Received: 4 January 2021 | Revised: 20 July 2021 | Accepted: 3 September 2021

DOI: 10.1111/jems.12456

ORIGINAL ARTICLE

Journal of Economics & Management Strategy WILEY

# Managing automation in teams

**Mustafa Dogan**[1] | **Pinar Yildirim**[2]

[1]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[2]The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania, USA

**Correspondence**
Mustafa Dogan, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
Email: mdogan@mit.edu

**Abstract**

In this paper, we study a principal's decision to introduce automation into a production process governed by a team of employees. When introduced, automation displaces an employee with a machine. This displacement increases efficiency as the machine carries out the tasks of the employee at a lower cost, and reduces the scope of moral hazard as the machine does not make unobserved effort choices. We show that, despite the direct benefits, a principal may prefer not to adopt automation due to its indirect costs. Before automation is introduced, the principal is able to take advantage of her ability to shape the interactions between the team members to manage the agency problem. Automation eliminates this ability and removes an incentive device at the principal's discretion, resulting in an indirect cost. On the one hand, adopting automation is always optimal when the principal incentivizes employees independently, abstaining from creating a team interaction. On the other hand, automation may be suboptimal when the principal incentivizes employees by encouraging them to compete via a "relative performance evaluation" contract or to cooperate via a "joint performance evaluation" contract. We offer two extensions to test the robustness of these findings qualitatively. First, the findings carry through if we consider alternative effects of automation, where it impacts employees symmetrically without displacing any employee. Second, the findings also remain consistent when there are synergies between the efforts of team members.

## 1 | INTRODUCTION

Using robots and artificial intelligence (AI) to automate tasks which were once carried out by human employees is increasingly more common in retail, customer service, operations, medicine, aviation, and aerospace (Reed & Peshkin, 2008). According to some reports, the pace of adoption will increase, resulting in up to 40% of jobs being automated by 2030 (Frey & Osborne, 2017). Such rapid adoption seems to make sense given the direct benefits of automation, such as consistent work input and reduced costs of operation (Acemoglu & Restrepo, 2020). Given these direct benefits, is adopting automation always a good idea?

In this paper, we try to address this question by studying a principal's decision to introduce automation into a production process governed by a team of two employees. We build a model borrowing from Che and Yoo (2001), where employees privately decide either to work at some cost or to shirk. The principal only observes a noisy performance signal of their effort choices. To induce the employees to work, the principal uses incentive contracts that rely on performance measures. The performance measures of the two employees are correlated. As a result, the principal can set the compensation scheme of an employee in a way that depends on the performance evaluation of the

  

other employee. This gives the principal the ability to shape the interaction between the team members. She can encourage them to compete by using a "relative performance evaluation" (RPE) contract, where the compensation of an employee decreases with higher performance of his teammate. Alternatively, she can encourage them to cooperate by using a "joint performance evaluation" (JPE) contract, where the compensation of an employee increases with the higher performance of his teammate. Or she can abstain from creating interactions between the employees and compensate each employee based on their own "independent performance evaluation" (IPE).

When introduced, automation displaces an employee, creating a human–machine production team. The machine carries out identical tasks to that of the employee, and always exerts high effort but at a lower cost. Therefore, automation increases the efficiency and reduces the scope of moral hazard in the production system. In this setting, the principal decides whether to introduce automation into the system.

We show that, despite the direct benefits, a principal may prefer not to adopt automation due to its indirect costs. Before automation is introduced, the principal is able to take advantage of her ability to shape the interactions between the team members through her choice of compensation contracts, which helps her to manage the underlying agency problem. Automation eliminates this ability and removes an incentive device at the principal's discretion, resulting in an indirect cost. On the one hand, adopting automation is always optimal when the principal incentivizes employees independently, abstaining from creating a team interaction. On the other hand, automation may be suboptimal when the principal incentivizes employees by encouraging them to compete via an RPE contract, or by encouraging them to cooperate via a JPE contract.

In the extensions, we first consider an alternative setting where automation impacts employees symmetrically, without displacing any. Specifically, once automation is introduced into a production team, it reduces the cost of high effort for all employees, and at the same time increases the likelihood that they receive a favorable performance evaluation. While the reduced cost of effort is always desirable from the principal's perspective as it reduces the cost of compensation, the high-likelihood of obtaining a favorable performance evaluation reduces the principal's monitoring ability and creates a potential indirect negative effect. RPE contracts shield the principal from such negative effects, hence, it is optimal to adopt automation when operating under an RPE contract. On the contrary, JPE and IPE suffer from this negative effect, hence adoption may be suboptimal when operating under these contracts. This reaffirms our key finding from the baseline model.

Next, we consider a second alternative setting where there are synergies between the efforts of team members. In this setting, there are no separate performance indicators for employees, and a single team output serves as the only source of information about the employees' effort choices. We show that our main results hold in this setting as well.

These findings contribute to the literature in two distinct ways. First, although couched in the context of automation, our study relates to the long tradition of employee contracts and team dynamics (e.g., Baker et al., 2002; Bonatti & Rantakari, 2016; Chan et al., 2014; Rayo, 2007; Weng & Carlsson, 2015). The study of relative incentive schemes, specifically, the comparison between RPE and JPE schemes, go back to Hölmstrom (1982). This literature finds that RPE contracts help a principal to reward employees based on effort rather than luck (due to exogenous shocks that influence the employee output, partially independent of effort). JPE, on the other hand, promotes cooperation since an agent is rewarded only if his peer performs well (Alchian & Demsetz, 1972; Hölmstrom, 1982). Bonatti and Hörner (2011) examine moral hazard in teams in a dynamic setting and demonstrate that free-riding results in lower effort and procrastination. To study automation in a team setting, we borrow from Che and Yoo (2001), who study a principal's wage-setting problem when two agents are working together and a principal monitors their effort imperfectly. In this setting, the authors show that, while RPE is the optimal incentive scheme when employees interact only in a single period, JPE can also be an optimal incentive scheme over multiple periods. And while they mention the possibility of collusion that RPE contracts create, they do not characterize the optimal RPE that rules out collusion for the whole parameter space they study. To answer the questions of interest to us, we complete this characterization. Accounting for collusion reduces the appeal of RPE and increases the wages paid to employees under RPE. Moreover, if all firms had the option to automate tasks, firms using JPE and IPE would be more likely to adopt automation than firms which use RPE. In fact, as technology progresses further and becomes cheaper, hiring human teams using JPE may become extinct.

Second, this study contributes to the literature on automation and new technology adoption. There is a keen interest in understanding the impact of automation and AI in the workplace (Agrawal et al., 2016, 2017b; Brynjolfsson & McAfee, 2011; Moriarty & Swartz, 1989; Venkatraman, 1994). Dogan et al. (2021) study how strategic adoption of automation influences allocation of decision-making authority within an organization. A number of studies tried to focus on how new technologies alter the performance of employees. Mobius and Schoenle (2006) study automation's

impact on division of labor, and argue that it results in skill upgrading by employees. Aral et al. (2012) focus on a human capital management software adoption, and compare how the performance of employees using the new technology improves. Agrawal et al. (2017a) show the complementarity of employees and machines by arguing that machine predictions are more valuable when payoffs are better understood through human judgment. Acemoglu and Restrepo (2018) consider automation to be both a substitute and a complement to human work, but demonstrate empirically that automation displaced 670,000 employees between 1990 and 2007 (Acemoglu & Restrepo, 2020). Similar to these studies, we also consider how automation influences the performance of employees, but in the baseline model we focus on the employees whose jobs are not directly impacted by automation. Moreover, we focus on the impact of automation in a team setting, which is novel, based on our reading of the literature.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3.1 describes the principal's problem with automation, Section 3.2 describes it without automation. We characterize the optimal automation adoption decision in Section 3.3. In Section 4, we provide two alternative modeling settings, one with symmetric effects of automation (Section 4.1) and with complementarity between team member's effort (Section 4.2). Finally, in Section 5, we conclude.

## 2 | MODEL

Consider a firm which delivers a product or a service that is the outcome of a team production process. We assume that the team consists of two employees who are overseen by a principal. The principal decides whether to introduce automation into the production process or not, and sets the compensation contracts for any employee involved in the production.

Production process takes place over a discrete and infinite time horizon. Let $\delta$ be the common discount factor that the principal ("she") and the employees ("he") use to value their future payoffs. Production relies on a group of tasks being completed, and the principal oversees two employees who perform these tasks. Before the production starts, she decides whether to introduce automation into the process, the details of which will be described subsequently.

### 2.1 | Setup of production without automation

The production team consists of two employees. Each employee makes a binary effort choice in each period and can either "work" by setting his effort level to $e = 1$, or "shirk" by setting effort to $e = 0$. The cost of working is denoted by $c > 0$ and the cost of shirking is normalized to 0.

The employees' effort choices are not directly observed by the principal, but she observes an informative signal of their effort. The signals, similar to the effort choices, are binary and can be either "favorable" ($s = 1$) or "unfavorable" ($s = 0$). While the signal realization of the employee is independent of the effort choice of the other employee, there is an underlying common component influencing the realizations of both. We will refer to the common component as the aggregate factor which can be either "good" or "bad," with probabilities $\sigma < 1$ and $1 - \sigma$, respectively. $\sigma$ is common to both employees and corresponds to the influence of outside factors on their productivities. If the factor is good (has a positive effect on productivity), then the signals of both employees are favorable regardless of their effort choices. If the shock is bad (has no positive effect on productivity), then their signals depend on their effort choices. In this case, if a employee chooses to put effort ($e = 1$), then his performance signal will be favorable with probability $p_1$, and if he chooses to shirk ($e = 0$) then his performance signal will be favorable with probability $p_0$, where $1 > p_1 > p_0 > 0$. That is, when the aggregate factor is not good, the probability that an employee receives a favorable evaluation is higher if he exerts high effort. The distribution of the performance signals is summarized in Table 1.

Employees observe the effort choices of each other as they work closely. Each observes the working or shirking choice of his peer and chooses his own effort accordingly, in his best interest. What is in his best interest depends on the compensation contract set by the principal. The principal can set the compensation such that she can take advantage of an employee's ability to observe his peer's effort. Specifically, by setting the contract accordingly, the principal can shape the nature of the relation between the employees. Even though the performance evaluations of the employees are independent of each other, the principal can create an interaction between the employees by conditioning the compensation of one employee on the performance of the other. This introduces a possible channel to sustain a desirable outcome for the principal. Employees' ability to monitor each other and respond to a unilateral deviation helps the manager to keep the

**TABLE 1** Joint distribution of the signals, conditional on effort profile of employees

| | | Signal pairs | | | |
|---|---|---|---|---|---|
| | | **(1,1)** | **(1,0)** | **(0,1)** | **(0,0)** |
| Effort pairs | (1,1) | $\sigma + (1-\sigma)p_1^2$ | $(1-\sigma)p_1(1-p_1)$ | $(1-\sigma)(1-p_1)p_1$ | $(1-\sigma)(1-p_1)^2$ |
| | (1,0) | $\sigma + (1-\sigma)p_1 p_0$ | $(1-\sigma)p_1(1-p_0)$ | $(1-\sigma)(1-p_1)p_0$ | $(1-\sigma)(1-p_1)(1-p_0)$ |
| | (0,1) | $\sigma + (1-\sigma)p_0 p_1$ | $(1-\sigma)p_0(1-p_1)$ | $(1-\sigma)(1-p_0)p_1$ | $(1-\sigma)(1-p_0)(1-p_1)$ |
| | (0,0) | $\sigma + (1-\sigma)p_0^2$ | $(1-\sigma)p_0(1-p_0)$ | $(1-\sigma)(1-p_0)p_0$ | $(1-\sigma)(1-p_0)^2$ |

employees working. Interaction, therefore, introduces a capacity of mutual monitoring that the principal can exploit to mitigate the agency problem.

## 2.2 | Setup of production with automation

When automation is introduced into the system, it can have an asymmetric or symmetric effect on pair of employees. That is, automation may affect both employees equally, or may affect one of the employees more than another. In the baseline model, we focus on the asymmetric case, and consider a scenario where one employee is displaced by the automation technology, while the other remains unimpacted.[1] One employee is replaced with an automated machine, which performs the tasks of the displaced employee at a fraction of the cost, $\alpha c$, where $\alpha < 1$. The effort cost and the information structure that generates the performance measure stay the same for the remaining employee—his cost of effort is still $c$, and the probability of positive aggregate factor is still $\sigma$.

Thus, while automation introduces a cost efficiency into the system, it also narrows down the scope of between–employee interaction in the production process. This is the indirect cost of adopting automation: it removes peer monitoring between employees, and therefore reduces the principal's ability to mitigate the agency problem through contracts. The trade-off between the direct benefits and indirect costs of automation is the core of the discussion in this paper.

# 3 | ANALYSIS OF OPTIMAL COMPENSATION CONTRACTS

This section characterizes the optimal compensation contracts in both cases—production with and without automation. Proofs of this section are provided in Appendix A.

We make the assumption that high effort is sufficiently valuable for the principal such that her central focus is to find the optimal wage to induce high effort ($e = 1$) in every period from each employee that plays a role in production. We assume that employees have limited liability. Therefore, the payment that an employee receives after any contingency has to be nonnegative.

