks, accounting for the compensating effects of AI task completion and the induced monitoring response throughout the network. □

Proof of Proposition 9.

For the case where $x_i < 1$ for each worker $i \in \{1, \ldots, n\}$, we can write:

$$p_n - \zeta_i^x = (1 - x_i)\,(p_n - \Psi_i^x)$$

where

$$\Psi_i^x = x_{i+1}\, p_{n-1} + \sum_{k=1}^{n-i-1} \left[x_{i+1+k}\, p_{n-1-k} \prod_{j=i+1}^{i+k} (1 - x_j) \right] + p_{i-1} \prod_{j=i+1}^{n} (1 - x_j)\,.$$

Moreover, we can rewrite the principal's expected cost as:

$$\begin{aligned} \sum_{i=1}^{n} [x_i c + p_n(1 - x_i) w_i] &= \sum_{i=1}^{n} \left[x_i c + p_n(1 - x_i) \frac{c}{p_n - \zeta_i^x} \right] \\ &= \sum_{i=1}^{n} \left[x_i c + p_n \frac{c}{p_n - \Psi_i^x} \right]. \end{aligned}$$

First, note that Ψ_i^x depends on x_ℓ only if $\ell > i$. Moreover, the partial derivative $\frac{\partial \Psi_i^x}{\partial x_\ell}$ is proportional to the difference between $p_{n+i-\ell}$ and a weighted average of $p_{n+i-\ell-1}, \ldots, p_{i-1}$, and is therefore positive. Now consider increasing x_i. The principal's expected compensation cost increases due to the following strictly increasing components:

- the direct term $x_i c$, and
- the indirect terms of the form $p_n \frac{c}{p_n - \Psi_\ell^x}$ for $\ell < i$, each of which increases because Ψ_ℓ^x is strictly increasing in x_i.

Therefore, increasing x_i strictly increases the expected cost. It follows that setting all $x_i = 0$ minimizes the expected cost over all strategies satisfying $x_i < 1$ for each worker $i \in \{1, \ldots, n\}$. This proves that the strategy $(0, \ldots, 0)$ is optimal among such replacement strategies. Lemma 2 completes the proof. □

B.4 Proofs of Section 4.3

Proof of Proposition 10.

First, similar reasoning to that in Proposition 1 establishes, for a given replacement strategy, the optimal compensation scheme, and that this scheme induces a trigger strategy as an equilibrium. Moreover, analogous arguments to those in Lemma 1 ensure the existence of an optimal replacement strategy in this star network structure. To prove the proposition, we first state and prove two auxiliary claims.

Claim 1. *Given a fixed x_n, there exists a constrained optimal replacement strategy in which*

$$x_1 = x_2 = \cdots = x_{n-1}.$$

Proof of Claim 1.
– If $x_n = 1$, the claim trivially holds as we must have $x_1 = \cdots = x_{n-1} = 0$ in this case.
– If $x_n = 0$, then $\zeta_i^x = p_{n-2}$, for each $i \in \{1, \cdots, n-1\}$. The expected compensation cost of the principal under an arbitrary replacement strategy $(x_1, x_2, \cdots, x_{n-1}, 0)$ is:

$$\sum_{i=1}^{n-1} \left[x_i c + (1 - x_i)\, p_n \frac{c}{p_n - p_{n-2}} \right] + p_n \frac{c}{p_n - p_{n-1}}.$$

Consider modifying this replacement strategy into $(\tilde{x}, \cdots, \tilde{x}, 0)$ with:

$$\tilde{x} = \frac{x_1 + \cdots + x_{n-1}}{n-1}.$$

The expected compensation cost of the principal under this replacement strategy is:

$$(n-1) \left[\tilde{x} c + (1 - \tilde{x})\, p_n \frac{c}{p_n - p_{n-2}} \right] + p_n \frac{c}{p_n - p_{n-1}}.$$

