Misaligned agents

Pierre Chaigneau* Queen's University

Nicolas Sahuguet† HEC Montréal

February 3, 2026

Abstract

How should firms select and incentivize managers who allocate resources based on private information when performance is measured asymmetrically across dimensions? We show that firms optimally hire managers whose preferences are biased toward poorly measured dimensions. This allows high-powered incentives on well-measured dimensions while relying on managerial preferences to prevent neglect of poorly-measured objectives. This framework can explain apparent inconsistencies between firms' stated objectives and the explicit incentives they provide, such as why ESG-committed firms provide weak ESG-based compensation while hiring ESG-motivated CEOs.

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

*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

Firms routinely face decisions that require allocating resources across multiple dimensions whose performance is measured with very different degrees of precision. A board designing executive compensation may care both about financial performance and environmental impact; a hospital about treatment intensity and preventive care; a university about research output and teaching quality.

In all these settings, explicit incentives tend to be concentrated on the better-measured dimension, while the organization's stated objectives usually emphasize a broader set of goals. This pattern has been widely documented: for example, ESG-committed firms relying almost exclusively on equity-based pay, or universities rewarding publications much more sharply than teaching quality, but it is often interpreted as evidence of window dressing, mission drift, or incomplete contracting.

This paper proposes a different interpretation. We show that when performance is asymmetrically measured and agents possess private information relevant for resource allocation, it can be optimal for principals to appoint misaligned agents, specifically agents whose intrinsic preferences are biased toward the poorly measured dimension. Doing so allows the principal to use high-powered incentives on the well-measured dimension, while relying on the agent's preferences to prevent excessive neglect of the poorly measured dimension. In this sense, misalignment is an optimal feature of the governance system rather than a governance failure.

Our analysis builds on a simple principal-agent model in which a manager allocates a fixed resource across two dimensions. On each dimension, output is imperfectly measured, and the manager privately observes information about how resource allocation affects both actual output and measured performance. Only measured performance on each dimension is contractible. Crucially, the manager has intrinsic preferences over the allocation decision, captured by a preferred baseline allocation that may differ from the principal's. These preferences matter because of the manager's private information: she can anticipate how her allocation decision affects measured performance and therefore compensation, as in Baker (1992). For example, a CEO familiar with the firm's production processes 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.

The logic can be explained as follows. Strong incentives improve the manager's responsiveness to her private information. However, with imperfect performance measurement, they distort resource allocation toward 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, optimal incentive contracts place higher weight on well-measured dimensions and lower weight on poorly measured ones. This asymmetry, however, creates a

systematic bias in the manager's decision, such that she tends to allocate too many resources to the well-measured dimension. Increasing incentives on the poorly measured dimension is not a solution, as it would exacerbate gaming and misallocation.

Our main result shows how this distortion can be mitigated through managerial selection. When performance is relatively poorly measured on one dimension, the principal optimally appoints a manager who intrinsically values that dimension more than he does. This bias counteracts the incentive-induced tilt toward the well-measured dimension, allowing the principal to maintain high-powered incentives where they are most effective. Relative to hiring a perfectly aligned manager, appointing a misaligned one leads to more extreme incentive asymmetry: incentives are muted further on the poorly measured dimension, and strengthened on the well-measured one.

By contrast, the standard multitasking model predicts low-powered incentives across all tasks when tasks are substitutes (Holmström and Milgrom (1991)). 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.

The model generates sharp comparative statics. Misalignment is more valuable when the manager's private information is more important, when performance is very well-measured on one dimension but poorly measured on the other, and when deviations from the principal's preferred allocation are costly. Conversely, the principal prefers an aligned agent when performance measures are either uninformative or fully revealing, or when the manager has no informational advantage. These results delineate precisely when organizations should appoint misaligned agents.