## 3.1 | Optimal contract under automation

We first investigate the optimal compensation scheme in a division where automation is adopted. A compensation scheme, in this setting, is a two-dimensional vector specifying the amount of pay corresponding to each possible performance outcome of the remaining employee. Formally, let $(w_1, w_0)$ be the wage scheme, where $w_1$ and $w_0$ represent the pay to the employee if his performance signal is $s = 1$ and $s = 0$, respectively.

To induce high effort, the principal needs to ensure that the employee earns a higher payoff from exerting effort than from shirking:

$$\underbrace{(\sigma + (1-\sigma)p_1)w_1 + (1-\sigma)(1-p_1)w_0 - c}_{\text{Expected payoff from working}} \geq \underbrace{(\sigma + (1-\sigma)p_0)w_1 + (1-\sigma)(1-p_0)w_0.}_{\text{Expected payoff from shirking}} \tag{1}$$

The left-hand side of the inequality is the employee's expected payoff from working. He earns the wage $w_1$ if he receives a favorable evaluation with probability $\sigma + (1-\sigma)p_1$, and he earns the wage $w_0$ with the remaining probability. The right-hand side is the expected payoff from shirking. Note that if he shirks, he receives $w_1$ with probability $\sigma + (1-\sigma)p_0$ and receives $w_0$ with the complementary probability. The constraint in Equation (1) is equivalent to the following simpler expression, which we label as the incentive constraint $\mathcal{IC}$:

$$(1-\sigma)(p_1 - p_0)(w_1 - w_0) \geq c. \quad (\mathcal{IC})$$

To satisfy the constraint $\mathcal{IC}$, the more likely it is for the employee to get rewarded by luck (due to high $\sigma$ or low $(p_1 - p_0)$), the more the principal has to reward a favorable signal relative to the unfavorable signal (i.e., $w_1 - w_0$ has to be larger). So whenever the monitoring capacity of the principal is low, the employee is more expensive.

The principal's problem is to minimize her cost (i.e., the expected compensation paid to the employee), with respect to the constraint $\mathcal{IC}$:

$$\min_{w_1, w_0} \underbrace{(\sigma + (1-\sigma)\, p_1)w_1 \; + (1-\sigma)(1-p_1)w_0}_{\text{Expected compensation paid to the employee}} \quad \text{s.t.} \quad \mathcal{IC} \qquad (\mathcal{P})$$

Since the employee has limited liability, it is straightforward to see that under the optimal compensation scheme, the agent receives a positive payment only after a favorable performance signal, therefore $w_0 = 0$. Moreover, the incentive constraint is binding, we can derive the optimal value of $w_1$. The following proposition summarizes the optimal compensation for the employee in an automated team.

**Lemma 1.** *Under automation, the optimal incentive scheme that induces the remaining employee to work in every period satisfies* $w_1 = \hat{c}$ *and* $w_0 = 0$, *where* $\hat{c} \equiv \frac{c}{(1-\sigma)(p_1 - p_0)}$.

When production is automated, the employee is paid only when his evaluation is favorable and his compensation $w_1 = \hat{c}$ depends on the likelihood of a positive aggregate factor ($\sigma$) and the difference in the return from working and shirking under a bad aggregate factor ($p_1 - p_0$). It is harder for principal to detect shirking behavior when $\sigma$ is high, and when $p_1 - p_0$ is low. In these circumstances, the principal has to pay the employee more to incentivize working.

## 3.2 | Optimal compensation without automation

Now, we characterize the optimal compensation scheme for employees in the production system without automation. The principal sets the compensation scheme as a four dimensional vector $\boldsymbol{w} = (w_{11}, w_{10}, w_{01}, w_{00})$, where $w_{s_i s_j}$ is the amount paid to employee $i$ when his signal is $s_i$ and the other employee's signal is $s_j$.[2] The wage contract allows the principal to manipulate employee interaction to her benefit, but also creates the possibility of a collusive outcome that she would like to avoid.

In tradition with Gibbons and Murphy (1990), Choi (1993), and Che and Yoo (2001), we categorize contracts based on how an employee is rewarded depending on his peer's performance. Under JPE, an employee is rewarded for the good performance of his peer. Therefore, a JPE satisfies:

$$(w_{11}, w_{01}) > (w_{10}, w_{00}),$$

implying that $w_{11} \geq w_{10}$, and $w_{01} \geq w_{00}$, where at least one of these inequalities has to be strict. So the employee is paid more when his peer obtains a favorable evaluation.

Under RPE, an employee is rewarded for his superior performance relative to his peer. Therefore, an RPE satisfies:

$$(w_{10}, w_{00}) > (w_{11}, w_{01}).$$

Thus, the employee is paid more when his peer obtains an unfavorable evaluation.

Finally, if an employee's compensation is independent of his peer's evaluation, we refer to it as the IPE. An IPE satisfies:

$$(w_{10}, w_{00}) = (w_{11}, w_{01}).$$

The compensation scheme that the principal chooses puts the employees into a repeated interaction over the infinite horizon. She would like to choose a compensation scheme such that both employees working in every period is a subgame perfect equilibrium outcome, that is, $(e_1, e_2) = (1, 1)$, of the corresponding repeated game. An incentive scheme that gives rise to this outcome in a subgame perfect equilibrium, however, may also give rise to other subgame perfect equilibria that support collusive outcomes in which at least one employee does not work in some periods, which the principal would like to avoid. To address this issue, we use the following refined equilibrium concept defined by Che and Yoo (2001).[3]

**Definition 1.** For a given compensation scheme $\boldsymbol{w}$, a subgame perfect equilibrium of the corresponding game is called a "collusion-free equilibrium," if there does not exist another subgame perfect equilibrium with an outcome in which the total payoff of the employees is strictly higher.

There are several benefits of adopting this equilibrium concept. First, it allows us to draw a unique prediction regarding the outcome of the employees' repeated interactions. Moreover, it eliminates the possibility of collusion between the employees in any period on the path of play. Next, we focus on the principal's problem and characterize the optimal incentive scheme that induces the desired path of play, that is, repeated joint effort, as a collusion-free equilibrium outcome.

For a given wage $\boldsymbol{w}$ and effort profile $(e_i, e_j)$, the expected within-period compensation of employee $i$ is denoted by $\pi_i(e_i, e_j, \boldsymbol{w})$, and is equal to:

$$\begin{aligned}\pi_i(e_i, e_j, \boldsymbol{w}) = {} & [\sigma + (1-\sigma)p_{e_i}\, p_{e_j}]w_{11} \\ & + (1-\sigma)[p_{e_i}(1-p_{e_j})w_{10} + (1-p_{e_i})p_{e_j}\, w_{01} + (1-p_{e_i})(1-p_{e_j})w_{00}]\end{aligned}$$

The following lemma asserts that the optimal wage of the employee is set to zero if he receives an unfavorable performance signal regardless of the performance measure of his peer.

**Lemma 2.** *The optimal compensation of the employees satisfies $w_{01} = w_{00} = 0$. In words, the wage of an employee is zero if his evaluation is unfavorable.*

Lemma 2 gives the intuition that the principal prefers not to pay a positive amount to employees in case of an unfavorable evaluation, as it makes it difficult to incentivize them to work. It is important to note that this result guarantees that the optimal compensation will be one of JPE, RPE, or IPE depending on how $w_{11}$ and $w_{10}$ compare to each other. When $w_{11} > w_{10}$, $\boldsymbol{w}$ is a JPE; when $w_{11} < w_{10}$, $\boldsymbol{w}$ is an RPE, when $w_{11} = w_{10}$, it is an IPE. Lemma 2 reduces the principal's problem to finding the optimal values of $w_{11}$ and $w_{10}$. The principal has to choose these variables such that the repeated joint work is an outcome of a collusion-free equilibrium of the corresponding repeated interaction among the employees. The following constraint comprises a necessary condition for the principal's problem.

$$\underbrace{\frac{1}{1-\delta}(\pi(1,1,\boldsymbol{w}) - c)}_{\text{Avg. expected payoff from working}} \geq \underbrace{\pi(0,1,\boldsymbol{w}) + \frac{\delta}{1-\delta}\min\{\pi(0,0,\boldsymbol{w}), \pi(0,1,\boldsymbol{w})\}}_{\text{Avg. expected payoff from deviating}} \qquad (\mathcal{IC}_0)$$

The left-hand side of the constraint $(\mathcal{IC}_0)$ represents the discounted expected payoff an employee receives when he and his peer work in every period. The right-hand side is the minimum discounted expected utility that an employee may receive from deviating and shirking in every period. If the inequality does not hold, then it is not possible to sustain repeated joint work as a subgame perfect equilibrium, as deviation makes employees better off. Note that, this constraint is just a necessary condition and is not sufficient for the principal's problem. Therefore, we can define a *relaxed version* of the principal's problem, which is to minimize the expected compensation to induce repeated effort, as follows:

$$\min_{w_{11}, w_{10}} \pi(1,1,\boldsymbol{w}) \ \text{ s.t. } \mathcal{IC}_0. \ (\mathcal{P}_0)$$

Che and Yoo (2001) characterizes the solution to this problem as follows. For each $\sigma$, there exists a $\delta_\sigma$ such that if $\delta > \delta_\sigma$, then the JPE scheme $\boldsymbol{w}^J = \left(w_{11}^J, 0, 0, 0\right)$ with

$$w_{11}^J = \frac{\hat{c}}{p_1 + \delta p_0}$$

solves $(\mathcal{P}_0)$. Otherwise, if $\delta \leq \delta_\sigma$, then the RPE scheme $\boldsymbol{w}^S = \left(0, w_{10}^S, 0, 0\right)$ with

$$w_{10}^S = \frac{\hat{c}}{1 - p_1}$$

solves problem $(\mathcal{P}_0)$.

This solution, however, does not necessarily give us the optimal solution, as $(\mathcal{P}_0)$ is just a relaxed version of the principal's problem. We need to check if the incentive schemes $\boldsymbol{w}^J$ and $\boldsymbol{w}^S$ induce repeated joint work as a collusion-free equilibrium outcome.

### 3.2.1 | The contract $\boldsymbol{w}^J$ works

The JPE contract, $\boldsymbol{w}^J$, possesses all the desirable properties: it induces repeated joint work as a collusion-free equilibrium outcome. Specifically, (work, work)$^\infty$ can be sustained as a subgame perfect outcome, and it does not create any possibility of collusion. To see the former, notice that under $\boldsymbol{w}^J$, the interaction between the employees is equivalent to a prisoner's dilemma game in which (shirk, shirk) is the unique stage game Nash equilibrium. Moreover, as $\boldsymbol{w}^J$ satisfies the constraint $(\mathcal{IC}_0)$, initiating (shirk, shirk)$^\infty$ can deter unilateral deviations from the desired outcome (work, work)$^\infty$. Such a punishment is self-enforcing, since repetition of (shirk, shirk) is a stage game Nash equilibrium. This suggests that (work, work)$^\infty$ can be sustained as a subgame perfect equilibrium outcome. To see the latter, under $\boldsymbol{w}^J$, the effort pair (work, work) maximizes the employees' total payoff in any period compared to unilateral or jointly shirking. More precisely:

$$\underbrace{2\pi(1,1,\boldsymbol{w}^J) - 2c}_{\text{Total payoff from (work, work)}} \geq \underbrace{\pi(1,0,\boldsymbol{w}^J) - c + \pi(0,1,\boldsymbol{w}^J)}_{\text{Total payoff from (work, shirk)}},$$

and

$$\underbrace{2\pi(1,1,\boldsymbol{w}^J) - 2c}_{\text{Total payoff from (work, work)}} \geq \underbrace{2\pi(0,0,\boldsymbol{w}^J)}_{\text{Total payoff from (shirk, shirk)}}.$$

In consequence, whenever $\boldsymbol{w}^J$ comprises a solution to the relaxed problem $(\mathcal{P}_0)$, it also comprises an optimal contract for the principal. Such a contract is effective as long as both employees value their future earnings, that is, as long as the discount factor is high. Peer's deviation to shirking in response to one's own shirking acts as a disciplining mechanism to keep both employees working. Thus, even though the principal cannot directly monitor her employees, she benefits from their mutual monitoring capacity.