Note that, this is equal to the expected compensation cost prior to modification. This implies that there exists a constrained optimal replacement strategy satisfying $x_1 = \cdots = x_{n-1}$.
– If $x_n \in (0, 1)$, we first show that $(1 - x_i)\, p_n \frac{w_i^x}{c}$ is a strictly convex function of x_i. Note that:

$$\begin{aligned}
\frac{\partial \left((1 - x_i)\, p_n \frac{w_i^x}{c} \right)}{\partial x_i} &= p_n \left(-\frac{w_i^x}{c} + \frac{1}{c} (1 - x_i) \frac{\partial w_i^x}{\partial x_i} \right) \\
&= p_n \left(-\frac{1}{p_n - \zeta_i^x} + \frac{1 - x_i}{(p_n - \zeta_i^x)^2} \frac{\partial \zeta_i^x}{\partial x_i} \right) \\
&= p_n \left(-\frac{1}{p_n - \zeta_i^x} + \frac{1}{(p_n - \zeta_i^x)^2} \frac{x_n (p_{n-1} - p_{n-2})}{1 - x_i} \right) \\
&= p_n \left(-\frac{1}{p_n - \zeta_i^x} + \frac{\zeta_i^x - p_{n-2}}{(p_n - \zeta_i^x)^2} \right) \\
&= p_n \left(\frac{2\zeta_i^x - p_n - p_{n-2}}{(p_n - \zeta_i^x)^2} \right).
\end{aligned}$$

Taking the second order derivative, we obtain:

$$\begin{aligned}\frac{\partial^2\left(\left(1-x_i\right)p_n\frac{w_i^x}{c}\right)}{\partial x_i^2} &= p_n\frac{\partial\zeta_i^x}{\partial x_i}\frac{2\left(p_n-\zeta_i^x\right)^2+2\left(p_n-\zeta_i^x\right)\left(2\zeta_i^x-p_n-p_{n-2}\right)}{\left(p_n-\zeta_i^x\right)^4}\\ &= p_n\frac{\partial\zeta_i^x}{\partial x_i}\frac{2\left(p_n-\zeta_i^x\right)\left(\zeta_i^x-p_{n-2}\right)}{\left(p_n-\zeta_i^x\right)^4}.\end{aligned}$$

But, $x_n>0$ implies $\zeta_i^x-p_{n-2}>0$ and $\frac{\partial\zeta_i^x}{\partial x_i}>0$. Also, note that $p_n-\zeta_i^x>0$. Therefore, the second order derivative of $\left(1-x_i\right)p_n\frac{w_i^x}{c}$ with respect to x_i is positive, implying that $\left(1-x_i\right)p_n\frac{w_i^x}{c}$ is a strictly convex function of x_i.

Now, denoting

$$\Gamma\left(x_i\right)\equiv\frac{\left(1-x_i\right)p_nc}{p_n-p_{n-2}\left(1-\frac{x_n}{1-x_i}\right)-p_{n-1}\frac{x_n}{1-x_i}},$$

we can write the expected compensation cost of the principal under the replacement strategy $(x_1,\cdots,x_n)$ as:

$$\sum_{k\neq i,j}\left[x_kc+\left(1-x_k\right)p_nw_k^x\right]+x_ic+x_jc+\Gamma\left(x_i\right)+\Gamma\left(x_j\right).$$

Consider the following optimization problem:

$$\min_{\substack{x_i'\geq0\\x_j'\geq0}}\quad x_i'c+x_j'c+\Gamma\left(x_i'\right)+\Gamma\left(x_j'\right)\quad\text{s.t. }x_i'+x_j'=x_i+x_j.$$

We can equivalently write this optimization problem as:

$$\min_{x_i'}\quad\Gamma\left(x_i'\right)+\Gamma\left(x_i+x_j-x_i'\right)\quad\text{s.t. }0\leq x_i'\leq x_i+x_j.$$

Note that:

$$\frac{\partial^2\left(\Gamma\left(x_i'\right)+\Gamma\left(x_i+x_j-x_i'\right)\right)}{\partial x_i'^2}=\Gamma''\left(x_i'\right)+\Gamma''\left(x_i+x_j-x_i'\right)>0.$$

Thus, the first-order condition

$$\frac{\partial\left(\Gamma\left(x_i'\right)+\Gamma\left(x_i+x_j-x_i'\right)\right)}{\partial x_i'}=\Gamma'\left(x_i'\right)-\Gamma'\left(x_i+x_j-x_i'\right)=0$$

is uniquely satisfied by $x_i'=\frac{x_i+x_j}{2}$, ensuring that the minimum is uniquely achieved. This implies that for all $i,j\in\{1,\ldots,n-1\}$, we must have $x_i=x_j$ in the optimal replacement strategy. Consequently, it follows that $x_1=\cdots=x_{n-1}$. □

Claim 2. *The principal chooses to fully utilize AI capacity, setting $\bar{x}=1$, if the condition $p_{n-1}^2-p_np_{n-2}\leq0$ is satisfied.*

Proof of Claim 2.

