Misaligned agents

Pierre Chaigneau* Queen's University

Nicolas Sahuguet† HEC Montréal

January 1, 2026

Abstract

How should firms select and incentivize managers who allocate resources based on private information when their performance is imperfectly measured? We show that, contrary to common wisdom, firms may benefit from appointing managers with misaligned preferences. When performance can be measured more accurately along one dimension than another, firms optimally select managers who are biased toward the poorly measured dimension. This allows to provide strong performance-based incentives on the well-measured dimension while relying on managerial bias to counterbalance substitution effects. As a result, firms with worse performance measurement on some dimension will tend to hire more biased managers, such as managers with strong prosocial or environmental motivations. This framework can explain apparent inconsistencies between firms' stated objectives and the explicit incentives they provide, and it offers a novel perspective on the strategic roles of corporate culture and corporate social responsibility.

Keywords: biased agents, corporate social responsibility, high-powered incentives, executive compensation, interests alignment, personnel economics.

*Address: Smith School of Business, Goodes Hall, 143 Union Street West, K7L 2P3, Kingston, ON, Canada. Email: pierre.chaigneau@queensu.ca. Tel: 613 533 2312.

†Address: Applied Economics Department, HEC Montréal, 3000 chemin de la côte Sainte Catherine, H3T 2A7 Montréal, QC, Canada. Email: nicolas.sahuguet@hec.ca. Tel: 514 340 6031.

1 Introduction

Managers play a special role in organizations. Whereas most employees follow instructions and execute their assigned tasks, managers assign tasks and make decisions. Many of these decisions are resource allocation decisions. They make these decisions because they are better informed or better able to make sense of relevant information than their employers. This informational advantage emanates from expertise and circumstantial data. For example, a CEO can evaluate the merits of an investment project for her firm once she goes over a valuation analysis. Some agents also make decisions with resource allocation implications based on their expertise and information. For example, a medical doctor can diagnose an illness and prescribe medical interventions when she sees a patient. For short, we refer to all such agents as managers.

Economic agents have heterogeneous preferences, including social preferences (Wrzesniewski et al. (1997), Cassar and Meier (2018), Fehr and Charness (2025)),[1] so that they may have different preferences about resource allocation decisions. In particular, their preferences could differ from their employer's. For example, a potential CEO may have a preference for environmental projects whereas the firm's board is mostly concerned with financial returns. A medical doctor may be concerned about her patients' welfare whereas the hospital mostly cares about the cost of treatment.

The performance of a manager can usually be measured, although these measures are often imperfect. Moreover, managers with a deep understanding of the production process can largely anticipate how their resource allocation decisions will affect their measured performance, as modeled in Baker (1992). For example, a well-informed CEO can understand the profitability implications of cutting some greenhouse gases emissions, as well as the effect of these cuts on measures of the firm's environmental performance.

In this context, we study which type of manager an employer would rather appoint, i.e. which types of preferences regarding resource allocation he would prefer the manager to have. Our main result is that the employer will sometimes prefer that the manager *not* be perfectly aligned, all else equal. We show that the employer will prefer a manager who has a stronger preference than him for resource allocation on some dimension if performance is poorly measured on this dimension, or if it is very well-measured on the other dimension. For example, if a product manager can allocate resources to improve either the quality or the cost of a product, the employer ("principal") will appoint a manager who cares more than he does about

[1] Many empirical studies find that personal values affect decisions. For example, the personal values of social enterprises' CEOs affect their attention to social goals (Stevens et al. (2014)); corporate researchers prefer science-oriented research and are willing to take a pay cut to work on fundamental research (Stern (2004)); medical doctors are reluctant to consider the costs of treatment they prescribe because this is inconsistent with their "fundamental motivations" (Hall (1996)).

quality. We obtain this result in a simple model with one principal and one manager, so that it is not driven by market equilibrium effects. Instead, it is driven by the interaction between individual preferences and performance measurement.

The logic can be explained as follows. Because of her expertise and her experience on the job, the manager understands the production process and knows the characteristics of potential projects. As in Baker (1992), this privately informed manager can predict how various resource allocations will affect both actual output and measured performance on each dimension. Without explicit incentives, she would just follow her preferences. Providing explicit incentives can improve her decision by making her responsive to her information, but it will also result in gaming of the incentive scheme when output is imperfectly measured: she will allocate resources on the dimension where it easily improves measured performance even if it does not improve actual output correspondingly (Kerr (1975) and Baker (1992)). As a result, incentives will be more muted on the dimension with worse performance measurement, similar to Baker (1992). For example, if cost is more easily measured than quality, the manager's incentives will be higher-powered for cost performance than for quality performance.

These asymmetric incentives make the manager biased toward improving the better-measured dimension of performance (e.g. reducing cost) at the expense of neglecting the other dimension (e.g. improving quality), similar to the logic at play in multitasking models. Increasing the sensitivity of pay to performance on the other dimension is not a solution: this would make the manager more concerned with her poorly measured performance on that dimension, which would result in substantial gaming and resource misallocation.

However, this bias can be mitigated by appointing a manager who prefers to allocate resources to the poorly-measured dimension (e.g. a manager who intrinsically values quality). Specifically, when the quality of performance measurement differs across dimensions, we show that the principal will optimally appoint a manager who cares more than he does about the poorly measured dimension (e.g. product quality), and provide explicit incentives on the other dimension (e.g. product cost). This type of manager is "misaligned".

For example, a product manager with these preferences will have a natural tendency to improve quality. She will only reduce cost to the extent that this improves cost performance *and* that she has strong explicit incentives to do so. Providing her with strong explicit incentives for cost reduction allows to make her highly concerned about her performance on the dimension where it is well-measured. In turn, this makes her more responsive to her information on this dimension. At the same time, her intrinsic preference for quality prevents excessive cost reductions at the expense of quality. Relative to a hypothetical manager whose preferences were perfectly aligned with the principal's, a misaligned manager will have even lowered-powered incentives on quality performance, and higher-powered incentives on cost

performance.

In summary, we show that the principal will often rather appoint a misaligned manager. The explicit incentives of a misaligned manager will be muted on the poorly-measured dimension, and high-powered on the well-measured dimension. This distinguishes our model from three branches of the literature.

First, standard models of multitasking point out that it is detrimental to give explicit incentives on one dimension only, and they accordingly predict low-powered incentives for all tasks when tasks are substitutes.

Second, the principal-agent literature has shown that hiring aligned agents will generate efficiency gains. With imperfect or nonexistent performance measurement, hiring employees who share the same preferences as the principal allows to more efficiently align interests, for several reasons. First, a manager who intrinsically values the principal's objective is less prone to game incentive schemes. Second, an agent who internalizes the objective of the principal or follows norms that promote high effort does not need high-powered incentives (Akerlof and Kranton (2005), Fischer and Huddart (2008)). Third, when the principal has nonpecuniary preferences such as a "mission", hiring an agent who is motivated by this mission allows to align interests at a lower cost (Besley and Ghatak (2005, 2006, 2018), Akerlof and Kranton (2005)). For example, in the literature on prosocial motivation, prosocial organizations try to hire individuals with a greater prosocial motivation (Delfgaauw and Dur (2007, 2008)). On the contrary, we show that organizations will sometimes rather appoint misaligned agents.

Third, in models of multitasking with imperfect performance measurement, Prendergast (2007, 2008) shows that the principal can hire "biased agents" whose bias is suited to a particular task. The solution is to use job design by allocating different types of agents to different tasks within an organization. In our model, the agent only has one task – making a resource allocation decision – so that job design is not a solution.

Our main result relies on several important assumptions: the agent must have some private information about the production technology that is relevant for her action choice, and there should be performance measures which are informative but not fully revealing. Indeed, without asymmetric information, the organization does not use pay-for-performance, and it hires an aligned agent, who shares the same preferences as the principal. Likewise, when performance measurement is either completely unreliable or perfect, the principal hires an aligned agent.

Next, we study the equilibrium matching between managers and firms. We extend the model by considering an economy with multiple firms and multiple managers. We let firms differ in the quality of their performance measurement on some dimension, and we let managers have different preferred resource allocations. We show that a stable matching will be such that managers who have a stronger preference for allocating resources to some dimension will

be matched with firms in which performance is more poorly measured on this dimension. Intuitively, the firms with the worst performance measurement on this dimension have a comparative advantage in employing these more biased managers. The bias of these managers is relatively more valuable in those firms, so that they are willing to pay more to hire them.

An interpretation of the model is that the principal can create an organizational culture, including norms, which affect managers' preferred resource allocations. This can be for two reasons: either managers assimilate and internalize the organizational culture, or the organization attracts and hires agents who are a good fit.[2] In the former case, the organization trains its hires to influence their preferred resource allocation; in the latter case, the organization selects employees with the right preferences. In any case, the corporate culture is part of the overall incentive system.

The model predicts that the corporate culture will depend on the performance measurement technology. This implies that changes in performance measurement will result in changes to corporate cultures. Most notably, better measurement on some dimension will shift the corporate culture away from this dimension. The model also predicts that organizations will create cultures and norms that differ from what the principal (who can be viewed as the founders, owners, or directors) would prefer. What could at first glance be viewed as organizational capture by biased agents might actually be optimal from the principal's perspective.