### 3.2.2 | RPE contract $\boldsymbol{w}^S$ does not work

Despite solving the relaxed problem $(\mathcal{P}_0)$, the RPE contract $\boldsymbol{w}^S$ also creates a possibility of collusion for the employees, hence cannot be an optimal contract. Specifically, while both employees working in every period (work, work)$^\infty$ can be obtained as a subgame perfect equilibrium under $\boldsymbol{w}^S$, they can obtain a higher payoff if they collude over a number of strategy profiles that include unilateral or bilateral deviations from (work, work) on the path of play. For instance, they can choose to alternate between (work, shirk) and (shirk, work) profiles over consecutive periods. The discounted expected payoffs from this strategy for the employees are:

$$\underbrace{\frac{1}{1-\delta^2}(\pi(1,0,\boldsymbol{w}^S)-c)+\frac{\delta}{1-\delta^2}\pi(0,1,\boldsymbol{w}^S)}_{\text{employee } i\text{'s payoff from (work, shirk), (shirk, work),....play}},$$

and

$$\underbrace{\frac{\delta}{1-\delta^2}(\pi(1,0,\boldsymbol{w}^S)-c)+\frac{1}{1-\delta^2}\pi(0,1,\boldsymbol{w}^S)}_{\text{employee } i\text{'s payoff from (shirk, work), (work, shirk), ...play}}$$

for the employee who is working and who is shirking in the first period, respectively. One can easily see that, both of these expected payoffs are indeed larger than the resulting payoff from working in all periods, which equals $\frac{1}{1-\delta}(\pi(1,1,\boldsymbol{w}^S)-c)$. And this outcome can be supported in a subgame perfect equilibrium in which the deviations trigger the repeated play of (work, work), which is self-enforcing. Thus, the RPE contract $\boldsymbol{w}^S$ incentivizes collusion, resulting in alternating shirking in every other period.

The discussions so far regarding the optimal incentive scheme overlap with the analysis provided in Che and Yoo (2001). They pointed out that the solution to $(\mathcal{P}_0)$ does not give us the optimal compensation scheme when the principal wants to preclude collusion. They also argue that when $\boldsymbol{w}^S$ solves $(\mathcal{P}_0)$, one can find another RPE scheme that precludes collusion. However, they do not provide a full characterization of the optimal incentive scheme inducing repeated joint work as a collusion-free equilibrium outcome, which will be crucial for us when we study the principal's automation strategy. Thus, we complement their work and characterize the optimal incentive scheme over the entire parameter space.

To provide a full characterization of the solution to the principal's problem, we need to focus on the cases where the incentive scheme $\boldsymbol{w}^S$ solves $(\mathcal{P}_0)$. This requires us to include additional constraints to preclude the possibility of collusion in such cases. However, this is not a trivial task, since there are many potential forms of collusion that the employees can engage in, and the principal has to rule them all out. Our main strategy to tackle this issue is to consider a specific collusive outcome—*alternating shirking* (AS) in which the employees alternate between shirking and working—which suffices to prevent all other forms of collusion, as will be shown.

In the following, we concentrate on the principal's problem by restricting her choice to the set of contracts satisfying $w_{11} \leq w_{10}$. This is because, among the incentive schemes satisfying $w_{11} > w_{10}$, the optimal one is $\boldsymbol{w}^J$. While choosing the optimal incentive scheme among the ones satisfying $w_{11} \leq w_{10}$, on top of $(\mathcal{IC}_0)$, which is still a necessary condition, we will include additional constraints that prevent AS from being a collusive outcome.

Specifically, the principal has to choose a scheme such that either (1) alternating shirking cannot be supported as a subgame perfect equilibrium (SPE) outcome, or (2) the corresponding total payoff of the employees from alternating shirking is lower than that from joint repeated work. The following constraints address (1):

$$\underbrace{\frac{1}{1-\delta^2}\pi(0,1,\boldsymbol{w})+\frac{\delta}{1-\delta^2}(\pi(1,0,\boldsymbol{w})-c)}_{\text{Payoff from AS, working employee}} < \underbrace{\frac{1}{1-\delta}(\pi(1,1,\boldsymbol{w})-c)}_{\text{Payoff from deviation}} \qquad \left(\mathcal{IC}_1^{AS}\right)$$

$$\underbrace{\frac{1}{1-\delta^2}(\pi(1,0,\boldsymbol{w})-c)+\frac{\delta}{1-\delta^2}\pi(0,1,\boldsymbol{w})}_{\text{Payoff from AS, shirking employee}} < \underbrace{\pi(0,0,\boldsymbol{w})+\frac{\delta}{(1-\delta)}(\pi(1,1,\boldsymbol{w})-c)}_{\text{Payoff from deviation}} \qquad \left(\mathcal{IC}_{1'}^{AS}\right)$$

These two constraints assert that one of the employees has a profitable deviation from AS, hence it cannot be sustained as an SPE outcome.[4]

To address (2), the total payoff for employees from AS must be smaller than their total payoff from (work, work) in every period:

$$\underbrace{\pi(1,0,\boldsymbol{w})-c+\pi(0,1,\boldsymbol{w})}_{\text{Total payoff from AS}} \leq \underbrace{2(\pi(1,1,\boldsymbol{w})-c).}_{\text{Total payoff from repeated working}} \qquad \left(\mathcal{IC}_2^{AS}\right)$$

The optimal incentive scheme satisfies at least one of $(\mathcal{IC}_1^{AS})$, $(\mathcal{IC}_{1'}^{AS})$, and $(\mathcal{IC}_2^{AS})$. We define the following global constraint:

$$\mathcal{IC}^{AS} \equiv \left(\mathcal{IC}_1^{AS}\right) \vee \left(\mathcal{IC}_{1'}^{AS}\right) \vee \left(\mathcal{IC}_2^{AS}\right), \qquad (\mathcal{IC}^{AS})$$

and then define the following problem:

$$\min_{w_{11},w_{10}} \pi(1,1,\boldsymbol{w}) \quad \text{s.t.} \quad \mathcal{IC}_0, (\mathcal{IC}^{AS}), w_{11} \leq w_{10}. \qquad (\mathcal{P}_{AS})$$

Notice that this problem is another relaxed version of the principal's problem (conditional on $w_{11} \leq w_{10}$), since it only accounts for a specific form of collusion. The following lemma, which details the solution to this problem, demonstrates that the constraints that are taken into account in problem $(\mathcal{P}_{AS})$ are sufficient to preclude all collusive outcomes when the incentive scheme satisfies $w_{11} \leq w_{10}$. Therefore, the solution to $(\mathcal{P}_{AS})$ coincides with the solution of the principal's problem.

**Lemma 3.** *The optimal solution to $(\mathcal{P}_{AS})$ is either an RPE contract $\boldsymbol{w}^R = \left(0, w_{10}^R, 0, 0\right)$, where $w_{10}^R = \frac{\hat{c}}{1-(1+\delta)p_1}$, or an IPE contract with $\boldsymbol{w}^I = (\hat{c}, \hat{c}, 0, 0)$. Moreover, if the optimal solution to the principal's problem satisfies $w_{11} \leq w_{10}$, then that solution coincides with the solution of $(\mathcal{P}_{AS})$.*

Lemma 3 does not only provide a solution to $(\mathcal{P}_{AS})$, but also indicates that precluding AS is sufficient to preclude all possible collusive outcomes. If the solution of $(\mathcal{P}_{AS})$ is the IPE contract $\boldsymbol{w}^I$, then employees' wages only depend on their own performance outcome. Collusion is not relevant in this case. If the solution is $\boldsymbol{w}^R$, collusion is again prevented, but this time by design.[5] In particular, the value of $w_{10}^R$ is set large enough to attract the employees to work all the time. In this regard, the difference $w_{10}^R - w_{10}^S$ can be considered as the cost of precluding collusion for the principal who would like to induce repeated joint work from the employees in an RPE scheme. Under $\boldsymbol{w}^R$, the corresponding collusion-free equilibrium features repetition of (work, work) on and off the equilibrium path. This comprises a subgame perfect equilibrium because (work, work) is an equilibrium of the stage game. In this pay scheme, the principal basically puts employees into a race and effectively asks them to compete with each other to have a better performance signal than their peer. The stakes are so strong that the employees never engage in collusion. As a consequence, the employees always prefer to work and compete with each other to receive a positive evaluation.

### 3.2.3 | Summary of optimal incentive schemes

The following proposition summarizes the optimal contract conditional on the value of the discount factor $\delta$. The principal either puts employees in competition via a collusion-proof RPE, or chooses a JPE contract to take advantage of the mutual monitoring capacity of the employees, or incentivizes them independently by choosing an IPE scheme which exterminates team interaction.

**Proposition 1.** *Let $\delta_J \equiv \frac{\sigma(1-p_1)}{(\sigma+(1-\sigma)p_1)p_0}$ and $\delta_R \equiv \frac{\sigma(1-p_1)}{(\sigma+(1-\sigma)p_1)p_1}$. Without automation, the optimal contract is:*

$$\boldsymbol{w} = \begin{cases} \text{JPE with } \boldsymbol{w}^J & \text{if } \delta > \delta_J \\ \text{IPE with } \boldsymbol{w}^I & \text{if } \delta \in (\delta_R, \delta_J] \\ \text{RPE with } \boldsymbol{w}^R & \text{if } \delta \leq \delta_R. \end{cases}$$

The proposition follows from a comparison of the expected costs of the incentive schemes $\boldsymbol{w}^J$, $\boldsymbol{w}^I$, and $\boldsymbol{w}^R$. The thresholds that determine the optimality of each scheme are functions of the discount factor and the probability of a positive aggregate factor. Figure 1 illustrates this result. As pointed out earlier, when collusion is not a concern, the optimal scheme can either be the JPE scheme ($\boldsymbol{w}^J$) or the RPE scheme ($\boldsymbol{w}^S$). Consideration of collusion results in three changes. First, there appears a region where the IPE becomes the optimal scheme. Second, the parameter range that supports the JPE contract $\boldsymbol{w}^J$ as the optimal scheme expands. Third, as alluded to earlier, the optimal RPE cheme that rules out collusion, $\boldsymbol{w}^R$, requires the principal to pay a higher compensation. As a consequence, the parameter range that supports an RPE contract as the optimal scheme narrows down.

To see why a specific scheme becomes optimal, let's first consider the relative benefits and costs of each scheme. The JPE scheme $\boldsymbol{w}^J$ results in cooperation of the employees, where the employees peer-monitor each other and deter deviations from

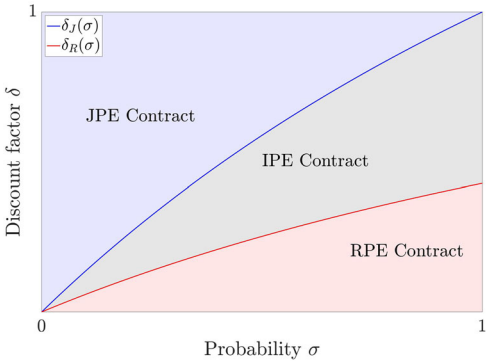


**FIGURE 1** Optimal contract without automation depending on $\sigma$ and $\delta$. IPE, independent performance evaluation; JPE, joint performance evaluation; RPE, relative performance evaluation [Color figure can be viewed at wileyonlinelibrary.com]

repeated joint work by the threat of initiating (shirk, shirk) afterwards. Such peer sanctions become particularly effective when employees value future interactions more (high $\delta$). The RPE scheme $\boldsymbol{w}^R$, in contrast to the JPE scheme, puts the employees in competition. Since such competition creates a risk of collusion, the principal increases the stakes under $\boldsymbol{w}^R$ to make sure that this risk is eliminated. This strategy is advantageous as long as the employees do not value future payoffs too much (low $\delta$), since otherwise the required increase of stakes is too high. When $\delta$ is in the intermediate range, both the JPE and the RPE schemes lose their appeal, resulting in the IPE scheme ($\boldsymbol{w}^I$) to be optimal.

## 3.3 | When would the principal adopt automation?

Now that we know the optimal compensation structures with and without automation, we can characterize the principal's optimal automation decision.

**Proposition 2.** *The principal adopts automation if and only if* $\delta \in [\underline{\delta}, \bar{\delta}\,]$, *where*

$$\bar{\delta} \equiv \min\left\{1, \frac{\sigma(2-p_1)+(1-\sigma)p_1^2-\alpha(1-\sigma)(p_1-p_0)p_1}{p_0[\sigma+(1-\sigma)((1+\alpha)p_1-\alpha p_0)]}\right\},$$

*and*

$$\underline{\delta} \equiv \max\left\{0, \frac{(1-p_1)[\sigma-(1-\sigma)((1-\alpha)p_1+\alpha p_0)]}{p_1[\sigma+(1-\sigma)((1+\alpha)p_1-\alpha p_0)]}\right\},$$

*and* $\underline{\delta} < \bar{\delta}$. *Otherwise, if* $\delta \notin [\underline{\delta}, \bar{\delta}\,]$, *she does not adopt automation and uses*

(i) *the JPE contract* $\boldsymbol{w}^J$ *if* $\delta > \bar{\delta}$,
(ii) *the RPE contract* $\boldsymbol{w}^R$ *if* $\delta < \underline{\delta}$.

Proposition 2 states that automation with asymmetric effects is valuable when the principal cannot take advantage of the employee interaction. On the one hand, automation is attractive because of its consistent high input and low operational costs. On the other hand, keeping the peer is attractive as the principal benefits from mutual monitoring and employee interaction. When automation diminishes the principal's monitoring capacity, the remaining employee can become more costly to incentivize. As a result, the principal may choose not to adopt automation but to keep the peers. When she does, however, she only uses a JPE or an RPE scheme, but not IPE.