Suppose, for contradiction, that the condition $p_{n-1}^2 - p_n p_{n-2} \leq 0$ is satisfied, yet the principal does not fully utilize the AI capacity. From the previous claim, we know that for a given replacement probability x_n of the central worker, the rest of the optimal replacement strategy can, without loss of generality, be restricted to strategies where the replacement probability is identical for all the peripheral workers, i.e., $x_i = x$ for each worker $i \leq n-1$, whereby $x \leq \frac{1-x_n}{n-1}$.

Take such a replacement strategy $(x, x, \cdots, x, x_n)$ for some $x \leq \frac{1-x_n}{n-1}$. Ignoring the dependence on the specific values of x and x_n, let w denote the corresponding wage of a peripheral worker and let ζ represent a peripheral worker's belief about the project's success rate when deviating under the trigger strategy profile. These are given by:

$$\zeta = p_{n-2} + \frac{x_n}{1-x}(p_{n-1} - p_{n-2}), \quad w = \frac{c}{p_n - \zeta}.$$

The principal's expected cost for a peripheral node is:

$$xc + (1-x)p_n w.$$

Substituting the expression for w and differentiating with respect to x, we obtain:

$$\begin{aligned}
\frac{\partial\left(xc + (1-x)p_n w\right)}{\partial x} &= c\left(1 - \frac{p_n}{p_n - \zeta} + \frac{(1-x)p_n}{(p_n-\zeta)^2}\frac{\partial \zeta}{\partial x}\right) \\
&= c\left(-\frac{\zeta}{p_n - \zeta} + \frac{p_n}{(p_n-\zeta)^2}\frac{x_n(p_{n-1} - p_{n-2})}{1-x}\right) \\
&= c\left(-\frac{\zeta}{p_n - \zeta} + \frac{p_n(\zeta - p_{n-2})}{(p_n-\zeta)^2}\right) \\
&= c\left(\frac{\zeta^2 - p_n p_{n-2}}{(p_n-\zeta)^2}\right).
\end{aligned}$$

Since $\zeta < p_{n-1}$, the condition $p_{n-1}^2 - p_n p_{n-2} \leq 0$ ensures that the derivative above is negative. Thus, increasing x slightly—which is feasible since the capacity is not fully utilized—leads to a reduction in expected compensation cost. This contradicts optimality, implying that the principal must fully utilize the AI capacity. □

Claims 1 and 2 imply that given x_n, the optimal replacement strategy satisfies:

$$x_i = x = \frac{1 - x_n}{n-1}, \quad \forall i \leq n-1.$$

The principal's expected compensation cost for a given x_n, denoted by W_{x_n} (with a slight abuse of notation), satisfies:

$$\frac{W_{x_n}}{c} = 1 + (n-1)(1-x)p_n\frac{1}{p_n - \zeta} + (1-x_n)p_n\frac{1}{p_n - p_{n-1}}.$$

Differentiating with respect to x_n, we obtain:

$$\frac{\partial \frac{W_{x_n}}{c\,p_n}}{\partial x_n} = -(n-1)\frac{1}{p_n-\zeta}\frac{\partial x}{\partial x_n} + (n-1)(1-x)\frac{1}{(p_n-\zeta)^2}\frac{\partial \zeta}{\partial x_n} - \frac{1}{p_n-p_{n-1}}$$
$$-(n-1)\frac{1}{p_n-\zeta}\frac{\partial x}{\partial x_n} + (n-1)(1-x)\frac{1}{(p_n-\zeta)^2}\left[\frac{p_{n-1}-p_{n-2}}{1-x} + \frac{(p_{n-1}-p_{n-2})x_n}{(1-x)^2}\frac{\partial x}{\partial x_n}\right] - \frac{1}{p_n-p_{n-1}}.$$

After simplifications, and defining $R \equiv p_n - \zeta$, this reduces to:

$$\frac{\partial \frac{W_{x_n}}{c\,p_n}}{\partial x_n} = \frac{1}{R} + \frac{(n-1)(p_{n-1}-p_{n-2})}{R^2} - \frac{(p_{n-1}-p_{n-2})x_n}{(1-x)R^2} - \frac{1}{p_n-p_{n-1}}.$$