In particular, our model suggests a new theory of Corporate Social Responsibility (CSR) and Diversity, Equity and Inclusion (DEI) commitments. Both are controversial and puzzling, to a large extent because many organizations seemingly went too far in this direction at the expense of other goals including profitability. While diversity and corporate social responsibility can be measured, these measures are notoriously bad or at least incomplete (Berg, Kölbel, and Rigobon (2022)), and they are often easy to game (Edmans (2021)).

Our model suggests that owners who would like to encourage *some degree* of diversity or social responsibility may create an organizational culture that is *heavily* biased in this direction when good measures of the organization's other goals are available. For corporations, the latter primarily includes measures of firm value. Firm value will typically be well-measured in publicly listed firms, but also in many private firms whose implicit valuation is regularly established in funding rounds or equity sales. These firms will offer high-powered financial incentives, which partially offset their strong culture. In summary, when financial performance is better measured than CSR, firms that would like to encourage some degree of CSR will

[2]The assimilation argument suggests that organizations will prefer to hire young managers, whose preferences may be more malleable (Gregg, Seibt, and Banaji (2006), Malmendier and Nagel (2011)), all the more that organizations have different cultures. It also suggests that the employees of organizations that impart cultures with desirable attributes (e.g. ethical behavior, focus on quality, etc.) will be in high-demand, so that jobs at these organizations will be in high-demand among young employees with career concerns. We leave a formal investigation of these arguments to future research.

appoint managers with a CSR bias, and give them incentives which are only high-powered on the financial dimension.

The model can help make sense of apparent inconsistencies between firms' stated objectives and the explicit incentives they provide. For example, Walker (2022) notes that firms' commitments to ESG (Environmental, Social, and Governance) in the early 2020s contrast with very low-powered ESG-based incentives for their top executives, which he describes as "window dressing". As another example, Vladasel et al. (2024) find that many social enterprises simultaneously offer a social mission and monetary rewards for (more easily measured) commercials tasks. The model also predicts that even a slight shift in investors' preferences toward causes such as CSR will result in a significantly larger change in corporate culture. What could be viewed as overshooting can actually be consistent with interest alignment with imperfect performance measurement. For example, this can help explain the rise in left-leaning corporate speech between 2012 and 2022, which is only partially explained by nonpecuniary investor preferences (Cassidy and Kempf (2025)).[3]

A majority of CEOs think that "top talent prefers to work for organizations with social values which are aligned to their own."[4] Our model shows how organizations should establish their values or culture to attract the most valuable type of manager given asymmetries in their performance measurement system.

2 Related literature

There is a large literature on nonpecuniary objectives, especially prosocial motivation (Francois and Vlassopoulos (2008), Besley and Ghatak (2018), Cassar and Meier (2018)). These papers point out that non-pecuniary benefits can be derived from actions or from outcomes. For example, the agent derives private benefits from some actions ("warm glow") in Besley and Ghatak (2005), and the cost of effort is affected by the agent's type in Delfgaauw and Dur (2007). The agent's nonpecuniary benefits can substitute for monetary compensation (Besley and Ghatak (2005, 2018), Dixit (2005)), and for incentive provision (Besley and Ghatak (2005), Akerlof and Kranton (2005), Ellingsen and Johannesson (2008), Makris (2009)). Cassar (2019) finds experimental evidence consistent with these predictions.

A closely related branch of the literature considers intrinsic motivation or pro-social motivation as a potential solution to the alignment problem between principal and agent when

[3]Ferreira and Nikolowa (2025) provide an alternative explanation which does not involve incentives or interest alignment. In their model, firms benefit from hiring workers with nonpecuniary preferences (e.g. a preference for working for an environmentally responsible firm). They show that, to extract more surplus from workers, firms would rather hire workers with more extreme preferences.

[4]Source: 2016 survey by PricewaterhouseCoopers, cited in Cassar and Meier (2018).

performance is poorly measured. This can rationalize the use of low-powered incentives in public services (Francois (2000)). When workers are heterogeneous, a key result in this literature is that employees with a purpose are hired by organizations with a mission that matches this purpose (Besley and Ghatak (2005, 2006), Francois (2007), Kosfeld and Von Siemens (2011), Henderson and Van den Steen (2015)).[5] Cassar and Armouti-Hansen (2020) study the consequences of misalignment between the principal and agent, when they disagree about the purpose or "mission" of the project. In models of identity and incentive provision, what matters is "how employees think of themselves in relation to the firm", which again suggests that employees' identities should match their organization's goals (Akerlof and Kranton (2008), Kampkötter, Petters, and Sliwka (2011)). On the contrary, Prendergast (2007, 2008) argues that, when contracts are incomplete, firms will hire agents who are intrinsically motivated for some tasks ("biased agents") to work on a narrow set of tasks. The idea is to use job design to allocate different agents to different tasks.[6]

It is already well-understood that poor performance measurement will reduce the effectiveness and efficiency of performance-based incentives (Baker (1992)), which makes it all the more useful to rely on intrinsic motivation instead (Besley and Ghatak (2018)). Likewise, multitasking in environments with poor performance measurement, and the reliance on intrinsic motivation in this context, has already been studied (Dixit (2002), Bénabou and Tirole (2016), Besley and Ghatak (2018)). Our contribution is to study the consequences of imperfect performance measurement along several dimensions for the principal's preferences about the agent's type when tasks cannot be unbundled.[7]

Bénabou and Tirole (2016) study a multitasking model in which one task is measurable and contractible, the other is not, and the agent can have intrinsic motivation for the former. In their model, increased competition for workers with heterogeneous ability leads to a rise in the use of incentive pay for screening purposes. This results in a shift of effort toward the more easily quantifiable task. Dewatripont, Jewitt, and Tirole (1999) consider a model of career concerns with multitasking. They show that additional tasks make it harder to learn the agent's ability, which reduces effort. This suggests a beneficial effect from hiring specialists with narrow jobs.

[5]For example, in the baseline model of Besley and Ghatak (2018), agents have different degrees of pro-social motivation, i.e., differentiation is vertical only. By contrast, in Besley and Ghatak (2005), there are two types of intrinsic motivations and two types of missions (horizontal differentiation), but also agents with pecuniary preferences only (vertical differentiation). Henderson and Van den Steen (2015) define an organization's purpose or mission as a concrete "objective for the firm that reaches beyond profit maximization."

[6]Prendergast (2007) analyzes how bureaucrats with different biases are assigned to different jobs. For example, in tax or immigration enforcement, the interests of the "principal" and "clients" are not aligned, so the principal will hire a bureaucrat biased against clients.

[7]Another perspective is to study the effect of explicit incentives on the agent's preferences (Kreps (1997, 2023)).

Another branch of the literature studies subjective performance evaluation when output is observable but not contractible (Baker, Gibbons, and Murphy (1994), MacLeod (2003)). Public economics has studied the implications of immeasurability and uncontractability for incentive provision. Hart, Shleifer, and Vishny (1997) summarize the debate about public services provision by private suppliers or by government employees by pointing out that private suppliers usually deliver a lower quality service at a lower cost. This is because quality cannot be fully specified in a contract – it is noncontractible. In their model, the service provider can exert effort on two dimensions to improve the quality of the service and reduce its cost, and cost reductions lead to lower quality. They conclude that public provision of public services is preferable when quality matters more than cost. They do not consider preferences over agents' types, or matching of agents to various tasks.

Finally, the results speak to a large empirical literature that studies complementarities between human resource management practices, including incentive pay, training, and screening (Ichniowski and Shaw (2003), Lazear and Gibbs (2014)).

3 Model

We consider a principal-agent model of resource allocation. A principal hires an agent who receives some information about the productivity of potential allocations of the organization's resources. The principal lacks the information necessary to simply tell the agent how to allocate resources. Likewise, the principal cannot use job design to allocate one agent to one task.[8] Indeed, the agent has only one task, making a resource allocation decision. She does so on the basis of her incentives and her private information.[9]

The principal and the agent have preferences about the resource allocation decision. Specifically, the agent incurs a private cost from deviating from a baseline resource allocation. This implies that the agent has a preferred resource allocation that she would put in place in the absence of explicit incentives. The organization also incurs a cost from deviations from a potentially different baseline resource allocation. If the principal and agent's baseline resource allocations are the same, we say that they have congruent or aligned preferences.

The organization produces unverifiable output V_i on two dimensions $i = 1, 2$. On each dimension, its performance is measured by the metric P_i. Output and performance are affected

[8]Holmström (2017) argues that, when faced with a multitasking problem, principals can often use job design, including the distribution of tasks across agents. He argues that this allows to incentivize the easy-to-measure tasks with high-powered incentives, while the hard-to-measure tasks can be allocated to other agents.