Figure 1,2 demonstrates how the preference for automation and the choice of optimal contract vary with $\sigma$ and $\delta$. Compared to the case without automation, when a firm optimally adopts asymmetric automation, three changes take

place. First, when the IPE scheme is the optimal compensation, the principal would always prefer to adopt automation and displace an employee. In this case, there is no indirect cost of automation, as there was no peer interaction to begin with. Second, when the optimal compensation scheme is JPE, the principal would choose to keep the employees only when $\delta$ is high and $\sigma$ is low. This is because the indirect cost of automation of removing the interaction is high when employees care about their future earnings, and to earn success they must work. Third, when the RPE scheme is optimal, the principal also keeps employees and chooses not to automate if $\sigma$ is sufficiently high and $\delta$ is sufficiently low. In these cases, the indirect cost of automation is again high, however, for different reasons. Specifically, when $\sigma$ is high, having a better performance measure compared to one's peer is a clear indicator of working. And automation amounts to losing this information, resulting in excessive pay. In such circumstances, the principal prefers not to adopt automation and uses an RPE scheme. RPE is also more likely to survive when $\delta$ is low, since in these cases collusion is a less serious concern and the optimal RPE is relatively cheaper.[6]

## 4 | EXTENSIONS

This section proposes two extensions to affirm the robustness of our baseline analysis. We first consider a setting in which automation affects both employees symmetrically rather than replacing one of them (Section 4.1). Next, we introduce an alternate model of teamwork with complementarities between the efforts of the team members (Section 4.2). We show that our main message carries through both settings, and automation adoption may result in suboptimal outcomes despite the direct benefits that it provides. Proofs of this section are given in Appendix B.

### 4.1 | Symmetric effects of automation

Thus far, we modeled automation to have asymmetric effects over employees—it displaces one of the employees while the remaining employee stays unaffected. However, it is possible that automation impacts all employees similarly, without displacing any of them. In this extension, we consider this possibility by focusing on an alternate setting where automation adoption does not change the number of employees, but impacts the production team in two distinct ways. First, the cost of work becomes $c_A$ for the employees, where $c_A < c$. Second, automation boosts the efficiency of the system, and as a result the likelihood of a good aggregate factor becomes $\sigma_A > \sigma$. Thus, automation introduces a direct cost efficiency into the system by reducing the cost of effort, but it narrows down the monitoring ability of the principal as well.

To characterize the optimal automation decision in this case, we need to know the optimal compensation scheme that the principal chooses when she adopts automation. The characterization of the optimal contract here follows Proposition 1, except $c$ is replaced by $c_A$ and $\sigma$ is replaced by $\sigma_A$. The question is, then, under what circumstances is it optimal to adopt automation, and how does this decision depend on the underlying contract that the principal utilizes when automation is not feasible? The next proposition provides a perspective on this.

**Proposition 3.** *The principal adopts automation in all regions where previously RPE contracts were optimal. In other regions (when IPE or JPE contracts were optimal), the principal may choose not to adopt automation.*

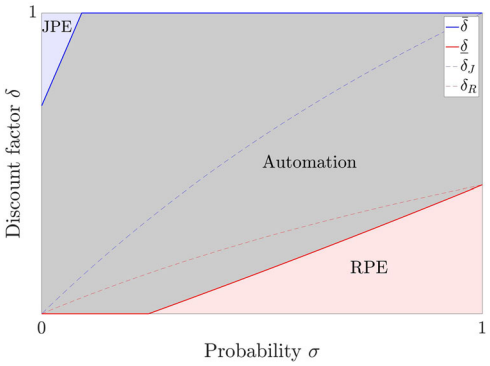


**FIGURE 2** Optimal automation adoption decision. Solid lines characterize the automation decision. Dashed lines replicate the lines given in Figure 1. JPE, joint performance evaluation; RPE, relative performance evaluation [Color figure can be viewed at wileyonlinelibrary.com]

Automation introduces a direct advantage for the principal, as it reduces the employees' cost of effort. But it also creates a potential disadvantage for her, as it negatively affects her monitoring ability. Proposition 3 stems from the fact that automation adoption affects the costs of RPE, JPE, and IPE contracts disproportionally. First, RPE contracts always become cheaper after automation adoption. This is because, under RPE contracts, the reduction on principal's monitoring capacity does not affect her as she only pays the employees when one of them has a better performance measure than the other. As a result, automation has only positive effects on RPE contracts. This explains the first part of the proposition. Second, JPE, and IPE contracts may become costlier after automation as the reduced monitoring capacity of the principal negatively affects her, since she pays the employees when they have a favorable evaluation. As a result, the principal may abstain from adopting automation in these circumstances.

Proposition 3 was about adoption of automation, and it did not provide a full picture of the contracts that the principal would use in case she adopts automation. It turns out with automation, the reduction in the cost of a JPE or an IPE contract, if any, is always less than that of an RPE contract. As a result, a principal who was using an RPE contract before automation will always adopt automation and continue to use the RPE contract. Moreover, one that is using a JPE or an IPE contract, if adopting automation, may switch to the RPE contract. Proposition 4 summarizes these arguments.

**Proposition 4.** *Differently from when automation has asymmetric effects, with symmetric automation, the parameter space over which RPE contracts are preferred expands.*

## 4.2 | Effort complementarity

In the previous sections, we have abstracted away from the complementarity of the effort of the employees, which is a natural characteristic of teamwork. In this section, we will focus on how the complementarity between peers' effort affects the principal's preference for automation. Since the baseline model is developed to study contracts, it is not ill-suited to investigate the effect of complementarity. Thus, we consider an alternative framework with the following features. First, we assume that peer effort is complementary. Second, we allow employees to choose their efforts from a continuum of options. Third, we allow the principal to adjust the effort provided by the machine in case automation is adopted. Specifically, each employee chooses an effort level from a continuum, $e \in [0, 1]$. The cost of effort is assumed to be a quadratic function of the effort level, $c(e) = \frac{c}{2}e^2$. Here, as it was in the asymmetric automation case, automation replaces an employee with a machine whose effort $e_m \in [0, 1]$ can be adjusted by the principal, and costs the firm $c_m(e_m) = \frac{\alpha c}{2}e_m^2$, where $\alpha \in [0, 1]$. Thus, automation offers higher efficiency as it was in the baseline model.

We assume that the efforts of the employees remain unobservable to the principal and jointly determine the output of the firm, which is a failure or a success. Formally, an effort profile $(e_1, e_2)$ by employees 1 and 2, respectively, results in success with probability $f(e_1, e_2) = pe_1e_2$ and in failure with the remaining probability. When production is automated instead, the probability of success is $f(e, e_m) = pee_m$, where $e$ is the effort of the employee and $e_m$ is the effort of the machine. We normalize the payoff of the principal from success and failure to 1 and 0, respectively. The principal chooses the production regime and the compensation scheme to maximize her expected profit and evaluates an employee's performance based on the outcome. This functional form indicates that the effort of both employees are crucial for production: success cannot be achieved if one member puts in no effort, regardless of the effort of the other member.

### 4.2.1 | Optimal production plan under automation

Under automation, the principal specifies an amount $w$ to pay the employee in case of success, and chooses the effort level of the machine, that is, $e_m$. Given $w$ and $e_m$, the employee decides his effort level $e$.

**Proposition 5.** *In an automated system, the optimal production plan is characterized by*:

$$(e, e_m) = \begin{cases} (0,0) & \text{if } \kappa < \sqrt{2\alpha} \\ \left(\frac{\kappa}{2}, 1\right) & \text{if } \kappa \in [\sqrt{2\alpha}, 2] \\ (1,1) & \text{if } \kappa > 2 \end{cases} \qquad w = \begin{cases} \frac{1}{2} & \text{if } \kappa \leq 2 \\ \frac{1}{\kappa} & \text{if } \kappa > 2 \end{cases}$$

Under automation, the productivity threshold under which the principal prefers to shut down production is $\sqrt{2\alpha}$. Only above this threshold can she anticipate a positive profit. In this case, it is always optimal to adjust the machine effort to its maximal capacity. However, the principal may not always want the employee to exert full effort, because to motivate maximal effort she has to leave a positive rent. When the productivity ($\kappa$) is sufficiently high, it becomes profitable for the principal to leave rent and induce the maximal effort.

### 4.2.2 | Optimal production plan without automation

Without automation, the principal sets a compensation $w$ to pay each employee in case of success. As before, we adopt the collusion-free equilibrium as the solution concept. Specifically, given the choice of $w$, the strategy profile that the employees follow comprises a collusion-free equilibrium. In the appendix, we show that the corresponding collusion-free equilibrium outcome of the repeated interaction following the principal's choice of $w$ can only be in two different forms: either the employees exert their maximal effort all the time, or they exert no effort at all. That is, it is not possible to have a collusion-free equilibrium outcome in which the employees choose an interior effort level. Therefore, the principal's problem reduces to finding a $w$ that induces $(1,1)^\infty$ as a collusion-free equilibrium outcome, conditional on keeping operations open. In such an equilibrium, deviations trigger a continuation play in which the employees choose their minimal efforts forever, which is self-enforcing. This is the most extreme punishment that the employees can impose on each other.

Let $\mathbf{e_i}(w, e_j)$ be the best response of employee $i$, conditional on wage $w$ and the effort choice of his peer $e_j$, which follows from the following problem:

$$\max_{e_i} pe_i e_j w - \frac{c}{2} e_i^2.$$

The first-order condition leads to $\mathbf{e_i}(w, e_j) = min\{1, \kappa w e_j\}$. Due to the symmetry, we suppress the notation, and use $\mathbf{e}(w, e)$ instead. In an optimal production plan that keeps the production open, the following condition must be satisfied:

$$wp - \frac{c}{2} \geq (1-\delta)\left[wp\mathbf{e}(w,1) - \frac{c}{2}(\mathbf{e}(w,1))^2\right].$$

This incentive constraint ensures that the employees do not deviate from exerting the maximal effort. The left-hand side of the inequality is the expected utility of an employee on the equilibrium path, whereas the right-hand side is his expected payoff resulting from deviation. In the current period, the employee deviates to his best response, and receives a positive payoff. In the remaining periods, he gets punished and receives a zero payoff. If the best response effort level is equal to 1, then this incentive constraint is automatically satisfied. Therefore, for this condition to be nonredundant, we must have $\mathbf{e}(w,1) = \min\{1, \kappa w\} = \kappa w$. In this case, the above incentive constraint becomes:

$$wp - \frac{c}{2} \geq (1-\delta)\frac{w^2 p^2}{2c}.$$

The principal's problem can be written as follows:

$$\max_w (1-2w)p \ \text{ s.t. } \ wp - \frac{c}{2} \geq (1-\delta)\frac{w^2 p^2}{2c}$$

The objective function is equal to $(1-2w)p$, because both employees will get paid $w$ if a success takes place. From this problem, it is evident that the principal chooses the minimum $w$ satisfying the incentive constraint. If this $w$ brings

a positive profit, she keeps production running and induces the maximal effort from the employees; otherwise she closes the production. The next proposition formalizes this intuition.

**Proposition 6.** *The optimal production plan without automation satisfies*:

$$(e_1, e_2) = \begin{cases} (0,0) & \text{if } \kappa \leq \dfrac{2}{1+\sqrt{\delta}} \\ (1,1) & \text{if } \kappa > \dfrac{2}{1+\sqrt{\delta}} \end{cases} \qquad w = \begin{cases} 0 & \text{if } \kappa \leq \dfrac{2}{1+\sqrt{\delta}} \\ \dfrac{1}{\kappa(1+\sqrt{\delta})} & \text{if } \kappa > \dfrac{2}{1+\sqrt{\delta}} \end{cases}$$

We already know that the principal either closes the production or induces maximal effort from both employees as a collusion-free equilibrium outcome. Proposition 6 characterizes the critical value of the effective productivity, below which shutting production down is the optimal choice for the principal. This critical value decreases with $\delta$, because if the employees are more patient, the punishment that they impose on each other is more deterrent.

### 4.2.3 | Optimal automation adoption

Given the described production plans, the firm profit with and without automation ($\pi_a$, and $\pi_h$, respectively) are:

$$\pi_a = \begin{cases} 0 & \text{if } \kappa < \sqrt{2\alpha} \\ \dfrac{p\kappa}{4} - \dfrac{\alpha c}{2} & \text{if } \kappa \in [\sqrt{2\alpha}, 2] \\ p - \dfrac{(2+\alpha)c}{2} & \text{if } \kappa > 2 \end{cases} \qquad \pi_h = \begin{cases} 0 & \text{if } \kappa \leq \dfrac{2}{1+\sqrt{\delta}} \\ p - \dfrac{2c}{(1+\sqrt{\delta})} & \text{if } \kappa > \dfrac{2}{1+\sqrt{\delta}} \end{cases}$$

Proposition 7 follows from the comparison of the profits given above.

**Proposition 7.** *When effort of team members are complementary, the principal adopts automation if* $\delta < \frac{(2-\alpha)^2}{(2+\alpha)^2}$, *and keeps the employees otherwise.*

As in the baseline model, with complementarity, automation may still not be preferred in all circumstances—despite its higher efficiency. To see the intuition, suppose that the effective productivity is sufficiently large, that is, $\kappa > 2$, so that the principal optimally induces the maximal effort from the employees regardless of her automation adoption decision. In this case, if the discount factor is greater than $\frac{(2-\alpha)^2}{(2+\alpha)^2}$, the principal keeps the employees since they are sufficiently patient and the punishment they impose on each other prevents shirking. As the cost of operation under automation increases, the discount factor $\delta$ at which this switch occurs becomes smaller. Thus, if automation offers less efficiency, the principal has a higher willingness to keep employees.