Therefore, the sign of $\frac{\partial \frac{W_{x_n}}{c\,p_n}}{\partial x_n}$ matches the sign of:

$$(p_n-p_{n-1})(1-x)R + (n-1)(p_n-p_{n-1})(p_{n-1}-p_{n-2})(1-x) - (p_{n-1}-p_{n-2})(p_n-p_{n-1})x_n - (1-x)R^2,$$

which in turn has the same sign as:

$$-(1-x)R^2 + (p_n-p_{n-1})(1-x)R + (n-2)(p_{n-1}-p_{n-2})(p_n-p_{n-1}).$$

Moreover, note that

$$1-x = \frac{(n-2)(p_{n-1}-p_{n-2})}{R+(n-1)(p_{n-1}-p_{n-2})-(p_n-p_{n-2})}.$$

Thus, the sign of the above expression matches the sign of:

$$-(n-2)(p_{n-1}-p_{n-2})R^2 + 2(n-2)(p_{n-1}-p_{n-2})(p_n-p_{n-1})R$$
$$+(n-2)(p_{n-1}-p_{n-2})(p_n-p_{n-1})\left[(n-1)(p_{n-1}-p_{n-2})-(p_n-p_{n-2})\right].$$

We know that $R \in [p_n-p_{n-1}, p_n-p_{n-2}]$ and that the expression above attains its minimum at $R = p_n-p_{n-2}$. Evaluating at $R = p_n-p_{n-2}$, we get:

$$-(n-2)(p_{n-1}-p_{n-2})(p_n-p_{n-2})^2 + 2(n-2)(p_{n-1}-p_{n-2})(p_n-p_{n-1})(p_n-p_{n-2})$$
$$+(n-2)(p_{n-1}-p_{n-2})(p_n-p_{n-1})\left[(n-1)(p_{n-1}-p_{n-2})-(p_n-p_{n-2})\right].$$

Dividing this expression by $(n-2)(p_{n-1}-p_{n-2})$, we obtain:

$$(p_{n-1}-p_{n-2})\left[(n-2)(p_n-p_{n-1})-(p_{n-1}-p_{n-2})\right],$$

which is positive.

Therefore, the derivative of the expected cost with respect to x_n is positive. As a result, the optimal strategy requires $x_n = 0$. Furthermore, as shown in the proof of Claim 1, all replacement strategies satisfying $\sum_{i=1}^{n-1} x_i = 1$ are optimal. □

Appendix C Complementary Analysis

This appendix section provides a complementary analysis to the main findings in the document. In what follows, we use *O-ring* production function and, for clarity and conciseness, continue with the setting of three workers.

AI's impact on intra-team hierarchy of payoffs. Proposition 11 presents additional results on how the optimal AI replacement affects the payoff hierarchy, showing that the middle worker receives the highest payoffs considering both wages and the frequency and the use of AI.

Proposition 11. *The optimal AI replacement reshapes the hierarchy of payoffs within the team: the middle worker receives the highest payoff, followed by the end-most worker, and then the front-most worker. Formally:*

$$(1-x_2^*)\left(p_3 w_2^{x^*}-c\right) > (1-x_3^*)\left(p_3 w_3^{x^*}-c\right) > (1-x_1^*)\left(p_3 w_1^{x^*}-c\right).$$

Proof of Proposition 11.
Denoting $\beta \equiv \sqrt{1+\alpha}$, we get:

$$\begin{aligned}
\frac{(1-x_1^*)\left(p_3 w_1^{x^*}-c\right)}{c} &= \frac{\left(\beta^2-1\right)^2}{\beta\left(\beta+1\right)\left(2-\beta^2\right)} \\
\frac{(1-x_2^*)\left(p_3 w_2^{x^*}-c\right)}{c} &= \frac{1}{\beta\left(2-\beta^2\right)} - 1 \\
\frac{(1-x_3^*)\left(p_3 w_3^{x^*}-c\right)}{c} &= \frac{\beta-1}{2-\beta^2}.
\end{aligned}$$

Therefore:

$$\frac{(1-x_2^*)\left(p_3 w_2^{x^*}-c\right)}{c} - \frac{(1-x_3^*)\left(p_3 w_3^{x^*}-c\right)}{c} = \frac{1}{\beta\left(2-\beta^2\right)} - 1 - \frac{\beta-1}{2-\beta^2} = \frac{\left(\beta+1\right)\left(\beta-1\right)^2}{\beta\left(2-\beta^2\right)} > 0,$$

and

$$\frac{(1-x_3^*)\left(p_3 w_3^{x^*}-c\right)}{c} - \frac{(1-x_1^*)\left(p_3 w_1^{x^*}-c\right)}{c} = \frac{\beta-1}{2-\beta^2} - \frac{\left(\beta^2-1\right)^2}{\beta\left(\beta+1\right)\left(2-\beta^2\right)} = \frac{\left(\beta^2-1\right)\left(-\beta^2+\beta+1\right)}{\beta\left(\beta+1\right)\left(2-\beta^2\right)},$$

which is positive as $\beta \in \left(1, \sqrt{2}\right)$. This concludes the proof. □