[9]This modeling assumption needs not be taken literally. For example, the agent may not possess any special information, but her qualifications and her on-the-job experience may allow her to draw the implications of some publicly available information to make better resource allocation decisions.

by a resource allocation decision, a. It can be interpreted as the investment of time, money, and other organizational resources on each dimension. As in Baker (1992), output and performance are affected by shocks which are privately observed by the agent at the time of making the resource allocation decision. Let $\boldsymbol{\epsilon} \equiv \{\epsilon_1^v, \epsilon_2^v, \epsilon_1^p, \epsilon_2^p\}$.

$$V_1(a, \boldsymbol{\epsilon}) = a\epsilon_1^v \tag{1}$$
$$V_2(a, \boldsymbol{\epsilon}) = (1-a)\epsilon_2^v \tag{2}$$
$$P_1(a, \boldsymbol{\epsilon}) = a\epsilon_1^p \tag{3}$$
$$P_2(a, \boldsymbol{\epsilon}) = (1-a)\epsilon_2^p \tag{4}$$

For $i \in \{1, 2\}$ and $j \in \{v, p\}$, the random variable ϵ_i^j has a standard deviation σ_i^j, and a mean normalized such that: $\mathbb{E}\left[\epsilon_i^j\right] = 1$.[10] Define:

$$\rho_i \equiv \frac{cov(\epsilon_i^v, \epsilon_i^p)}{\sigma_i^v \sigma_i^p} \tag{5}$$

We assume that performances and outcomes are independent across dimensions, so that:

$$cov(\epsilon_i^p, \epsilon_j^v) = 0 \text{ and } cov(\epsilon_i^p, \epsilon_j^p) = 0 \qquad \forall i \neq j. \tag{6}$$

The objective function of the principal is:

$$\mathbb{E}\left[V_1(a, \boldsymbol{\epsilon}) + V_2(a, \boldsymbol{\epsilon}) - W(a, \boldsymbol{\epsilon}) - \frac{k}{2}(a - a^\star)^2\right] \tag{7}$$

where $W(a, \boldsymbol{\epsilon})$ is the manager's payoff, and $k > 0$. The last term is the cost to the organization of deviations from a baseline resource allocation $a^\star$.

As in Baker (1992), the principal provides a linear incentive contract such that the agent's payoff is equal to:

$$W(a, \boldsymbol{\epsilon}) = S + b_1 P_1(a, \boldsymbol{\epsilon}) + b_2 P_2(a, \boldsymbol{\epsilon}) \tag{8}$$

Linear contracts are a standard assumption in this literature (Baker (1992), Besley and Ghatak (2005), Fischer and Huddart (2008), Bénabou and Tirole (2016)).

Given a contract $\{S, b_1, b_2\}$, the agent's information $\boldsymbol{\epsilon}$, and resource allocation a, the ex-

[10] Baker (1992) makes a similar assumption by equating the expectation of the performance P to the expectation of the outcome V.

pected utility of the agent is:

$$\mathbb{E}\left[S + b_1 P_1(a, \boldsymbol{\epsilon}) + b_2 P_2(a, \boldsymbol{\epsilon}) - \frac{c}{2}(a - \overline{a})^2\right] \tag{9}$$

where $c > 0$ parameterizes the cost to the agent of deviating from her preferred resource allocation $\overline{a}$. The agent has reservation utility H, and chooses a to maximize her expected utility.

3.1 Discussion of modeling assumptions

We consider a model of resource allocation, in which the agent must choose how to allocate a finite resource, such as time or money, to different uses. This is different from a multitasking model, in which the agent must choose his effort on different tasks. In both models, the decisions to allocate more on various dimensions may be substitutes. There are also important differences between these two models. First, in our model of resource allocation, there is only one action, and preferences over possible actions are simply described by one easily interpretable parameter. It is then straightforward to compare the preferences of the principal and agent. Second, in a multitasking model, individual tasks could in principle be allocated to different agents. Indeed, the division of tasks has been recommended as solutions to multitasking problems (Prendergast (2007, 2008), Holmström (2017)). By contrast, the resource allocation model involves a single task. Third, multitasking models assume that effort is privately costly for the agent, so that inefficiencies take the form of low effort, and effort could be elicited via audits (Sinclair-Desgagné (1999)).[11] This framework may be practically less relevant for some occupations, including top managers and startup employees, who are usually motivated to work very hard even without explicit monetary incentives (Edmans, Gosling, and Jenter (2023)).[12]

There are several interpretations for the cost of deviating from a baseline resource allocation $a^\star$ in the principal's objective function in equation (7). First, it could be a reduced-form representation of the cost of various resource allocations given the firm's production technology. Second, $a^\star$ can represent an approved resource allocation for planning purposes, and any deviation will involve additional costs. Third, under a literal interpretation, $a^\star$ represents the principal's preferences over the resource allocation decision. They may depend on the personal preferences of the organization's directors (e.g. Kim, Minton, and Williamson (2025)) or on the organization's ownership. Owners with non-financial objectives, including organizations with

[11] Relatedly, we show below that, if the performance measurement technology allows to perfectly monitor the resource allocation, then the principal optimally appoints an aligned agent.

[12] According to Martin Mignot from Index Ventures, the 996 work culture is the new startup standard. The 996 model involves working 12 hours a day for six days per week. Source: PitchBook newsletter, June 22 2025.

"missions" such as foundations, may be biased towards specific actions or resource allocations (Besley and Ghatak (2005)). Alternatively, some owners may be concerned about synergies between the organization and other ones in their investment portfolio.

We assume that the agent has preferences over the resource allocation decision (similar to Besley and Ghatak (2005, 2018)). The agent's objective function in equation (9) is in the spirit of Fischer and Huddart (2008) and Akerlof and Kranton (2005). In Fischer and Huddart (2008), agents find it costly to deviate from the norm, and the equivalent of $\overline{a}$ in their model is a "norm parameter". In Akerlof and Kranton (2005), the agent incurs a disutility if and only if her effort deviates from her "ideal effort", similarly to the last term of the agent's utility in equation (9). In practice, there is evidence that many economic agents have "professional motivations", i.e. preferences over actions (e.g. Hall (1996), Stevens et al. (2014), Chalfin and Gonçalves (2023)). These nonpecuniary preferences are heterogeneous in the population: not all workers care about nonpecuniary factors to the same extent, and different workers care about different aspects of their jobs (Cassar and Meier (2018), Wrzesniewski et al. (1997)).

4 Analysis

In this section, we study which type of agent the principal would rather hire.

We define the first-best as the outcome when the principal can offer a forcing contract after observing the state $\boldsymbol{\epsilon}$ and the agent's resource allocation decision a.

Lemma 1 *At the first-best, the principal hires an aligned agent:* $\overline{a} = a^{\star}$.

Intuitively, the principal will choose the resource allocation that he prefers. The only problem is therefore to minimize the cost of the agent's compensation. It is cheaper to hire an agent who shares the same preferences. Indeed, the agent will need to be compensated for the deviation between the chosen resource allocation allocation and her preferred resource allocation.

We now study the equilibrium outcome when the agent's resource allocation decision a is unobservable, i.e. the second-best outcome. We solve the model backwards, starting with the agent's resource allocation decision, then deriving the optimal contract, and finally the agent's bias $\overline{a}$ that the principal prefers. We start by deriving several preliminary results.

Lemma 2 *For a given contract* $\{S, b_1, b_2\}$*, the agent optimally chooses the following resource allocation:*

$$a = \overline{a} + \frac{1}{c}\left(b_1\epsilon_1^p - b_2\epsilon_2^p\right) \tag{10}$$

The optimally chosen resource allocation is additive in two terms. The first is the agent's baseline resource allocation $\overline{a}$. This is what she would choose in the absence of explicit incentives. The second term captures the deviation from this baseline level due to her information and incentives. The higher the sensitivity of pay to performance on dimension 1 (as measured by b_1) and the more allocating resources to this dimension improves performance (depending on ϵ_1^p), the more the agent allocates resources to the first dimension by increasing a. Likewise, the higher the sensitivity of pay to performance on dimension 2 (as measured by b_2) and the more allocating resources to this dimension improves performance (depending on ϵ_2^p), the more the agent allocates resources to the second dimension by *decreasing* a.

We start by focusing on special cases that allow to better understand the logic of the model. The proofs are special cases of the proofs of Lemma 3 and Proposition 1.

Case 1 *If there is no asymmetric information between the principal and the agent about the optimal resource allocation (i.e.* $\sigma_1^v = \sigma_2^v = 0$*), then the principal hires an aligned agent* $(\overline{a} = a^\star)$ *and* $b_1 = b_2 = 0$.

When $\sigma_1^v = \sigma_2^v = 0$, the agent's private information is not relevant for outcomes. Then, the only role of performance-based incentives is to elicit the principal's preferred allocation, $a = a^\star$, in a framework with symmetric information. This can simply be achieved by hiring an agent with the same preferences, i.e. $\overline{a} = a^\star$, and not providing any explicit incentives, i.e. $b_1 = b_2 = 0$.

Case 2 *If performance measures are independent of output (i.e.* $\rho_1 = \rho_2 = 0$*), then the principal hires an aligned agent* $(\overline{a} = a^\star)$*, and* $b_1 = b_2 = 0$.

When performance is independent from output on both dimensions, "performance measures" are pure noise. This is equivalent to a case without performance measurement. Then, there will not be any pay-for-performance: $b_1 = b_2 = 0$. To align interests as much as possible, the principal will optimally hire an aligned agent by choosing $\overline{a} = a^\star$.