## 5 | CONCLUSION

Automation of tasks is expected to radically transform work environments (Frey & Osborne, 2017). Motivated with this expectation, this paper asks two simple but important questions: When automation displaces a worker in a team of workers and has clear operational benefits, will a principal always prefer to adopt it? In which settings is adoption of automation more likely?

Our findings highlight an indirect cost of automation adoption that has not been highlighted before in the literature. Automation may cause the principal to lose an important tool to manage the agency problem, which is her ability to shape interactions between the workers. As highlighted by Hölmstrom (1982) and Che and Yoo (2001), interactions between team members equip the principal with a tool to write compensation contracts and motivate workers to work. In the absence (or in environments with partial loss) of team interactions, the cost of incentivizing the

remaining worker is higher for a principal, making the overall system more costly to operate in some cases. The overall message of this paper is, therefore, that despite the direct benefits of automation, its adoption may not always lower operating costs in a work setting. This message is robust to the consideration of automation effects that do not displace a worker, but rather reduce the workload of all members, as well as when there are complementarities between the effort choices of team members.

In a broad stroke, the literature focusing on automation generally highlights its benefits: automation can increase productivity (Gaimon, 1985), reduce workers' workload, and reduce error rate and waste (Wilson, 2016). At the same time, two downsides have been suggested. At a macro level, automation can displace workers at a substantial rate, resulting in wage losses or loss of employment (Acemoglu & Restrepo, 2020, 2018). At a microlevel, consumers may suffer from algorithm aversion—reject decisions from an automated system (Dietvorst et al., 2015, 2018). Another potential concern regarding automation in a work setting is its impact on the underlying agency problem. Even though automation reduces the scope of moral hazard problem, it may also reduce the number of tools at the discretion of a principal to manage it. This finding generates an empirical question for labor economists actively working on the implications of automation. Specifically, findings in the labor literature point out to an increase in the wages of high-skill labor, who are also more likely to be retained after automation (Acemoglu & Restrepo, 2020). While the explanation may be the increase in productivity due to automation, our study highlights a second channel which may result in an increase: reduction in the principal's capacity of handling the agency problem.

A second outcome of our analysis concerns what types of firms may be more likely to adopt automation. In our baseline model, we find that firms that have optimally chosen to utilize IPE and JPE contracts are more likely to benefit from automation, therefore they choose to introduce it to their work environment. Firms operating with RPE contracts are less likely to adopt automation. With symmetric automation, the appeal of RPE contracts expands. Put differently, an empirical analysis that compares workers' contracts "before and after" the growth of automation over a period of time may find the share of competitive compensation contracts go up. We leave this as an interesting future exercise to interested researchers.

For managers, our findings have a simple call: think beyond the immediate benefits of automation. As automation is finding increasing use in marketing, operations, finance, and law, the managers of these work environments will have to make the call about whether to replace an employee with a machine. For instance, a call center employee may be replaced with an automated answering system. If multiple call center employees make a team and their interactions are valuable to the principal, should she choose to replace one of them with a machine? Our study suggests that automation may result in lower performance of the remaining employee unless he is adequately compensated, increasing the cost of the overall system. We hope that future research can build on the worker interaction implications of automation that we highlight.

## ACKNOWLEDGMENTS

This study was generously funded by the Wharton Dean's Research Fund and the Mack Institute of the University of Pennsylvania. A previous version of this paper was circulated under the title "Man versus Machine: When is automation inferior to human labor?" We thank the editor, the coeditor, and the anonymous referees for their valuable feedback. We also thank Chris Dellarocas, Anthony Dukes, Avi Goldfarb, Hanna Halaburda, Lorin Hitt, George Mailath, Steven Matthews, Mallesh Pai, Jiwoong Shin, K. Sudhir, Christophe Van den Bulte, Senthil Veeraraghavan, and John Zhang for their valuable feedback.

## ORCID

*Mustafa Dogan* https://orcid.org/0000-0001-9313-8911
*Pinar Yildirim* https://orcid.org/0000-0002-6667-9365

## ENDNOTES

[1]Examples of environments where humans and robots are in a team are rising in practice. For instance, BMW's manufacturing facilities are being transformed with humans collaborating with machines (Knight, 2020). Similarly, examples of human–robot teams can be seen in military (Military, 2020), as well as in medicine (Lohmeyer, 2020). Human-robot teams and machine collaboration also make up of an academic area of investigation, focusing on the development and performance improvement of human-robot teams (e.g., Lewis et al., 2010; Saenz et al., 2020; Shah et al., 2011).

[2]Note that this formulation requires the wage structure to be symmetric for the employees. This restriction, however, is innocuous given that the employees are identical.

[3]Note that JPE, RPE and IPE do not cover the entire space of contracts. For instance, a compensation scheme satisfying $w_{11} > w_{10}$, and $w_{01} < w_{00}$ is none of JPE, RPE, or IPE.

[4]The way these constraints are written assumes that any deviation from AS is punished by a repeated play of (work, work). While, in principle, precluding AS from being an SPE outcome requires considering all possible punishments, in the appendix we show that it suffices to consider repeated play of (work, work).

[5]In fact, the actual value of $w_{10}^R$ needs to be slightly larger than $\frac{\hat{c}}{1-(1+\delta)p_1}$. When $w_{10}^R$ is exactly equal to $\frac{\hat{c}}{1-(1+\delta)p_1}$, only a weak inequality version of the relevant constraint ($\mathcal{IC}_1^{AS}$) (which is based on a strict inequality) holds.

[6]While the derivations would be messier, the results of our study could be replicated for teams with more than two people. In such a model, we would create a signal structure that maintains the agency problem and also allow for loss of team interaction—which we show to be an effective tool to manage the agency problem—when there are fewer individuals in the production process. These two assumptions in model construction should result in similar qualitative findings.

[7]As the constraint ($\mathcal{IC}_1^{AS}$) is based on a strict inequality, the actual value of $w_{10}$ is slightly larger than $\frac{\hat{c}}{1-(1+\delta)p_1}$.

[8]A necessary condition for $\boldsymbol{w}^R$ to solve ($\mathcal{P}_{AS}$) is $[1-(1+\delta)p_1] > 0$ since otherwise $\boldsymbol{w}^I$ would dominate $\boldsymbol{w}^R$. Therefore, $w_{10}^R$ is well defined without loss of generality.

[9]Note that, $1-\delta-2\delta^2-(1-\delta^2)p_0+(\delta+\delta^2)p_1 > 0$ must hold, otherwise increasing $w_{00}$ does not help to satisfy ($\mathcal{IC}_{1'}^{AS}$), contradicting with $w_{00} > 0$.

[10]It must be $(\delta+\delta^2)p_1-(1-\delta^2)p_0-\delta^2 > 0$ as otherwise $A$ would be negative, contradicting with $w_{00} > 0$.

[11]These expressions are based on the fact that $\mathbf{e}_i(\boldsymbol{w}, e_1) = min\{1, w\kappa e_j\} = w\kappa e_j$. The second equality follows from the fact that $\mathbf{e}(\boldsymbol{w}, 1) = min\{1, w\kappa\} = w\kappa$.

## REFERENCES


Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, *108*(6), 1488–1542.

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

Agrawal, A., Gans, J., & Goldfarb, A. (2016). The simple economics of machine intelligence. Harvard Business Review.

Agrawal, A., Gans, J. S., & Goldfarb, A. (2017a). *Exploring the impact of artificial intelligence: Prediction versus judgment*. University of Toronto and NBER working paper.

Agrawal, A. K., Gans, J. S., & Goldfarb, A. (2017b). What to expect from artificial intelligence. *MIT Sloan Management Review*, *58*(3), 23.

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American Economic Review*, *62*(5), 777–795.

Aral, S., Brynjolfsson, E., & Wu, L. (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science*, *58*(5), 913–931.

Baker, G., Gibbons, R., & Murphy, K. J. (2002). Relational contracts and the theory of the firm. *Quarterly Journal of Economics*, *117*(1), 39–84.

Bonatti, A., & Hörner, J. (2011). Collaborating. *American Economic Review*, *101*(2), 632–63.

Bonatti, A., & Rantakari, H. (2016). The politics of compromise. *The American Economic Review*, *106*(2), 229–259.

Brynjolfsson, E., & McAfee, A. (2011). Race against the machine: how the digital revolution is accelerating innovation. *Driving Productivity, and Irreversibly Transforming Employment and The Economy*. Digital Frontier Press.

Chan, T. Y., Li, J., & Pierce, L. (2014). Compensation and peer effects in competing sales teams. *Management Science*, *60*(8), 1965–1984.

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

Choi, Y. K. (1993). Managerial incentive contracts with a production externality. *Economics Letters*, *42*(1), 37–42.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170.

Dogan, M., Jacquillat, A., & Yildirim, P. (2021). Strategic automation and decision-making authority. Available at SSRN 3226222.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*(January), 254–280.

Gaimon, C. (1985). The optimal acquisition of automation to enhance the productivity of labor. *Management Science*, *31*(9), 1175–1190.

Gibbons, R., & Murphy, K. J. (1990). Relative performance evaluation for chief executive officers. *ILR Review*, *43*(3), 30–S.

Hölmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, *13*(2), 324–340.

Knight, W. (2020). How human-robot teamwork will upend manufacturing.

Lewis, M., Wang, H., Chien, S.-Y., Scerri, P., Velagapudi, P., Sycara, K., & Kane, B. (2010). Teams organization and performance in multi-human/multi-robot teams. In *2010 IEEE International Conference on Systems, Man and Cybernetics*, IEEE (pp. 1617–1623).

Lohmeyer, R. (2020). Human-robot interaction in the O.R.: How surgeons and medical robots can work together.

Mobius, M., & Schoenle, R. (2006). *The evolution of work*. Technical report, National Bureau of Economic Research.
Moriarty, R. T., & Swartz, G. S. (1989). Automation to boost sales and marketing, Volume January-February. Harvard Business Review.
Rayo, L. (2007). Relational incentives and moral hazard in teams. *The Review of Economic Studies*, *74*(3), 937–963.
Reed, K. B., & Peshkin, M. A. (2008). Physical collaboration of human-human and human-robot teams. *IEEE Transactions on Haptics*, *1*(2), 108–120.
Saenz, M. J., Revilla, E., & Simón, C. (2020). Designing AI systems with human-machine teams. *MIT Sloan Management Review*, *61*(3), 1–5.
Shah, J., Wiken, J., Williams, B., & Breazeal, C. (2011). Improved human-robot team performance using CHASKI, a human-inspired plan execution system. In *Proceedings of the 6th International Conference on Human-robot Interaction* (pp. 29–36).
US Military Academy West Point (2020). Human-robot teaming.
Venkatraman, N. (1994). IT-enabled business transformation: From automation to business scope redefinition. *Sloan Management Review*, *35*(2), 73.
Weng, Q., & Carlsson, F. (2015). Cooperation in teams: The role of identity, punishment, and endowment distribution. *Journal of Public Economics*, *126*, 25–38.
Wilson, A. (2016). Drilling automation saves rig time and safeguards against human error. *Journal of Petroleum Technology*, *68*(09), 95–97.

**How to cite this article:** Dogan, M., & Yildirim, P. (2021). Managing Automation in Teams. *Journal of Economics & Management Strategy*, 1–25. https://doi.org/10.1111/jems.12456

## APPENDIX A: PROOFS OF THE STATEMENTS FROM SECTION 3

We find it more convenient to provide the proof of Lemma 3 before the proof of Lemma 2.