AI's impact on payoffs. Proposition 12 demonstrates that the front-most worker may experience a loss or gain in payoff, and provides the conditions for each outcome under the O-ring production function. Recall that, in the main text, we have already demonstrated that the optimal AI adoption

always resulted in higher (lower) payoffs for the middle (end-most) worker. This proposition therefore completes the earlier payoff analysis.

Proposition 12. *The front-most worker's payoff declines following AI replacement,*

$$(1-x_1^*)\left(p_3 w_1^{x^*}-c\right) < p_3 w_1^0 - c,$$

if and only if effort complementarity is weak enough—that is, $\alpha \geq \bar{\alpha}$ for some threshold $\bar{\alpha} \in (0,1)$.

Proof of Proposition 12.
Note that:

$$\begin{aligned}\frac{p_3 w_1^0 - c}{c} - \frac{(1-x_1^*)\left(p_3 w_1^{x^*}-c\right)}{c} &= \frac{\left(\beta^2-1\right)^3}{(2-\beta^2)\left[\left(\beta^2-1\right)^2+\beta^2\right]} - \frac{\left(\beta^2-1\right)^2}{\beta(\beta+1)(2-\beta^2)} \\ &= \frac{\left(\beta^2-1\right)^2}{\beta(\beta+1)(2-\beta^2)\left[\left(\beta^2-1\right)^2+\beta^2\right]}\left(\beta(\beta+1)\left(\beta^2-1\right) - \left[\left(\beta^2-1\right)^2+\beta^2\right]\right).\end{aligned}$$

The second line is a positive value multiplied with $\beta^3-\beta-1$, which is positive if and only if β is larger than $\sqrt[3]{\frac{1}{2}+\sqrt{\frac{23}{108}}}+\sqrt[3]{\frac{1}{2}-\sqrt{\frac{23}{108}}} \approx 1.318$ (by Cardano's formula). □

As is indicated in Proposition 6, the optimal AI deployment on the one hand increases the wage of the front-most worker. On the other hand, he faces replacement. Proposition 12 implies that the latter effect dominates when α is large (tasks are highly independent of each other).

AI's impact on intra-team inequality of payoffs. Proposition 13 explores how AI alters intra-team payoff inequality and shows that, similarly to wage inequality, the inequality of payoffs also declines with optimal AI adoption.

Proposition 13. *The intra-team payoff inequality decreases with optimal AI adoption, relative to the intra-team inequality without AI adoption. That is:*

$$\left[(1-x_2^*)\left(p_3 w_2^{x^*}-c\right) - (1-x_1^*)\left(p_3 w_1^{x^*}-c\right)\right] < \left[\left(p_3 w_3^0 - c\right) - \left(p_3 w_1^0 - c\right)\right].$$

Proof of Proposition 13.
Note that:

$$\frac{1}{c}\left[(1-x_2)\left(p_3 w_2^x - c\right) - (1-x_1)\left(p_3 w_1^x - c\right)\right] - \frac{1}{c}\left[\left(p_3 w_3^0 - c\right) - \left(p_3 w_1^0 - c\right)\right]$$

$$= \left[\frac{1}{\beta\left(2-\beta^2\right)} - 1 - \frac{\left(\beta^2-1\right)^2}{\beta\left(\beta+1\right)\left(2-\beta^2\right)}\right] - \left[\frac{1}{2-\beta^2} - \frac{1}{\left(2-\beta^2\right)\left[\left(\beta^2-1\right)^2+\beta^2\right]}\right]$$

$$= \frac{\beta-1}{2-\beta^2} - \frac{\beta^2\left(\beta^2-1\right)}{\left(2-\beta^2\right)\left[\left(\beta^2-1\right)^2+\beta^2\right]}$$

$$= \frac{\beta-1}{\left(2-\beta^2\right)\left[\left(\beta^2-1\right)^2+\beta^2\right]}\left\{\left(\beta-1\right)\left[\left(\beta-1\right)\left(\beta^2+\beta-1\right)-3\right]-1\right\}$$

which is negative as $\beta \in \left(1, \sqrt{2}\right)$. This completes the proof. □