Case 3 *If performance measures and output are the same (i.e.* $\sigma_1^p = \sigma_1^v$*,* $\sigma_2^p = \sigma_2^v$*, and* $\rho_1 = \rho_2 = 1$*) then the principal hires an aligned agent* $(\overline{a} = a^\star)$*, and* $b_1 = b_2 = \frac{1}{1+k/c}$.

When performance measures perfectly reveal output, the principal can achieve the first-best outcome at the second-best. The agent is not entirely residual claimant because $k > 0$. Intuitively, the principal suffers a disutility whenever the resource allocation deviates from the baseline allocation $a^\star$. The principal does not just want to maximize the sum of outputs, he

also wants to minimize deviations from the baseline resource allocation. The latter is achieved by hiring an aligned agent with lower-powered incentives.[13]

Case 4 *If performance on some dimension reveals the resource allocation but is unrelated to output (i.e. $\sigma_1^p = 0$), then the principal hires an aligned agent ($\overline{a} = a^\star$).*

Some performance measurement technologies are perfectly informative about the agent's choice of a. For example, some monitoring technologies allow to know the time spent by an office worker in front of her desk, as opposed to collaborative meetings (see Raith (2008), where the contract can also be contingent on inputs). For $\sigma_1^p = 0$, performance on the first dimension perfectly reveals the choice of a, and it is unrelated to the shock ϵ_1^v that drives output on this dimension. Therefore, the principal can incentivize her preferred choice of a without preventing the agent from using her private information. The principal offers state-contingent incentives to the agent on the first dimension to partly offset incentives on the second dimension (it is optimal to offer incentives on the second dimension so that the agent is responsive to the signal ϵ_2^p that she gets on ϵ_2^v). In this case, the principal will prefer to hire an agent with congruent preferences.

We now analyze the general case.

Lemma 3 *The optimal sensitivity of pay to performance on both dimensions is:*

$$b_1 = \frac{c\left(\sigma_2^{p2}\left(-k\left(\bar{a}-a^\star\right)\right)+\rho_2\sigma_2^p\sigma_2^v+\rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2}+1\right)\right)}{(c+k)\left(\sigma_1^{p2}\sigma_2^{p2}+\sigma_1^{p2}+\sigma_2^{p2}\right)} \tag{11}$$

$$b_2 = \frac{c\left(\sigma_1^{p2}k\left(\bar{a}-a^\star\right)+\rho_1\sigma_1^p\sigma_1^v+\rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2}+1\right)\right)}{(c+k)\left(\sigma_1^{p2}\sigma_2^{p2}+\sigma_1^{p2}+\sigma_2^{p2}\right)} \tag{12}$$

To better understand this result, it is helpful to consider these intermediate expressions for b_1 and b_2:

$$b_1 = \frac{1}{\sigma_1^{p2}+1}\left(\frac{\rho_1\sigma_1^v\sigma_1^p-k(\overline{a}-a^\star)}{1+\frac{k}{c}}+b_2\right) \tag{13}$$

$$b_2 = \frac{1}{\sigma_2^{p2}+1}\left(\frac{\rho_2\sigma_2^v\sigma_2^p+k(\overline{a}-a^\star)}{1+\frac{k}{c}}+b_1\right) \tag{14}$$

This shows that pay-performance sensitivity (PPS) on dimension i is additive in three terms.

[13]Interestingly, this is no longer true if performance measures are perfectly informative ($\rho_1 = \rho_2 = 1$) but do not have the same volatility as output. Intuitively, in this case, optimal pay-for-performance varies across performance measures, so that it is generally suboptimal to hire a perfectly aligned agent.

First, it is increasing in $\rho_i \sigma_i^v \sigma_i^p$, similar to equation (8) in Baker (1992). This captures the correlation ρ_i between performance and output, but also uncertainty σ_i^v about the production technology, and performance variability σ_i^p. Intuitively, pay will be very sensitive to performance on a dimension if performance is well-measured on this dimension: it is worth giving the agent strong incentives to be responsive to her private information on this dimension. On the contrary, pay-for-performance can backfire when performance is poorly measured: the manager will then improve measured performance without correspondingly improving output.[14] In the special case with deterministic performance on dimension i ($\sigma_i^p = 0$), performance only reflects resource allocation a, and it does not account for its productivity ϵ_i^v on this dimension.

Second, PPS depends on the difference between the resource allocation preferred by the agent, $\overline{a}$, and the one preferred by the principal, $a^\star$. Intuitively, the principal uses PPS to better align the agent's interests with respect to the resource allocation decision. For example, when the agent is biased toward dimension 2 relative to the principal, i.e. $\overline{a} < a^\star$, PPS on dimension 1 increases, and PPS on dimension 2 decreases.

Third, PPS on a dimension is increasing in PPS on the other dimension. This is a standard effect when tasks are substitutes so that the balance of incentives matters. Suppose that PPS increases on some dimension because of one of the two effects described above. To not excessively discourage resource allocation on the other dimension, PPS on this other dimension increases as well. Thus, improvements in performance measurement on one dimension increase PPS on both dimensions – a spillover effect.

The fixed wage S is set to keep the agent at her reservation utility given the sensitivity of pay to performance derived in Lemma 3.

Lemma 4 *The principal's objective function can be rewritten as:*

$$\begin{aligned} & 1 + \frac{1}{c}\left(b_1\left(\mathbb{E}\left[\epsilon_1^p \epsilon_1^v\right] - 1\right) + b_2\left(\mathbb{E}\left[\epsilon_2^p \epsilon_2^v\right] - 1\right)\right) \\ & -\frac{1}{2c}\left(b_1^2 \mathbb{E}\left[\epsilon_1^{p2}\right] + b_2^2 \mathbb{E}\left[\epsilon_2^{p2}\right] - 2 b_1 b_2\right) - H - \frac{k}{2}\mathbb{E}\left[\left(\overline{a} - a^\star + \frac{b_1 \epsilon_1^p - b_2 \epsilon_2^p}{c}\right)^2\right] \end{aligned} \tag{15}$$

where b_1 and b_2 are described in Lemma 3.

We now consider which preferred resource allocation $\overline{a}$ of the agent maximizes the principal's objective function. We refer to this value of $\overline{a}$ as the one that the principal "prefers". To better understand some of the effects at play, we start by deriving the resource allocation that

[14] This is one of the main limitations of pay-for-performance identified by personnel economics. As pointed out by Lazear and Shaw (2007), loan officers compensated based on the number of loans may accept excessively risky loans, while manufacturers paid on piece rates might sell more but at the cost of neglecting maintenance and inventories (Freeman and Kleiner (2005)).

the principal prefers in the hypothetical case when PPS is exogenously given, i.e. holding b_1 and b_2 constant.

Lemma 5 *Holding b_1 and b_2 constant, the principal prefers to appoint an agent with:*

$$\overline{a} = a^\star + \frac{b_2 - b_1}{c} \tag{16}$$

When b_1 and b_2 are exogenously given, the principal hires an aligned agent if and only if PPS is the same on both dimensions. Intuitively, if PPS is higher on one dimension, then the agent will be biased toward allocating more resources to this dimension. To offset this bias, it is optimal to hire an agent whose preference for allocating resources to the other dimension exceeds the principal's. Of course, the principal also controls the PPS b_1 and b_2, and he will choose $\overline{a}$ taking into account its effect on the PPS.

We can now state our main result, which relates the principal's preference over $\overline{a}$ to exogenous parameters.

Proposition 1 *The principal prefers to appoint an agent with:*

$$\overline{a} = a^\star + \frac{\rho_2 \sigma_2^p \sigma_2^v \sigma_1^{p2} - \rho_1 \sigma_1^p \sigma_1^v \sigma_2^{p2}}{c\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right) + k\sigma_1^{p2}\sigma_2^{p2}} \tag{17}$$

This implies that the principal will rather appoint a misaligned agent with $\overline{a} < a^\star$ if ρ_1 is sufficiently high or if ρ_2 is sufficiently low. The principal will rather appoint an agent with a stronger preference for resource allocation on a dimension if performance is sufficiently poorly measured on that dimension, or if it is sufficiently well-measured on the other dimension.

Example 5 *(ESG-based compensation) Let the first dimension be profitability, and the second dimension be ESG. First, suppose that profitability is very well-measured because the firm is publicly listed and investors are well-informed: ρ_1 is high. Second, suppose that profitability can vary a lot across potential investments: σ_1^v is high. In this case, when both ρ_1 and σ_1^v are high, it is especially important for the manager to be responsive to the productivity signal ϵ_1^p. This is achieved with a high sensitivity of pay to performance on the first dimension, i.e. a high b_1 (see equation (11)). Third, suppose that ESG performance is poorly measured, because metrics are highly imperfect and partial: ρ_2 is low. In this context, the firm will optimally hire a manager with a bias for ESG investments, i.e. $\overline{a} < a^\star$. This further reduces the sensitivity of pay to ESG performance, b_2 (see equation (12)).*

Normalizing performance measures' volatilities, which involves some loss of generality, allows to derive a more specific result.