*Proof of Lemma* 3. The problem $(\mathcal{P}_{AS})$ is defined as follows:

$$\min_{w_{11},w_{10}} [\sigma + (1-\sigma)p_1]w_{11} + [(1-\sigma)p_1(1-p_1)]w_{10} \quad \text{s.t.} \quad (\mathcal{IC}_0), (\mathcal{IC}^{AS}), w_{11} \le w_{10},$$
$$\text{where } (\mathcal{IC}^{AS}) \equiv \left(\mathcal{IC}_1^{AS}\right) \vee \left(\mathcal{IC}_{1'}^{AS}\right) \vee \left(\mathcal{IC}_2^{AS}\right).$$

When $w_{11} \le w_{10}$, we can rewrite $(\mathcal{IC}_0)$, $(\mathcal{IC}_1^{AS})$, $(\mathcal{IC}_{1'}^{AS})$, and $(\mathcal{IC}_2^{AS})$ as follows:

$$w_{10} + p_1(w_{11} - w_{10}) \ge \hat{c}, \qquad \left(\mathcal{IC}_0\right)$$
$$w_{10} + (1+\delta)p_1(w_{11} - w_{10}) > \hat{c} \qquad \left(\mathcal{IC}_1^{AS}\right)$$
$$[(\delta+\delta^2)p_1 - (1-\delta^2)p_0]w_{11} + [(\delta+\delta^2-1) - (\delta+\delta^2)p_1 + (1-\delta^2)p_0]w_{10} > (\delta+\delta^2-1)\hat{c}, \qquad \left(\mathcal{IC}_{1'}^{AS}\right)$$
$$w_{10} + 2p_1(w_{11} - w_{10}) \ge \hat{c}. \qquad \left(\mathcal{IC}_2^{AS}\right)$$

Since $w_{11} \le w_{10}$, satisfying $(\mathcal{IC}_2^{AS})$ is always costlier than satisfying $(\mathcal{IC}_1^{AS})$ as $2p_1 > (1+\delta)p_1$. Therefore, fulfilling $(\mathcal{IC}^{AS})$ in problem $(\mathcal{P}_{AS})$ requires to satisfy either $(\mathcal{IC}_1^{AS})$ or $(\mathcal{IC}_{1'}^{AS})$. Note also that, $(\mathcal{IC}_{1'}^{AS})$ depends on $\delta$, and it can be further written as follows:

$$\left(\mathcal{IC}_{1'}^{AS}\right) = \begin{cases} w_{10} + \left(\frac{1-\delta^2}{1-\delta-\delta^2}p_0 - \frac{\delta+\delta^2}{1-\delta-\delta^2}p_1\right)(w_{11} - w_{10}) < \hat{c} & \text{if } \delta+\delta^2 < 1, \\ w_{11} - w_{10} > 0 & \text{if } \delta+\delta^2 = 1, \\ w_{10} + \left(\frac{\delta+\delta^2}{\delta+\delta^2-1}p_1 - \frac{1-\delta^2}{\delta+\delta^2-1}p_0\right)(w_{11} - w_{10}) > \hat{c} & \text{if } \delta+\delta^2 > 1. \end{cases} \tag{A1}$$

The solution to $(\mathcal{P}_{AS})$ lies in the shaded region in Figure A1. All the lines that define the constraints $(\mathcal{IC}_0)$, $(\mathcal{IC}_1^{AS})$, and $(\mathcal{IC}_{1'}^{AS})$ pass through the point $(w_{11}, w_{10}) = (\hat{c}, \hat{c})$, and they never intersect again since each has a different slope. The frontier of the constraint $(\mathcal{IC}_0)$ is the line connecting points $(\hat{c}, \hat{c})$ and $(0, 1/(1-p_1))$, and that an incentive scheme $(w_{11}, w_{10})$ that lies on the right side of this frontier satisfies $(\mathcal{IC}_0)$. Then, the solution to $(\mathcal{P}_{AS})$ (satisfying either $[\mathcal{IC}_1^{AS}]$

or $[\mathcal{IC}_{1'}^{AS}]$ on top of $[\mathcal{IC}_0]$ and $w_{11} \leq w_{10}$) is obtained either by (i) setting $(w_{11}, w_{10}) = (\hat{c}, \hat{c})$, or by (ii) setting $w_{11} = 0$. If the former is the solution, then we have an IPE since the agents' payments are independent of each other. This scheme clearly satisfies all the constraints.

If the optimal solution is obtained by setting $w_{11} = 0$, the relevant constraint for $(\mathcal{IC}^{AS})$ is $(\mathcal{IC}_1^{AS})$, that is, $(\mathcal{IC}_{1'}^{AS})$ is redundant. To see this, we consider three cases separately.

- When $\delta + \delta^2 > 1$, $(\mathcal{IC}_{1'}^{AS})$ is costlier than $(\mathcal{IC}_1^{AS})$ as $\left(\frac{\delta+\delta^2}{\delta+\delta^2-1}p_1 - \frac{1-\delta^2}{\delta+\delta^2-1}p_0\right) > (1+\delta)p_1$.
- When $\delta + \delta^2 = 1$, $(\mathcal{IC}_{1'}^{AS})$ requires $w_{11} - w_{10} > 0$, which would violate $w_{11} \leq w_{10}$.
- When $\delta + \delta^2 < 1$, we have $(\frac{\delta+\delta^2}{\delta+\delta^2-1}p_1 - \frac{1-\delta^2}{\delta+\delta^2-1}p_0) < (1+\delta)p_1$. Suppose $(\mathcal{IC}_{1'}^{AS})$ is the relevant constraint for $(\mathcal{IC}^{AS})$. The solution then satisfies $(\mathcal{IC}_0)$ and $(\mathcal{IC}_{1'}^{AS})$ at the same time. However, the fact that $w_{11} = 0$ makes this impossible.

Hence, if the optimal solution is obtained by setting $w_{11} = 0$, then the value of $w_{10}$ must be set equal to the minimum level satisfying $(\mathcal{IC}_1^{AS})$, and $(\mathcal{IC}_0)$: $w_{10} = \frac{\hat{c}}{1-(1+\delta)p_1}$.[7]

Next, we will show that the optimal solution to principal's problem, if satisfies $w_{11} \leq w_{10}$, then coincides with the solution to $(\mathcal{P}_{AS})$.

First, we know that the incentive scheme $\boldsymbol{w}^I$ induces repeated (work, work) as a collusion-free equilibrium outcome. Now, suppose that $\boldsymbol{w}^I$ solves $(\mathcal{P}_{AS})$. If the optimal solution to principal's problem satisfies $w_{11} \leq w_{10}$, then this solution must be $\boldsymbol{w}^I$. Otherwise, we would contradict with $\boldsymbol{w}^I$ being solution to $(\mathcal{P}_{AS})$.

Second, suppose that $\boldsymbol{w}^R$ solves $(\mathcal{P}_{AS})$. If we can show that $\boldsymbol{w}^R$ induces repeated (work, work) as a collusion-free equilibrium outcome, by the same logic above, we can conclude that, if the optimal solution to principal's problem satisfies $w_{11} \leq w_{10}$, then this solution must be $\boldsymbol{w}^R$.[8] To this end, we will show that repeated (work, work) is the unique subgame perfect equilibrium (SPE) outcome under $\boldsymbol{w}^R$; hence it is also the (unique) collusion-free equilibrium outcome.

First step: It is well-known in the literature that, we can restrict attention into the equilibria in which the agents condition their actions only on the actions of the previous periods. In principle, actions can depend on the realized performance signals as well. However, each such equilibrium can be modified into another equilibrium where the actions just depend on the earlier actions. This is because, the agents can use a public randomization device to coordinate over any probability distribution of action profiles reached by conditioning their actions on the realized signals. Any such modification of a given strategy profile (through public randomization) does not effect the incentives since the agents only care about their continuation utilities when choosing their actions.

Second step: Under $\boldsymbol{w}^R$, the agents would never shirk together in an SPE. To induce (shirk, shirk) in some period, each agent's continuation utility must be at least $U_{00}$, which satisfies:

$$(1-\delta)\pi(0,0,\boldsymbol{w}^R) + \delta U_{00} = (1-\delta)\pi(1,0,\boldsymbol{w}^R) + \delta\pi(1,1,\boldsymbol{w}^R).$$

However, under $\boldsymbol{w}^R$, there does not exist a strategy profile that delivers a continuation utility greater than or equal to $U_{00}$ for both agents. To see this, fix a strategy profile, and let $\lambda_{e_1e_2}$ be the discounted probability that the effort pair $(e_1, e_2)$ is chosen following the initial period at which the agents play (shirk, shirk). For instance, if the agents alternate between (work, shirk), and (shirk, work) afterwards, then we have $\lambda_{11} = 0$, $\lambda_{10} = \frac{1}{1+\delta}$,

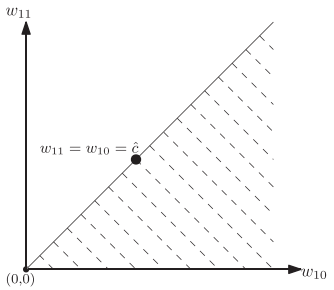


FIGURE A1 The region satisfying $w_{11} \leq w_{10}$, $w_{11} \geq 0$, and $w_{10} \geq 0$

$\lambda_{01} = \frac{\delta}{1+\delta}$, and $\lambda_{00} = 0$. It is crucial to note that, $\lambda_{11} + \lambda_{10} + \lambda_{01} + \lambda_{00} = 1$, since all the terms are multiplied with $1 - \delta$. Then the expected continuation utilities of Agent 1, and Agent 2 become:

$$U_1 = \lambda_{11}[\pi(1, 1, \boldsymbol{w}^R) - c] + \lambda_{10}[\pi(1, 0, \boldsymbol{w}^R) - c] + \lambda_{01}\pi(0, 1, \boldsymbol{w}^R) + \lambda_{00}\pi(0, 0, \boldsymbol{w}^R)$$
$$U_2 = \lambda_{11}[\pi(1, 1, \boldsymbol{w}^R) - c] + \lambda_{01}[\pi(1, 0, \boldsymbol{w}^R) - c] + \lambda_{10}\pi(0, 1, \boldsymbol{w}^R) + \lambda_{00}\pi(0, 0, \boldsymbol{w}^R)$$

By using the fact that $\lambda_{00} = 1 - \lambda_{11} - \lambda_{10} - \lambda_{01}$, we obtain:

$$U_1 + U_2 - 2U_{00} = \frac{2c}{(p_1 - p_0)(1 - (1+\delta)p_1)}\left[\lambda_{11}(\delta p_1 - p_0) + (\lambda_{10} + \lambda_{01})[(1+\delta)p_1 - 2p_0] - \frac{2(p_1 - p_0)}{\delta}\right].$$

It is clear that $U_1 + U_2 - 2U_{00} < 0$ regardless of the $\lambda_{11}$, $\lambda_{10}$, and $\lambda_{01}$ values. Therefore, there does not exist a strategy profile bringing a continuation utility higher than $U_{00}$ to both agents at the same time. Hence the agents never shirk together under $\boldsymbol{w}^R$.

Third step: The action pair (shirk, work) (or [work, shirk]) never appears in an SPE when the incentive scheme is $\boldsymbol{w}^R$. ($\mathcal{IC}_1^{AS}$) implies that, to convince the first agent to stick with (shirk, work), his continuation utility must larger than or equal to $U_{01}$, where:

$$U_{01} = \frac{1}{1+\delta}[\pi(1, 0, \boldsymbol{w}^R) - c] + \frac{\delta}{1+\delta}[\pi(0, 1, \boldsymbol{w}^R)].$$

However, as we shall see, when an agent's continuation utility is larger than $U_{01}$, it is impossible to sustain the other agent's incentive constraints. Recall by definition of $\boldsymbol{w}^R$ that:

$$\pi(1, 1, \boldsymbol{w}^R) - c = \frac{\delta}{1+\delta}[\pi(1, 0, \boldsymbol{w}^R) - c] + \frac{1}{1+\delta}[\pi(0, 1, \boldsymbol{w}^R)].$$

Then, by using the facts $\lambda_{00} = 0$, and $\lambda_{11} = 1 - \lambda_{10} - \lambda_{01}$, we obtain:

$$U_1 = \frac{\delta + \lambda_{10} - \delta\lambda_{01}}{1+\delta}[\pi(1, 0, \boldsymbol{w}^R) - c] + \frac{1 - \lambda_{10} + \delta\lambda_{01}}{1+\delta}\pi(0, 1, \boldsymbol{w}^R),$$
$$U_2 = \frac{\delta + \lambda_{01} - \delta\lambda_{10}}{1+\delta}[\pi(1, 0, \boldsymbol{w}^R) - c] + \frac{1 - \lambda_{01} + \delta\lambda_{10}}{1+\delta}\pi(0, 1, \boldsymbol{w}^R).$$

Thus, keeping the first agent's payoff higher than $U_{01}$ requires $\lambda_{10} - \delta\lambda_{01} > 1 - \delta$. In consequence, the other agent's payoff must be smaller than $\pi(1, 1, \boldsymbol{w}^R) - c$, which is the maximum value that $U_2$ can take conditional on $\lambda_{10} - \delta\lambda_{01} \geq 1 - \delta$ (achieved when $\lambda_{10} = \frac{1}{1+\delta}$, $\lambda_{01} = \frac{\delta}{1+\delta}$, and $\lambda_{11} = 0$).

Suppose the agents play (shirk, work) at period 1 (without loss of generality). We will obtain a contradiction by showing that no continuation play can support this play in an SPE.