Corollary 1 *Given $\sigma_1^p = \sigma_2^p = 1$, the principal appoints an agent with $\overline{a} < a^\star$ if and only if $\rho_2\sigma_2^v < \rho_1\sigma_1^v$.*

When $\rho_1 > \rho_2$, performance is better measured on the first dimension. In this case, the principal would rather have $\bar{a} < a^\star$, i.e. appoint an agent whose bias is to allocate resources to the second dimension. The optimal bias will be greater when output volatility on both dimensions is high, i.e., when the agent's private information matters more. Intuitively, in this case, explicit incentives are higher-powered to make the agent more responsive to her private information. To partly offset the effect of strong incentives on the first performance dimension on resource allocation, the principal appoints an agent who is more biased in favor of the second dimension.

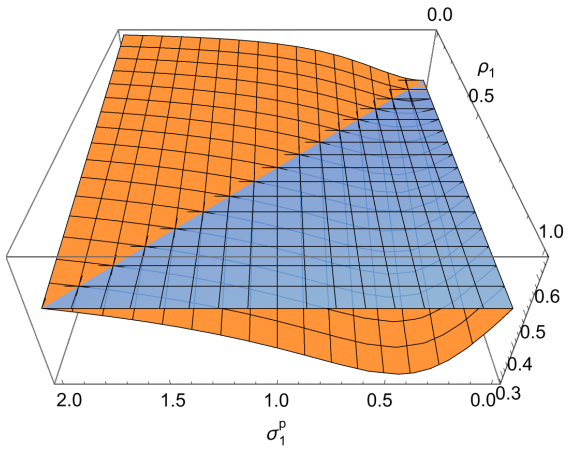

Figure 1: Orange surface: optimal $\overline{a}$ as a function of σ_1^p and ρ_1, for $a^\star = 0.5$, $\sigma_1^v = \sigma_2^v = \sigma_2^p = 1$, $\rho_2 = 0.5$, $c = k = 1$, and $\boldsymbol{\epsilon}$ follows a multivariate normal distribution. Translucent blue surface: $a^\star = 0.5$.

In summary, the relation between output and performance on both dimensions determines the appointment policy and the sensitivity of pay to performance on both dimensions. When performance is a good measure of output on dimension i, the agent should be highly responsive to her signal, ϵ_i^p. To avoid excessive resource allocation on the well-measured dimension at the expense of the poorly measured dimension because of the difference in the sensitivity of pay to performance on these two dimensions, the principal optimally appoints an agent who is biased in favor of the latter.

5 Matching

In this section, we extend the model to analyze the endogenous matching between principals and agents.[15] We consider an economy with a number N of principals and agents. Principals differ on ρ_1 (the quality of performance measurement on the first dimension), while agents differ on $\overline{a}$ and c (their preferred resource allocation, and the cost of deviating from their preferred resource allocation). The set of agents is denoted by $M \equiv \{m_1, m_2, \ldots, m_N\}$, and the set of firms is denoted by $F \equiv \{f_1, f_2, \ldots, f_N\}$. Agent m_i is characterized by $\{c_i, \overline{a}_i\}$; firm f_i is characterized by ρ_1^i.

An equilibrium of the model is such that: (i) all principal-agent matches are stable, i.e., no pair that includes an agent and a principal currently not matched to each other can increase their payoffs by leaving their respective match and forming a new match; (ii) all principals offer to the agent the optimal contract that maximizes the joint surplus; (iii) all agents choose the optimal resource allocation given their contracts. $\overline{a}_i$. To keep the model tractable, we henceforth assume $\sigma_1^v = 1, \sigma_2^v = 1, \sigma_1^p = 1, \sigma_2^p = 1$.

First, we consider the matching between principals with heterogeneous ρ_1^i and agents with heterogeneous $\overline{a}_i$.

Proposition 2 *Other things equal, agents with a high $\overline{a}_i$ are matched with principals with a low ρ_1^i.*

Holding $a^\star$ and ρ_2 constant, we get assortative matching between firms with lower ρ_1^i and agents with higher $\overline{a}_i$. Principals with poor performance measurement on dimension 1 are matched with agents who are biased towards allocating resources to this dimension.

Second, we consider the matching between principals with heterogeneous ρ_1^i and agents with heterogeneous c_i.

Proposition 3 *Other things equal, agents with a low c_i are matched with principals with a high ρ_1^i.*

When ρ_1^i is higher, the principal has access to a better performance measurement system. He can therefore provide better incentives, and values more highly an agent who is more responsive to incentives. An agent with a low c_i does not incur a large cost from deviations from her preferred resource allocation. She is therefore more responsive to incentives, and more valuable for this type of principal than for a principal with a lower ρ_1^i. In short, a principal with a better performance measurement technology is matched with an agent with weak preferences over the resource allocation.

[15] See Bandiera, Guiso, Prat and Sadun (2015) and Macho-Stadler and Pérez-Castrillo (2021) for a review of this literature.

6 Applications

We now discuss several applications of our stylized model. It is worth noting that the rise in incentive pay is relevant in many occupations as documented by Lazear and Shaw (2007) and Lemieux, MacLeod, and Parent (2009). It takes many forms, including bonuses, commissions, profit-sharing, equity-based pay, and piece-rate contracts.

6.1 ESG and equity-based executive compensation

In the private sector, firm value is a contractible performance measure that is very closely aligned with the objective of shareholders (Baker (1992)). Yet, some relevant business activities may not be reflected in firm value.[16] Firms which provide social initiatives and missions that are valuable to their workers tend to improve the satisfaction, motivation, attraction, and retention of their employees (Kitzmueller and Shimshack (2012), Henderson and Van den Steen (2015), Cassar and Meier (2018)), which improves future financial performance (Edmans (2011), Edmans et al. (2024)). However, many of these activities are usually poorly measured, and potential measures might be easy to game.

Our model suggests that this could be addressed by hiring managers who prefer to allocate resources on these dimensions, and then incentivize these managers on the financial dimension with explicit incentives based on firm value. These incentives will be all the more high-powered that stock market valuation provides a better measure of firm value, for example due to improvements in stock market efficiency. They will also be higher-powered when boards have a stronger preference for financial returns, e.g. when the shareholder value paradigm is prevalent.

6.2 Short-term and long-term performance

The resource allocation problem can involve investments that improve performance either in the short or in the long-term, such as investing in sales or in R&D. The principal (firm owner) might have preferences in that regard, for example because the firm might benefit from R&D tax credits or because he also owns downstream firms that would also benefit from technological advances. Likewise, the agent could have a preference for achieving milestones or for easily observable accomplishments that are more likely to improve her reputation.

[16] Even before the advent of ESG, the 2018 Deloitte Global Human Capital Trends report summarized: "Organizations are no longer assessed based only on traditional metrics such as financial performance, or even the quality of their products or services. Rather, organizations today are increasingly judged on the basis of their relationships with their workers, their customers, and their communities, as well as their impact on society at large." (as cited in Cassar and Meier (2018)).

Suppose that short-term output is realized and therefore perfectly measurable in the short-term. By contrast, since long-term output will only be realized later, only imperfect measures such as firm valuation can capture it in the short-term. These measures are especially imperfect when there is a lot of uncertainty about firm value, and no liquid market with well-informed investors. That is, performance measurement of long-term outcomes will be especially poor in young privately held firms – for example start-ups backed by venture capitalists.

The model predicts that these firms' governance systems will have two important features. First, they will provide high-powered incentives based on short-term performance – such as performance milestones in VC-backed firms, or short-term performance targets in CEO compensation contracts. Second, the founder or manager of a high-potential start-up will be "visionary", who is intrinsically focused on harder to measure long-term outcomes and not overly concerned with short-term outcomes.

6.3 Product management

For many tasks, quality is not easily or immediately observable, but the cost or quantity produced is easily and immediately observable. This is an important consideration for product managers. This is also relevant in other occupations, such as craftsmen who build custom furniture or chefs who compose menus. When companies hire these managers and workers, they are often looking for candidates with a passion for quality. It is not enough for them to be qualified or skilled, they must also have the right "values". Indeed, an employee who is "passionate" about dimensions that are hard to measure, such as the quality or durability of a product or service, can be effectively incentivized with explicit incentives on the dimensions that are easy to measure such as cost or quantity.

6.4 Universities

A research university employs faculty members who are at the same time researchers and instructors. These faculty members face a resource allocation problem in how to allocate their time between teaching and research.

Some institutions that care more about the quantity of research produced will find it easy to incentivize on this dimension, for example with bonus pay that depends on publications. If in addition they mostly teach undergraduate students who may not recognize or value high-quality teaching, measures of teaching quality such as student evaluations will be uninformative and easy to game. In this case, explicitly incentivizing teaching could backfire. Accordingly, these schools often create a "teaching culture".[17]

[17]This can be achieved indirectly, including in hiring and promotion decisions, but also by celebrating

Other institutions care about research quality and societal impact, which are hard to measure especially in the short-term. If in addition they mostly teach mature students such as MBAs, MSc, and PhDs, who tend to be better able to recognize and appreciate valuable teaching, then teaching quality can usually be better measured. These schools or universities will create a research culture[18] and provide high-powered teaching incentives. The latter can be achieved by generous overload payments for extra teaching, including executive education.

In summary, universities will tend to create organizational cultures around the dimension that is easier to game, and they will provide explicit incentives on the dimension where performance is a good measure of the output they value.[19]

6.5 Healthcare

Medical doctors can allocate their time to actions, such as performing surgeries or prescribing drugs, that tend to improve a measurable short term outcome. They can alternatively allocate their time to better understand individual cases as well as provide authoritative guidance and advice to patients and their relatives. This tends to improve immeasurable long-term outcomes. Of course, in some cases surgery will be the best course of action for the patient, whereas in other cases it will not be, and the medical doctor who interacts with the patient is best placed to make this assessment.

The model suggests that hospitals should hire doctors with an intrinsic preference for better understanding individual cases as well as providing guidance and advice to patients, and incentivize them based on short-term outcomes.[20] By contrast, in recent years, various forms of performance pay in hospitals has arguably led to the overuse of revenue-generating tests and procedures, and to under-investments in low-margin work such as preventive care and cooperation between specialists (Gawande (2009)). Our model suggests that the blame may lie not in the introduction of financial incentives per se, but instead in the failure to preserve the culture of putting the patient first. According to Lester Dyke, a cardiac surgeon

excellent teaching rather than research. It can also be achieved more directly. For example, the Dean can ask a research-productive faculty member to focus on teaching rather than research.

[18] At one institution, a senior faculty member addressing a group of PhD students who were about to teach classes reminded them to not get carried away with teaching and not waste more time than necessary on this mundane activity at the expense of research. The department's newsletter focused almost exclusively on research, and so did departmental events including seminars, lecture series, and social events. The status of a faculty member would depend on his or her status as a researcher; no one was known or admired as a "star teacher".

[19] Alternatively, if all measures are easy to game, then incentives will be low-powered on all dimensions, as also predicted by other models.

[20] A concern is that this incentive scheme could lead to adverse selection, since it may be especially attractive to doctors who intrinsically prefer the types of actions that improve short-term outcomes. This suggests that screening by hospitals and (credible) signaling by medical students or residents may play a useful role. For example, medical students may donate time or money to their local churches or charities.

quoted by Gawande (2009), "We took a wrong turn when doctors stopped being doctors and became businessmen."

7 Conclusion

This paper provides a new theory of managerial selection and incentive provision. We consider a manager with private information in charge of making a resource allocation decision. When performance is imperfectly and asymmetrically measured, the principal will rather appoint a misaligned agent, i.e., an agent with different preferences about the resource allocation decision. We show that the principal will appoint an agent who is biased toward allocating resources to the dimension of performance which is more poorly measured. Thus, an improvement in the measurement technology will change the type of preferences valued by the principal.

The model can explain the concomitance of high-powered explicit incentives on a dimension of performance which is well-measured, and low-powered incentives on a poorly measured dimension of performance. The appointment of a misaligned manager accentuates the difference in pay-for-performance across dimensions.

Our results highlight a new reason for the importance of jointly determining compensation policy and hiring policy. For example, a company that optimally hires biased agents and compensates them with incentive pay should not suddenly stop using incentive pay, even if it would be warranted by changes in the economic environment. Indeed, agents would then only rely on their biases to make decisions.

Our results echo the notion that an important part of education is to endow agents destined for careers in some field with preferences appropriate for this field. They imply that there is a role for inculcating some values during the education process, for screening on the basis of values (Van den Steen (2010)), for creating contexts or a corporate culture that elicit specific values (Akerlof and Kranton (2005)), and for signaling these values (Bénabou and Tirole (2006)). Our contribution to this literature is to point out that the most relevant values depend on the asymmetric nature of performance measurement.

8 References

Akerlof, G.A. and Kranton, R.E., 2005. Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19(1), 9-32.

Akerlof, G.A. and Kranton, R.E., 2000. Economics and identity. *Quarterly Journal of Economics*, 115(3), 715-753.

Akerlof, G.A. and Kranton, R.E., 2008. Identity, supervision, and work groups. *American Economic Review*, 98(2), 212-217.

Ashraf, N. and Bandiera, O., 2017. Altruistic capital. *American Economic Review*, 107(5), 70-75.

Baker, G.P., 1992. Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3), 598-614.

Baker, G., 2002. Distortion and risk in optimal incentive contracts. *Journal of Human Resources*, 37(4), 728-751.

Baker, G., Gibbons, R. and Murphy, K.J., 1994. Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics*, 109(4), 1125-1156.

Bandiera O., L. Guiso, A. Prat, R Sadun. 2015. Matching Firms, Managers and Incentives. *Journal of Labor Economics*, 33(3), 623-681.

Bartling, B., Fehr, E. and Schmidt, K.M., 2012. Screening, competition, and job design: Economic origins of good jobs. *American Economic Review*, 102(2), 834-864.

Bénabou, R. and Tirole, J., 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70(3), 489-520.

Bénabou, R. and Tirole, J., 2006. Incentives and prosocial behavior. *American Economic Review*, 96(5), 1652-1678.

Bénabou, R. and Tirole, J., 2016. Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, 124, 305-370.

Berg, F., Kölbel, J.F. and Rigobon, R., 2022. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315-1344.

Besley, T. and Ghatak, M., 2005. Competition and incentives with motivated agents. *American Economic Review*, 95(3), 616-636.

Besley, T. and Ghatak, M., 2006. Sorting with motivated agents: implications for school competition and teacher incentives. *Journal of the European Economic Association*, 4(2-3), 404-414.

Besley, T. and Ghatak, M., 2018. Prosocial motivation and incentives. *Annual Review of Economics*, 10(1), 411-438.

Bewley, T.F., 1995. A depressed labor market as explained by participants. *American Economic Review*, 85(2), 250-254.

Cassar, L. and Meier, S., 2018. Nonmonetary incentives and the implications of work as a source of meaning. *Journal of Economic Perspectives*, 32(3), 215-238.

Cassar, L., 2019. Job mission as a substitute for monetary incentives: Benefits and limits. *Management Science*, 65(2), 896-912.

Cassar, L. and Armouti-Hansen, J., 2020. Optimal contracting with endogenous project mission. *Journal of the European Economic Association*, 18(5), 2647-2676.

Cassidy, W.M. and Kempf, E., 2025. Partisan corporate speech. Working paper, National Bureau of Economic Research.

Chalfin, A. and Gonçalves, F.M., 2023. Professional motivations in the public sector: Evidence from police officers. Working paper, National Bureau of Economic Research.

Delfgaauw, J. and Dur, R., 2007. Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization*, 62(4), 605-624.

Delfgaauw, J. and Dur, R., 2008. Incentives and workers' motivation in the public sector. *Economic Journal*, 118(525), 171-191.

Dewatripont, M., I. Jewitt and J. Tirole, 1999. The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies. *Review of Economic Studies*, 66(1), 199-217.

Dixit, A., 2002. Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, 37(4), 696-727.

Dixit, A., 2005. Incentive contracts for faith-based organizations to deliver social services. *Economic theory in a changing world: Policy modelling for growth.* Oxford University Press.

Ederer, F., Holden, R. and Meyer, M., 2018. Gaming and strategic opacity in incentive provision. *RAND Journal of Economics*, 49(4), 819-854.

Edmans, A., 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3), 621-640.

Edmans, Alex, 2021. Why companies shouldn't tie CEO pay to ESG metrics. *Wall Street Journal*, June 27.

Edmans, A., Pu, D., Zhang, C. and Li, L., 2024. Employee satisfaction, labor market flexibility, and stock returns around the world. *Management Science*, 70(7), 4357-4380.

Edmans, A., Gosling, T. and Jenter, D., 2023. CEO compensation: Evidence from the field. *Journal of Financial Economics*, 150(3), 103718.

Ellingsen, T. and Johannesson, M., 2008. Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3), 990-1008.

Fehr, E. and Charness, G., 2023. Social preferences: fundamental characteristics and economic consequences. *Journal of Economic Literature*, 63(2), 440-514.

Ferreira, D. and Nikolowa, R., 2025. Polarization, purpose and profit. *Journal of Financial Economics*, 172, 104147.

Fischer, P. and Huddart, S., 2008. Optimal contracting with endogenous social norms. *American Economic Review*, 98(4), 1459-1475.

Francois, P., 2000. 'Public service motivation' as an argument for government provision. *Journal of Public Economics*, 78(3), 275-299.

Francois, P., 2007. Making a difference. *RAND Journal of Economics*, 38(3), 714-732.

Francois, P. and Vlassopoulos, M., 2008. Pro-social motivation and the delivery of social services. *CESifo Economic Studies*, 54(1), 22-54.

Freeman, R.B. and Kleiner, M.M., 2005. The last American shoe manufacturers: Decreasing productivity and increasing profits in the shift from piece rates to continuous flow production. *Industrial Relations*, 44(2), 307-330.

Frey, B.S., 1994. How intrinsic motivation is crowded out and in. *Rationality and Society*, 6(3), 334-352.

Frey, B.S., 1997. *Not Just for the Money: An Economic Theory of Personal Motivation.* Cheltenham: Edward Elgar.

Frey, B.S. and Oberholzer-Gee, F., 1997. The cost of price incentives: An empirical analysis of motivation crowding-out. *American Economic Review*, 87(4), 746-755.

Frey, B.S. and Jegen, R., 2001. Motivation crowding theory. *Journal of Economic Surveys*, 15(5), 589-611.

Frydman, C. and Saks, R.E., 2010. Executive compensation: A new view from a long-term perspective, 1936-2005. *Review of Financial Studies*, 23(5), 2099-2138.