First, we introduce some auxiliary notation: Let $\lambda^2_{e_1,e_2}$ be the probability that agents play $(e_1, e_2)$ at $t = 2$ conditional on having played (shirk, work) at period 1. More generally, $\lambda^t_{e_1,e_2}$ is the probability that the agents $(e_1, e_2)$ play $t$ conditional on having played (shirk, work) in all previous periods. That is, if probability of (shirk, work) is 0 at some period $\{1, ..., t-1\}$, then $\lambda^t_{e_1,e_2} = 0$. Second, there is no loss of generality to assume that, once agents play (work, work) in a period, they continue to do so afterwards. Third, we know from the above discussion that, if (work, shirk) is played at some period, then the first agent's continuation payoff must be lower than $\pi(1, 1, \boldsymbol{w}^R) - c$. Putting these together, we know that the first agent's continuation utility starting from $t = 2$ is:

$$U_1 < \sum_{t=2}^{\infty}\Lambda_{01}^{t-1}\delta^{t-2}\lambda_{11}^t[\pi(1, 1, \boldsymbol{w}^R) - c] + \sum_{t=2}^{\infty}\Lambda_{01}^{t-1}\delta^{t-2}\lambda_{01}^t(1-\delta)\pi(0, 1, \boldsymbol{w}^R)$$
$$+ \sum_{t=2}^{\infty}\Lambda_{01}^{t-1}\delta^{t-2}\lambda_{10}^t[(1-\delta)(\pi(1, 0, \boldsymbol{w}^R) - c) + \delta(\pi(1, 1, \boldsymbol{w}^R) - c)],$$

where $\Lambda_{01}^t = \prod_{k=1}^t \lambda_{01}^k$ is the probability that agents play (shirk, work) in each period between the first period, and period $t$, that is, $\Lambda_{01}^1 = 1$. Then by using, $\lambda_{10}^t = 1 - \lambda_{11}^t - \lambda_{01}^t$, and $\pi(1,1,\boldsymbol{w}^R) - c = \frac{\delta}{1+\delta}[\pi(1,0,\boldsymbol{w}^R) - c] + \frac{1}{1+\delta}[\pi(0,1,\boldsymbol{w}^R)]$, we rewrite this inequality:

$$U_1 < \sum_{t=2}^{\infty}\Lambda_{01}^{t-1}\delta^{t-2}\left(\frac{1}{1+\delta} - \frac{1-\delta}{1+\delta}\lambda_{11}^t - \frac{1}{1+\delta}\lambda_{01}^t\right)[\pi(1,0,\boldsymbol{w}^R) - c]$$
$$\sum_{t=2}^{\infty}\Lambda_{01}^{t-1}\delta^{t-2}\left(\frac{\delta}{1+\delta} + \frac{1-\delta}{1+\delta}\lambda_{11}^t + \frac{1-\delta^2-\delta}{1+\delta}\lambda_{01}^t\right)\pi(0,1,\boldsymbol{w}^R).$$

We know that $\Lambda_{01}^t = \Lambda_{01}^{t-1}\lambda_{01}^t$, therefore:

$$U_1 < \left(\frac{1}{1+\delta} - \frac{1-\delta}{1+\delta}\lambda_{11}^2 - \sum_{t=3}^{\infty}\Lambda_{01}^{t-1}\delta^{t-2}\left(\frac{1-\delta}{\delta(1+\delta)} + \frac{1-\delta}{1+\delta}\lambda_{11}^t\right)\right)[\pi(1,0,\boldsymbol{w}^R) - c]$$
$$+ \quad \left(\frac{\delta}{1+\delta} + \frac{1-\delta}{1+\delta}\lambda_{11}^2 + \sum_{t=3}^{\infty}\Lambda_{01}^{t-1}\delta^{t-2}\left(\frac{1-\delta}{\delta(1+\delta)} + \frac{1-\delta}{1+\delta}\lambda_{11}^t\right)\right)\pi(0,1,\boldsymbol{w}^R).$$

The right-side of this inequality is maximized by setting $\lambda_{11}^2 = \lambda_{01}^2 = 0$, and is equal to $U_{01}$ (since $\lambda_{01}^2 = 0$, $\Lambda_{01}^t = 0$ for each $t \geq 2$). Therefore, it is not possible to convince the first agent for (shirk, work) at $t = 1$, while respecting the other agent's incentives for the future periods. Hence, repeated (work, work) is the unique SPE, and is the (unique) collusion-free equilibrium under $\boldsymbol{w}^R$. □

*Proof of Proposition* 2. This immediately follows from the comparison between the optimal production plans for human and automated production teams. □

*Proof of Lemma* 2. The proof of Lemma 3 assumes $w_{01} = w_{00} = 0$. Here, we show that this assumption is indeed correct. Specifically, we show that all the problems that we obtain without assuming $w_{00} = w_{01} = 0$ are equivalent to their counterparts that assume $w_{00} = w_{01} = 0$.

Recall Problem $(\mathcal{P}_0)$, which is a relaxed version of the principal's problem:

$$\min_{\boldsymbol{w}} \quad \left[\sigma + (1-\sigma)p_1^2\right]w_{11} + (1-\sigma)p_1(1-p_1)(w_{10} + w_{01}) + (1-\sigma)(1-p_1)^2 w_{00}$$
$$\text{s.t. } (\mathcal{IC}_0): \frac{1}{1-\delta}(\pi(1,1,\boldsymbol{w}) - c) \geq \pi(0,1,\boldsymbol{w}) + \frac{\delta}{1-\delta}\min\{\pi(0,0,\boldsymbol{w}), \pi(0,1,\boldsymbol{w})\}$$

Constraint $(\mathcal{IC}_0)$ depends on how $\pi(0,0,\boldsymbol{w})$ compares to $\pi(0,1,\boldsymbol{w})$. When $\pi(0,0,\boldsymbol{w}) \geq \pi(0,1,\boldsymbol{w})$, $(\mathcal{IC}_0)$ is equivalent to $(\mathcal{IC}S)$. When $\pi(0,0,\boldsymbol{w}) < \pi(0,1,\boldsymbol{w})$, $(\mathcal{IC}_0)$ is equivalent to $(\mathcal{IC}J)$.

$$p_1(w_{11} - w_{01}) + (1-p_1)(w_{10} - w_{00}) \geq \hat{c}, \qquad (\mathcal{IC}_S)$$
$$(p_1 + \delta p_0)w_{11} + (1 - p_1 - \delta p_0)w_{10} + (\delta - p_1 - \delta p_0)w_{01} - (1 + \delta - p_1 - \delta p_0)w_{00} \geq \hat{c}. \qquad (\mathcal{IC}_J)$$

The solution to $(\mathcal{P}_0)$ clearly satisfies $w_{00} = 0$ as $w_{00}$ has negative coefficients in both $(\mathcal{IC}^{AS})$ and $(\mathcal{IC}_J)$, while it has a positive coefficient in the objective function. Moreover, both $w_{10}$, and $w_{01}$ have the same coefficients in the objective function, while the latter has a smaller coefficient in the constraints. Thus, the solution to $(\mathcal{P}_0)$ also satisfies $w_{00} = 0$. Finally, we know that if $\boldsymbol{w}^J$ is the solution to $(\mathcal{P}_0)$, then it is also the solution to the principal's problem. This implies that, when the solution to the principal's problem satisfies, $w_{11} > w_{10}$, it also satisfies $w_{01} = w_{00} = 0$. Hence, it remains to show that $w_{01} = w_{00} = 0$ also holds when the optimal solution satisfies $w_{11} \leq w_{10}$.

Consider problem $(\mathcal{P}_{AS})$ in the absence of assumption $w_{01} = w_{00} = 0$:

$$\min_{\boldsymbol{w}} \left(\sigma + (1-\sigma)p_1^2\right)w_{11} + (1-\sigma)p_1(1-p_1)(w_{10}+w_{01}) + (1-\sigma)(1-p_1)^2 w_{00} \qquad (\mathcal{P}')$$
$$\text{s.t. } (\mathcal{IC}_0), (\mathcal{IC}^{AS}), w_{11} \le w_{10}.$$

Now, in the absence of $w_{01} = w_{00} = 0$, the constraint $w_{11} \le w_{10}$ does not necessarily require $\pi(0,0,\boldsymbol{w}) \ge \pi(0,1,\boldsymbol{w})$. Therefore, $(\mathcal{IC}_0)$, which may be equivalent to $(\mathcal{IC}_{\mathcal{S}})$ or $(\mathcal{IC}_{\mathcal{J}})$. To figure out how things are different here in comparison to the proof of Lemma 3, which characterizes the optimal solution of $(\mathcal{P}_{AS})$, we need to have a better understanding of the constraints. We already know the expressions for $(\mathcal{IC}_{\mathcal{S}})$ or $(\mathcal{IC}_{\mathcal{J}})$. The constraints $(\mathcal{IC}_1^{AS})$, and $(\mathcal{IC}_2^{AS})$ (in the absence of $w_{01} = w_{00} = 0$) are given by:

$$w_{10} + (1+\delta)p_1(w_{11} - w_{10}) - w_{00} + (\delta - (1+\delta)p_1)(w_{01} - w_{00}) > \hat{c}, \qquad \left(\mathcal{IC}_1^{AS}\right)$$
$$w_{10} + 2p_1(w_{11} - w_{10}) - w_{00} + (1-2p_1)(w_{01} - w_{00}) \ge \hat{c}. \qquad \left(\mathcal{IC}_2^{AS}\right)$$

Moreover, in the absence of $w_{01} = w_{00} = 0$, $(\mathcal{IC}_{1'}^{AS})$ satisfies:

$$\mathcal{IC}_{1'}^{AS} = \begin{cases} w_{10} - w_{00} + \dfrac{[(1-\delta^2)p_0 - (\delta+\delta^2)p_1](w_{11}-w_{10}) + [(\delta+\delta^2)p_1 - (1-\delta^2)p_0 - \delta^2](w_{01}-w_{00})}{1-\delta-\delta^2} < \hat{c} \text{ if } \delta+\delta^2 < 1, \\ (p_1 - \delta p_0)(w_{11}-w_{10}) + (\delta^2 + \delta p_0 - p_1)(w_{01}-w_{00}) > 0 \text{ if } \delta+\delta^2 = 1, \\ w_{10} - w_{00} + \dfrac{[(\delta+\delta^2)p_1 - (1-\delta^2)p_0](w_{11}-w_{10}) + [\delta^2 + (1-\delta^2)p_0 - (\delta+\delta^2)p_1](w_{10}-w_{00})}{\delta+\delta^2-1} > \hat{c} \text{ if } \delta+\delta^2 > 1. \end{cases}$$

In the following, we show that the solution of $\mathcal{P}'_{AS}$ satisfies $w_{01} = w_{00} = 0$. Suppose that either $w_{01} \neq 0$, or $w_{00} \neq 0$ to get a contradiction. Then, constraint $(\mathcal{IC}^{AS})$ must be fulfilled through $(\mathcal{IC}_{1'}^{AS})$. Otherwise, if $(\mathcal{IC}^{AS})$ was fulfilled through $(\mathcal{IC}_1^{AS})$, or $(\mathcal{IC}_2^{AS})$, then we would have $w_{01} = w_{00} = 0$. The reasons for this are twofold. First, constraints $(\mathcal{IC}S)$, $(\mathcal{IC}J)$, $(\mathcal{IC}_1^{AS})$, and $(\mathcal{IC}_2^{AS})$ increase with $w_{00}$, yet the objective function decreases with $w_{00}$. Second, in all these constraints, $w_{10}$ has a weakly larger coefficient than that of $w_{01}$; while their coefficients are the same in the objective function. Next, we separately examine the three cases that define $(\mathcal{IC}_{1'}^{AS})$.

Case 1: $\delta + \delta^2 > 1$. Constraints $(\mathcal{IC}_{1'}^{AS})$, $(\mathcal{IC}S)$, and $(\mathcal{IC}J)$ increase with $w_{00}$, yet the objective function decreases with $w_{00}$. Moreover, in all these constraints, $w_{10}$ has a weakly larger coefficient than that of $w_{01}$; while they have the same coefficients in the objective function. Therefore, $w_{01} = w_{00} = 0$.

Case 2: $\delta + \delta^2 = 1$. The sign of $(\delta^2 + \delta p_0 - p_1)$ turns to be crucial in this case. We thus, further separate the discussion into three subcases.

• When $\delta^2 + \delta p_0 - p_1 < 0$. Constraint $(\mathcal{IC}_{1'}^{AS})$ requires $w_{00} > w_{01}$ as we already have $w_{11} \le w_{10}$. Then, the incentive scheme is an RPE, and hence $(\mathcal{IC}_0)$ is equivalent to $(\mathcal{IC}_{\mathcal{S}})$. Since increasing $w_{01}$, or $w_{00}$ hurts the objective function and the constraint $(\mathcal{IC}_{\mathcal{S}})$, the only reason for assigning a positive value for $w_{01}$ or $w_{00}$ is to fulfill $(\mathcal{IC}_{1'}^{AS})$. Then, it is clear to see that $w_{00}$ must be set (slightly above) $\frac{p_1 - \delta p_0}{p_1 - \delta^2 - \delta p_0}(w_{10} - w_{11})$, while $w_{01} = 0$. Consequently, $w_{11}$, and $w_{10}$ needs to minimize the objective function, conditional on the value of $w_{00}$, and subject to the constraint $(\mathcal{IC}_{\mathcal{S}})$:

$$\min_{w_{11} \le w_{10}} \left(\sigma + (1-\sigma)p_1^2\right)w_{11} + (1-\sigma)p_1(1-p_1)w_{10} + (1-\sigma)(1-p_1)^2 \frac{p_1 - \delta p_0}{p_1 - \delta^2 - \delta p_0}(w_{10} - w_{11})$$
$$\text{s.t. } p_1 w_{11} + (1-p_1)w_{10} - (1-p_1)\frac{p_1 - \delta p_0}{p_1 - \delta^2 - \delta p_0}(w_{10} - w_{11}) \ge \hat{c}$$

This is a linear problem with constraint $w_{11} \leq w_{10}$. Thus, we have either $w_{11} = w_{10}$, or $w_{11} = 0$. The latter would violate the constraint since it features a negative coefficient for $w_{10}$. If former is correct, then $w_{11} = w_{10} = \hat{c}$, which gives rise to a contradiction as it requires $w_{01} = w_{00} = 0$.