Gawande, A., 2009. The cost conundrum. *Annals of Medicine: The Cost Conundrum: Reporting & Essays: The New Yorker*, June 1.

Georgellis, Y., Iossa, E. and Tabvuma, V., 2011. Crowding out intrinsic motivation in the public sector. *Journal of Public Administration Research and Theory*, 21(3), 473-493.

Gregg, A.P., Seibt, B. and Banaji, M.R., 2006. Easier done than undone: asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, 90(1), 1-20.

Hall, M.A, 1996. Motivating Physicians With Financial Incentives, in *Making Medical Spending Decisions: The Law, Ethics, and Economics of Rationing Mechanisms.* Oxford University Press.

Hart, O., Shleifer, A. and Vishny, R.W., 1997. The proper scope of government: theory and an application to prisons. *Quarterly Journal of Economics*, 112(4), 1127-1161.

Henderson, R. and Steen, E.V.D., 2015. Why do firms have "purpose"? The firm's role as a carrier of identity and reputation. *American Economic Review*, 105(5), 326-330.

Holmström, B., 1979. Moral hazard and observability. *Bell Journal of Economics*, 10(1), 74-91.

Holmström, B., 2017. Pay for performance and beyond. *American Economic Review*, 107(7), 1753-1777.

Holmström, B. and Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7, 24-52.

Ichniowski, C. and Shaw, K., 2003. Beyond incentive pay: Insiders' estimates of the value of complementary human resource management practices. *Journal of Economic Perspectives*, 17(1), 155-180.

James Jr, H.S., 2005. Why did you do that? An economic examination of the effect of extrinsic compensation on intrinsic motivation and performance. *Journal of Economic Psychology*, 26(4), 549-566.

Jones, D.B., Tonin, M., Vlassopoulos, M. and Winichakul, K.P., 2023. Paying for what kind of performance? Performance pay, multitasking, and sorting in mission-oriented jobs. *Games and Economic Behavior*, 142, 480-507.

Kampkötter, P., Petters, L.M. and Sliwka, D., 2021. Employee identification and wages–on the economics of "Affective Commitment". *Journal of Economic Behavior & Organization*, 188, 608-626.

Kerr, S., 1975. On the folly of rewarding A, while hoping for B. *Academy of Management Journal*, 18(4), 769-783.

Kim, S., Minton, B.A. and Williamson, R.G., 2025. Climate Boards: Do Natural Disaster Experiences Make Directors More Prosocial? Working paper, National Bureau of Economic Research.

Kitzmueller, M. and Shimshack, J., 2012. Economic perspectives on corporate social responsibility. *Journal of Economic Literature*, 50(1), pp.51-84.

Kosfeld, M. and Von Siemens, F.A., 2011. Competition, cooperation, and corporate culture. *RAND Journal of Economics*, 42(1), 23-43.

Kőszegi, B., 2014. Behavioral contract theory. *Journal of Economic Literature*, 52(4), 1075-1118.

Kreps, D.M., 1997. Intrinsic motivation and extrinsic incentives. *American Economic Review*, 87(2), 359-364.

Kreps, D., 2023. *Arguing about tastes: Modeling how context and experience change economic preferences*. Columbia University Press.

Laffont, J.J. and Martimort, D., 2009. *The Theory of Incentives: The Principal-Agent Model*. Princeton university press.

Lazear, E.P. and Gibbs, M., 2014. Personnel economics in practice. John Wiley & Sons.

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

Legros P. and A. Newman 2007. Beauty is a beast, frog is a prince: Assortative matching with non-transferabilities. *Econometrica*, 75(4), 1073-1102.

Lemieux, T., MacLeod, W.B. and Parent, D., 2009. Performance pay and wage inequality. *Quarterly Journal of Economics*, 124(1), 1-49.

Macho-Stadler, I., and Pérez-Castrillo, D. (2021). Agency theory meets matching theory. *SERIEs*, 12(1), 1-33.

MacLeod, W.B., 2003. Optimal contracting with subjective evaluation. *American Economic Review*, 93(1), 216-240.

Malmendier, U. and Nagel, S., 2011. Depression babies: do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics*, 126(1), 373-416.

Makris, M., 2009. Incentives for motivated agents under an administrative constraint. *Journal of Economic Behavior & Organization*, 71(2), 428-440.

Prendergast, C., 2007. The motivation and bias of bureaucrats. *American Economic Review*, 97(1), 180-196.

Prendergast, C., 2008. Intrinsic motivation and incentives. *American Economic Review*, 98(2), 201-205.

Raith, M., 2008. Specific knowledge and performance measurement. *RAND Journal of Economics*, 39(4), 1059-1079.

Sinclair-Desgagné, B., 1999. How to Restore Higher-Powered Incentives in Multitask Agencies. *Journal of Law, Economics, and Organization*, 15(2), 418-433.

Sliwka, D., 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, 97(3), 999-1012.

Stern, S., 2004. Do scientists pay to be scientists? *Management Science*, 50(6), 835-853.

Stevens, R., Moray, N., Bruneel, J. and Clarysse, B., 2015. Attention allocation to multiple goals: The case of for-profit social enterprises. *Strategic Management Journal*, 36(7), 1006-1016.

Terviö, Marko. 2008. The Difference That CEOs Make: An Assignment Model Approach. *American Economic Review*, 98(3), 642-668.

Van den Steen, E., 2010. On the origin of shared beliefs (and corporate culture). *RAND Journal of Economics*, 41(4), 617-648.

Vladasel, T., Parker, S.C., Sloof, R. and Van Praag, M., 2024. Revenue drift, incentives, and effort allocation in social enterprises. *Journal of Economics & Management Strategy*, 33(3), 630-651.

Walker, D.I., 2022. The Economic (In)Significance of Executive Pay ESG Incentives. *Stanford Journal of Law, Business & Finance*, 27, 317-350.

Weisbrod, B., 1988. *The Nonprofit Economy*. Harvard University Press.

Wrzesniewski, A., McCauley, C., Rozin, P. and Schwartz, B., 1997. Jobs, careers, and callings: People's relations to their work. *Journal of Research in Personality*, 31(1), pp.21-33.

9 Appendix

Proof of Lemma 2:

The agent chooses a to maximize her objective function in equation (9). At the time of choosing her actions, the agent observes ϵ_1^p and ϵ_2^p, so that her objective function can be rewritten as:

$$S + b_1 a \epsilon_1^p + b_2(1-a)\epsilon_2^p - \frac{c}{2}(a-\overline{a})^2 \tag{18}$$

This objective function is concave in a, so that the optimum is given by the first-order condition:

$$b_1\epsilon_1^p - b_2\epsilon_2^p - c(a-\overline{a}) = 0 \tag{19}$$

Rearranging gives the expression in Lemma 2. ■

Proof of Lemma 3:

Plugging equation (8) into equation (7), the objective function of the principal is:

$$\max_{b,S} \mathbb{E}\left[V_1(a,\epsilon) + V_2(a,\epsilon) - S - b_1 P_1(a,\epsilon) - b_2 P_2(a,\epsilon) - \frac{k}{2}(a-a^\star)^2\right] \tag{20}$$

The participation constraint and incentive constraint are respectively:

$$S + b_1 P_1(a,\epsilon) + b_2 P_2(a,\epsilon) - \frac{c}{2}(a-\overline{a})^2 \geq H \tag{21}$$

$$b_1\epsilon_1^p - b_2\epsilon_2^p - c(a-\overline{a}) = 0 \quad \Leftrightarrow \quad a = \overline{a} + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c} \tag{22}$$

Substituting in equation (20) for the binding participation constraint in equation (21), the principal's optimization program rewrites as:

$$\max_{b} \mathbb{E}\left[V_1(a,\epsilon) + V_2(a,\epsilon) - \frac{c}{2}(a-\overline{a})^2 - H - \frac{k}{2}(a-a^\star)^2\right] \tag{23}$$

This objective function can be rewritten as:

$$\mathbb{E}\left[\left(\overline{a} + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\epsilon_1^v + \left(1 - \overline{a} - \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\epsilon_2^v - \frac{c}{2}\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)^2 - H - \frac{k}{2}\left(\overline{a} - a^\star + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)^2\right]$$

The first-order condition with respect to b_1 is:

$$\begin{aligned}
& \mathbb{E}\left[\frac{\epsilon_1^p}{c}\epsilon_1^v - \frac{\epsilon_1^p}{c}\epsilon_2^v - c\epsilon_1^p\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c^2}\right) - k\frac{\epsilon_1^p}{c}\left(\overline{a} - a^\star + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\right] = 0 \\
\Leftrightarrow \quad & \mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right] - \mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right] - b_1\mathbb{E}\left[\epsilon_1^{p2}\right] + b_2\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right] - k\left((\overline{a} - a^\star)\mathbb{E}\left[\epsilon_1^p\right] + \frac{b_1}{c}\mathbb{E}\left[\epsilon_1^{p2}\right] - \frac{b_2}{c}\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right]\right) = 0 \\
\Leftrightarrow \quad & b_1 = \frac{\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right] - \mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right] - k(\overline{a} - a^\star)\mathbb{E}\left[\epsilon_1^p\right] + b_2\left(1 + \frac{k}{c}\right)\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right]}{\left(1 + \frac{k}{c}\right)\mathbb{E}\left[\epsilon_1^{p2}\right]} \\
\Leftrightarrow \quad & b_1 = \frac{\rho_1\sigma_1^v\sigma_1^p - k(\overline{a} - a^\star)}{\left(1 + \frac{k}{c}\right)\left(\sigma_1^{p2} + 1\right)} + b_2\frac{1}{\sigma_1^{p2} + 1} \qquad (24)
\end{aligned}$$

Likewise, the first-order condition with respect to b_2 is:

$$\begin{aligned}
& \mathbb{E}\left[-\frac{\epsilon_2^p}{c}\epsilon_1^v + \frac{\epsilon_2^p}{c}\epsilon_2^v + c\epsilon_2^p\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c^2}\right) + k\frac{\epsilon_2^p}{c}\left(\overline{a} - a^\star + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\right] = 0 \\
\Leftrightarrow \quad & b_2 = \frac{\rho_2\sigma_2^v\sigma_2^p + k(\overline{a} - a^\star)}{\left(1 + \frac{k}{c}\right)\left(\sigma_2^{p2} + 1\right)} + b_1\frac{1}{\sigma_2^{p2} + 1} \qquad (25)
\end{aligned}$$

Solving equations (24) and (25) for b_1 and b_2 yields:

$$\begin{aligned}
b_1 &= \frac{c\left(\sigma_2^{p2}\left(-k\left(\bar{a} - a^\star\right)\right) + \rho_2\sigma_2^p\sigma_2^v + \rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2} + 1\right)\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)} \qquad (26) \\
b_2 &= \frac{c\left(\sigma_1^{p2}k\left(\bar{a} - a^\star\right) + \rho_1\sigma_1^p\sigma_1^v + \rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2} + 1\right)\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)} \qquad (27)
\end{aligned}$$

■

Proof of Lemma 4:

Plugging in equation (22) gives:

$$a = \overline{a} + \frac{\sigma_2^{p2}\left(-k\left(\bar{a} - a^\star\right)\right) + \rho_2\sigma_2^p\sigma_2^v + \rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2} + 1\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)}\epsilon_1^p - \frac{\sigma_1^{p2}k\left(\bar{a} - a^\star\right) + \rho_1\sigma_1^p\sigma_1^v + \rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2} + 1\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)}\epsilon_2^p(2$$

The principal's objective function is:

$$\begin{aligned}
& \mathbb{E}\left[\left(\overline{a}+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)\epsilon_1^v+\left(1-\overline{a}-\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)\epsilon_2^v\right. \\
& \left.-\frac{c}{2}\left(\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2-H-\frac{k}{2}\left(\overline{a}-a^\star+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\right] \\
= & \left(\overline{a}\mathbb{E}\left[\epsilon_1^v\right]+\frac{b_1\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right]-b_2\mathbb{E}\left[\epsilon_2^p\epsilon_1^v\right]}{c}\right)+\left((1-\overline{a})\mathbb{E}\left[\epsilon_2^v\right]-\frac{b_1\mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right]-b_2\mathbb{E}\left[\epsilon_2^p\epsilon_2^v\right]}{c}\right) \\
& -\frac{1}{2c}\mathbb{E}\left[(b_1\epsilon_1^p-b_2\epsilon_2^p)^2\right]-H-\frac{k}{2}\mathbb{E}\left[\left(\overline{a}-a^\star+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\right] \\
= & 1+\frac{1}{c}\left(b_1\left(\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right]-1\right)+b_2\left(\mathbb{E}\left[\epsilon_2^p\epsilon_2^v\right]-1\right)\right) \\
& -\frac{1}{2c}\left(b_1^2\mathbb{E}\left[\epsilon_1^{p2}\right]+b_2^2\mathbb{E}\left[\epsilon_2^{p2}\right]-2b_1b_2\right)-H-\frac{k}{2}\mathbb{E}\left[\left(\overline{a}-a^\star+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\right] \qquad (29)
\end{aligned}$$

■

Proof of Lemma 5:

We differentiate the objective function of the principal in Lemma 4 with respect to $\overline{a}$ holding b_1 and b_2 constant:

$$k\left(a^\star-\overline{a}+\frac{-b_1+b_2}{c}\right) \qquad (30)$$

Since the objective function of the principal is concave in $\overline{a}$, setting this derivative to zero gives the optimal value of $\overline{a}$ in Lemma 5. ■

Proof of Proposition 1:

The proof combines the results in Lemmas 3 and 5 and reorganizes to get equation (17).

■

Proof of Case 4:

This proof builds on the proof of Proposition 1. Differentiating the principal's objective function with respect to $\overline{a}$ for $\sigma_1^p=0$ gives:

$$\frac{ck}{c+k}\left(a^\star-\bar{a}\right) \qquad (31)$$

This expression is positive if and only if $a^\star-\bar{a}$. Thus, if $\bar{a}<a^\star$, the principal would increase its objective function by increasing $\bar{a}$; if $\bar{a}>a^\star$, the principal would increase its objective

function by decreasing $\bar{a}$. Therefore, the principal's objective function is maximized at $\bar{a} = a^\star$.
■

Proof of Corollary 1:

Differentiating the principal's objective function with respect to $\overline{a}$ for $\sigma_1^p = 1$ and $\sigma_2^p = 1$ gives:

$$\frac{k\left((a^\star - \bar{a})\left(9c^2 + 12ck + 3k^2\right) + (c+k)(3\rho_2\sigma_2^v - 3\rho_1\sigma_1^v)\right)}{9(c+k)^2} \tag{32}$$

■

Proof of Propositions 2 and 3:

The contract includes a fixed wage, and the monetary benefit and allocation cost are separable in the agent's utility function. As a consequence, we have a matching model with transferable utility. The matching pattern depends on the cross-derivative of the surplus with respect to the two matching variables (Legros and Newman (2007)). The surplus created by a match is given by:

$$\mathbb{E}\left[V_1(a,\boldsymbol{\epsilon}) + V_2(a,\boldsymbol{\epsilon}) - \frac{k}{2}\left(a - a^\star\right)^2 - \frac{c}{2}\left(a - \overline{a}\right)^2\right]$$

With normalized volatilities, we have the following:

$$b_1 = \frac{c(-k(\overline{a} - a^\star) + 2\rho_1 + \rho_2)}{3(c+k)} \tag{33}$$

$$b_2 = \frac{c(k(\overline{a} - a^\star) + \rho_1 + 2\rho_2)}{3(c+k)} \tag{34}$$

The surplus when the contract is chosen optimally and the resource allocation is optimal is given by:

$$S = 1 + \frac{b_1\rho_1 + b_2\rho_2}{c} - \frac{c}{2}\mathbb{E}\left[\frac{\left(b_1\epsilon_1^p - b_2\epsilon_2^p\right)^2}{c}\right] - \frac{k}{2}\mathbb{E}\left[\left(\overline{a} - a^\star + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)^2\right] \tag{35}$$

Denoting $\Delta \equiv \overline{a} - a^\star$, we have:

$$S = 1 + \frac{b_1\rho_1 + b_2\rho_2}{c} - k\frac{\Delta^2}{2} - \Delta\frac{k(b_1 - b_2)}{c} - (b_1^2 + b_2^2 - b_1b_2)\frac{c+k}{c^2}. \tag{36}$$

We thus have:

$$\begin{aligned}
\frac{\partial b_1}{\partial \Delta} &= -\frac{ck}{3(c+k)} \\
\frac{\partial b_2}{\partial \Delta} &= \frac{ck}{3(c+k)} \\
\frac{\partial b_1}{\partial \rho_2} &= \frac{c}{3(c+k)}\frac{\partial b_1}{\partial \rho_1} = \frac{2c}{3(c+k)} \\
\frac{\partial b_2}{\partial \rho_2} &= \frac{2c}{3(c+k)}\frac{\partial b_2}{\partial \rho_1}\frac{c}{3(c+k)} \\
\frac{\partial b_1}{\partial c} &= \frac{k(2\rho_1+\rho_2-k\Delta)}{3(c+k)^2} \\
\frac{\partial b_2}{\partial c} &= \frac{k(\rho_1+2\rho_2+k\Delta)}{3(c+k)^2}
\end{aligned}$$

and

$$\begin{aligned}
\frac{dS}{d\Delta} &= \frac{k(\rho_2-\rho_1)}{3(c+k)} - k\Delta \\
\frac{d^2S}{d\Delta d\rho_1} &= -\frac{k}{3(c+k)} \\
\frac{d^2S}{d\Delta d\rho_2} &= \frac{k}{3(c+k)} \\
\frac{dS}{d\rho_2} &= \frac{b_2}{c} \\
\frac{dS}{d\rho_1} &= \frac{b_1}{c} \\
\frac{d^2S}{d\rho_2 dc} &= -\frac{b_2}{c(c+k)} \\
\frac{d^2S}{d\rho_1 dc} &= -\frac{b_1}{c(c+k)}
\end{aligned}$$

The signs of the cross derivatives $\frac{d^2S}{d\Delta d\rho_1}$ and $\frac{\partial b_2}{\partial c}$ immediately give the results. ■