• When $\delta^2 + \delta p_0 - p_1 > 0$. Similar arguments provided in the previous subcase lead to a contradiction here as well.

• When $\delta^2 + \delta p_0 - p_1 = 0$. We have a contradiction since $(\mathcal{IC}_{1'}^{AS})$ is $w_{11} - w_{10} > 0$ in this case.

Case 3: $\delta + \delta^2 < 1$. Due to the linearity of the problem, and given the constraint $w_{11} \leq w_{10}$, we have either $w_{11} = w_{10}$, or $w_{11} = 0$. If the former, the best we can do is to set $w_{11} = w_{10} = \hat{c}$, and $w_{01} = w_{00} = 0$, which is a contradiction. If the latter, then $w_{10}$ must be sufficiently large so that satisfying $(\mathcal{IC}_{1'}^{AS})$ requires to assign a positive value for $w_{01}$, or $w_{00}$ in this case. Specifically, we need $w_{10} = \frac{1-\delta-\delta^2}{1-\delta-\delta^2-(1-\delta^2)p_0+(\delta+\delta^2)p_1}\hat{c} + A$, for some positive $A$. Given that $(\mathcal{IC}_{1'}^{AS})$ is linear, and that satisfying this constraint is the only reason to set a positive value for $w_{01}$ and $w_{00}$, only one of them really needs to be positive. Here, we consider the case with $w_{00} > 0$; the other case similarly follows. To satisfy $(\mathcal{IC}_{1'}^{AS})$, $w_{00}$ needs to be large enough:[9]

$$w_{00} > A\frac{1-\delta-\delta^2}{1-\delta-2\delta^2-(1-\delta^2)p_0+(\delta+\delta^2)p_1}. \tag{A2}$$

As the incentive scheme is RPE in this case, $(\mathcal{IC}_0)$ is equivalent to $(\mathcal{IC}_{\mathcal{S}})$, which requires:

$$w_{00} \leq A - \frac{\frac{p_1-(1-\delta^2)p_0}{1-p_1}}{1-\delta-\delta^2-(1-\delta^2)p_0+(\delta+\delta^2)p_1}\hat{c}. \tag{A3}$$

Equations (A2) and (A3) leads to

$$A\frac{(\delta+\delta^2)p_1-(1-\delta^2)p_0-\delta^2}{1-\delta-2\delta^2-(1-\delta^2)p_0+(\delta+\delta^2)p_1} > \hat{c}\,\frac{\frac{p_1-(1-\delta^2)p_0}{1-p_1}}{1-\delta-\delta^2-(1-\delta^2)p_0+(\delta+\delta^2)p_1}.$$

This defines a lower bound for $A$, which in turn implies $w_{10} > \frac{1-\delta^2+(\delta+\delta^2)p_1-(1-\delta^2)p_0}{(1-p_1)((\delta+\delta^2)p_1-(1-\delta^2)p_0-\delta^2)}$.[10] But, then there is no need to set $w_{00} > 0$, since $w_{10}$ is sufficiently large to satisfy $(\mathcal{IC}_1^{AS})$ is satisfied at this value when $w_{00} = 0$. Therefore, decreasing $w_{00}$ to 0 improves the objective function without violating $(\mathcal{IC}_{\mathcal{S}})$. This makes $(\mathcal{IC}_{1'}^{AS})$ redundant as $(\mathcal{IC}_1^{AS})$ is already satisfied, and gives a contradiction.

In sum, solution to problem $\mathcal{P}'$ coincides with that of $(\mathcal{P}_{AS})$. Therefore, $w_{01} = w_{00} = 0$ also holds, when the optimal solution satisfies $w_{11} \leq w_{10}$. □

## APPENDIX B: PROOFS OF THE STATEMENTS FROM SECTION 4

*Proof of Propositions 3 and* 4. The expected costs of the JPE contract $\boldsymbol{w}^J$, the IPE contract $\boldsymbol{w}^I$, and the RPE contract $\boldsymbol{w}^R$—which we denote by $C^J$, $C^I$, and $C^R$, respectively—satisfy:

$$C^J = 2\frac{1}{1-\delta}\left[\sigma + (1-\sigma)p_1^2\right]\frac{\hat{c}}{p_1+\delta p_0}$$

$$C^I = 2\frac{1}{1-\delta}[\sigma + (1-\sigma)p_1]\hat{c}$$

$$C^R = 2\frac{1}{1-\delta}(1-\sigma)p_1(1-p_1)\frac{\hat{c}}{1-(1+\delta)p_1}$$

After automation adoption, the costs of these contracts become $C_A^J$, $C_A^I$, and $C_A^R$, respectively. These expressions that govern these costs are similar to above except that now $\sigma$ is replaced with $\sigma_A$ and $c$ is replaced with $c_A$ in each.

Denoting $\beta = \frac{1-\sigma_A}{1-\sigma} < 1$ (hence $\sigma_A = 1 - \beta + \beta\sigma$), and $\gamma = \frac{c_A}{c} < 1$, we can see that $\frac{\hat{c}_A}{\hat{c}} = \frac{\frac{\gamma c}{\beta(1-\sigma)(p_1-p_0)}}{\frac{c}{(1-\sigma)(p_1-p_0)}} = \frac{\gamma}{\beta}$ since $\hat{c} = \frac{c}{(1-\sigma)(p_1-p_0)}$. Then we have:

$$r_J = \frac{C_A^J}{C^J} = \frac{1-\beta(1-\sigma)\left(1-p_1^2\right)}{1-(1-\sigma)\left(1-p_1^2\right)}\frac{\gamma}{\beta},$$
$$r_I = \frac{C_A^I}{C^I} = \frac{1-\beta(1-\sigma)(1-p_1)}{1-(1-\sigma)(1-p_1)}\frac{\gamma}{\beta},$$
$$r_R = \frac{C_A^R}{C^R} = \gamma < 1$$

It is clear that $r_J > r_I > r_R$. Therefore, a principal who uses an RPE contract before automation, never switches to another type of contract even if she adopts automation. Moreover, as $r_R < 1$, optimal RPE contract becomes cheaper after automation, hence the principal always adopts automation when she was using an RPE contract before automation. This does not hold for the other cases (when she was using IPE or JPE contracts) since $r_I$ and $r_J$ can be greater than 1.

Overall, a principal using an RPE contract before, adopts automation and continues to use an RPE contract. Moreover, a principal using a JPE or an IPE contract may adopt automation and switch to an RPE contract. Thus, use of RPE contracts expands with automation. □

*Proof of Proposition* 5. For given $w$ and $e_m$, the agent has the following problem:

$$\max_e pwee_m - \frac{c}{2}e^2.$$

The solution to this problem gives us his best response function: $\mathbf{e}(w, e_m) = min\{1, \kappa we_m\}$. Then we can write the principal's problem as follows:

$$\max_{w,e_m} (1-w)p\mathbf{e}(w, e_m)e_m - \frac{\alpha c}{2}e_m^2.$$

The net value of the success is now equal to $1 - w$, since the principal pays $w$ to agent in case of a success. This term is multiplied with the probability of success, $p\mathbf{e}(w, e_m)e_m$. Finally, since the principal assumes the cost of operating machinery, we subtract it from her revenue.

We first analyze the optimal plan to induce the maximal effort, $e = 1$, from the agent. The wage must be set $w = \frac{1}{\kappa e_m}$, and hence the principal's problem boils down to:

$$\max_{e_m} \left(1 - \frac{1}{\kappa e_m}\right)pe_m - \frac{\alpha c}{2}e_m^2.$$

The derivative w.r.t $e_m$ is $p - \alpha c e_m$ which is always positive, hence $e_m = 1$, and $w = \frac{1}{\kappa}$.

If principal induces the agent to choose $e < 1$, from the agent's best response function we know that $e = \kappa we_m$. Hence the wage must be set $w = \frac{e}{\kappa e_m}$ and the principal's problem becomes:

$$\max_{w,e_m} (1-w)wp\kappa e_m^2 - \frac{\alpha c}{2}e_m^2.$$

The derivative (w.r.t. $e_m$) is $2(1-w)wp\kappa e_m - \alpha c e_m$, which is linear in $e_m$. Therefore it is optimal to set $e_m$ at its maximum capacity, as long as production is on. Then by plugging in $e_m = 1$, and taking the first-order condition with respect to $w$, one can see that it is optimal to set $w = \frac{1}{2}$. This, in turn, induces the agent to set his effort level to $e = \frac{\kappa}{2}$. Putting the findings together, there are three possibilities for the principal: (i) shutting down the production, (ii) inducing the agent to choose $e = \frac{\kappa}{2} < 1$ with $e_m = 1$, and $w = \frac{1}{2}$, (iii) inducing the agent to exert maximal effort $e = 1$ with $e_m = 1$, and $w = \frac{1}{\kappa}$. The resulting profit for these options are are 0, $\frac{p\kappa}{4} - \frac{\alpha c}{2}$, and $p - c - \frac{\alpha c}{2}$, respectively. Comparing these values, we obtain the result. □

**Lemma 4.** *Without automation, for a given wage w, a collusion-free equilibrium outcome has to be in one of the following forms:*

(i) *employees choose* $(e_1, e_2) = (1, 1)$ *in all periods.*
(ii) *employees choose* $(e_1, e_2) = (0, 0)$ *in all periods.*

*Proof of Lemma* 4. First, note that choosing (0, 0) in every period is an SPE. Therefore, it is natural to expect this to be a collusion-free equilibrium when $w$ is not sufficiently large.

Suppose that, for some $w$, there is a collusion-free equilibrium inducing positive effort at some periods. We want to show that, in this collusion-free equilibrium, the agents choose their maximal effort levels in every period. First, there is at least a period in which the agents receive a positive payoff. This, however, requires that the effort pair $(e_1, e_2) = (1, 1)$ maximizes the agents' total payoffs in the stage game due to increasing marginal returns. Therefore, showing that the repetition of this effort pair in every period can be sustained as an SPE would be sufficient to complete our proof. This stems from the fact that there cannot be any other SPE bringing a higher total payoff to agents. To get a contradiction, suppose that choosing $(e_1, e_2) = (1, 1)$ in every period cannot be sustained even with the most severe punishment: exerting minimal effort in all the remaining periods. That is, deviating to the best response in a period is profitable deviation for agent $i$:

$$wp - \frac{c}{2} < (1-\delta)\left[wp\mathbf{e}(w,1) - \frac{c}{2}(\mathbf{e}(w,1))^2\right].$$

It must be $\mathbf{e}(\boldsymbol{w}, 1) = min\{1, w\kappa\} = w\kappa$, because otherwise the agents are already best responding to each other and do not want to deviate. Therefore, we must have

$$wp < \frac{c}{2} + (1-\delta)\frac{w^2p^2}{2c}.$$

However, we already know that the collusion-free equilibrium that we have started with induces an effort profile $(e_1, e_2) > > (0, 0)$ at least for some periods. Therefore this profile must also be sustained as an SPE with the most severe punishment. In other words, the following incentive constraints for agent 1 and 2 must be satisfied respectively.[11]

$$wpe_1e_2 - \frac{c}{2}e_1^2 \geq (1-\delta)\left[\frac{w^2p^2}{2c}e_2^2\right]$$
$$wpe_1e_2 - \frac{c}{2}e_2^2 \geq (1-\delta)\left[\frac{w^2p^2}{2c}e_1^2\right]$$

By reorganizing these two inequalities one can get get:

$$wp \geq \frac{c}{2}\left(\frac{e_1^2 + e_2^2}{2e_1e_2}\right) + (1-\delta)\frac{w^2p^2}{2c}\left(\frac{e_1^2 + e_2^2}{2e_1e_2}\right)$$

But this contradicts with $wp < \frac{c}{2} + (1-\delta)\frac{w^2p^2}{2c}$, because $e_1, e_2 \leq 1$, and hence $\frac{e_1^2+e_2^2}{2e_1e_2} \geq 1$. □

*Proof of Proposition* 6. The optimal plan either induces the maximal effort in every period, or shuts down the production. The incentive constraint will be relevant only if it induces effort. Denoting $x = wp$, one can rewrite the incentive constraint as follows:

$$-(1-\delta)x^2 + 2cx - c^2 \geq 0$$

The principal will choose the minimum possible value of $w$ satisfying the above condition. $x = \frac{c}{1+\sqrt{\delta}}$ is the minimum value of $x$ at which the above inequality is satisfied. Therefore it is optimal to set $w = \frac{c}{p(1+\sqrt{\delta})}$. Thus, the principal has a positive profit only if $\kappa \geq \frac{2}{1+\sqrt{\delta}}$, where $\kappa = \frac{p}{c}$. Otherwise it is optimal to shut down the production. □