We then extend the analysis to an economy with heterogeneous firms and managers. Firms differ in the quality of their performance measurement, while managers differ in their intrinsic biases. In a matching equilibrium with transferable utility, we show that firms with worse measurement on a given dimension are matched with managers who are more biased toward that dimension. Intuitively, this bias is more valuable in these firms, so that they are willing to pay more to hire them. The model also predicts that agents who are biased toward the poorly measured dimension will earn the highest compensation, both because firms compete for them and because they must be compensated for optimally choosing allocations they personally dislike. Alternatively, the model can generate non-monotonic agency rents when firms prefer managers with intermediate biases.

The model delivers testable implications for incentive design. Managers who are more

biased toward a given dimension receive lower-powered incentives on that dimension. For example, executives who are more biased toward ESG should have lower-powered ESG-based incentives. Depending on the relative strength of incentive-adjustment and sorting effects, they may receive higher or lower-powered incentives on the other dimension. These predictions help rationalize observed patterns in executive compensation, public-sector pay, and professional organizations.

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.

Finally, the framework offers a novel perspective on organizational culture. If we interpret culture as the prevailing managerial bias within the organization, our model predicts that improved performance measurement on some dimension, such as increased financial transparency or new ESG metrics, will shift the corporate culture away from this dimension. The model also rationalizes apparent overcommitment to poorly measured objectives such as CSR or ESG (Berg, Kölbel, and Rigobon (2022)), which may reflect optimal responses to imperfect measurement rather than ideological capture by biased agents.

2 Related literature

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 principal-agent literature has shown that hiring aligned agents will generate efficiency gains. With imperfect or nonexistent performance measurement, hiring employees who share

[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. (2015)); 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)).

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.

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 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)).[2] 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

[2]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."

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.[3]

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.[4]

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. By contrast, we derive our main result in a simple model with one principal and one agent, so that it is not driven by market equilibrium effects. 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. Ederer, Holden, and Meyer (2018) study the advantages of opaque incentive schemes when agents facing multiple tasks have private information.

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.[5]

[3]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.

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

[5]Matching between principals and agents has been studied in a large literature reviewed in Macho-Stadler

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.[6] 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: the agent has only one task, making a resource allocation decision.[7]

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 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 $\boldsymbol{\epsilon} \equiv \{\epsilon_1^v, \epsilon_2^v, \epsilon_1^p, \epsilon_2^p\}$ which are privately observed by the agent at the time of making the resource allocation decision. Output and performance on both dimensions are:

$$V_1(a, \boldsymbol{\epsilon}) = a\epsilon_1^v \qquad \text{and} \qquad V_2(a, \boldsymbol{\epsilon}) = (1-a)\epsilon_2^v \tag{1}$$

$$P_1(a, \boldsymbol{\epsilon}) = a\epsilon_1^p \qquad \text{and} \qquad P_2(a, \boldsymbol{\epsilon}) = (1-a)\epsilon_2^p \tag{2}$$

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

and Pérez-Castrillo (2021).

[6]This asymmetric information assumption needs not be taken literally. 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. For example, a CEO can evaluate the merits of an investment project for her firm; a medical doctor can diagnose an illness and prescribe medical interventions when she sees a patient.

[7]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.

normalized such that: $\mathbb{E}\left[\epsilon_i^j\right] = 1$.[8] Define:

$$\rho_i \equiv \frac{cov(\epsilon_i^v, \epsilon_i^p)}{\sigma_i^v \sigma_i^p} \tag{3}$$

We assume that shocks to performances and outcomes are independent across dimensions:[9]

$$cov(\epsilon_i^p, \epsilon_j^v) = 0 \text{ and } cov(\epsilon_i^p, \epsilon_j^p) = 0 \qquad \forall i \neq j. \tag{4}$$

The objective function of the principal is:

$$\mathbb{E}\left[V_1(a, \boldsymbol{\epsilon}) + V_2(a, \boldsymbol{\epsilon}) - W(a, \boldsymbol{\epsilon}) - \frac{k}{2}(a - a^\star)^2\right] \tag{5}$$

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 $\{S, b_1, b_2\}$ such that the agent's payoff is equal to:

$$W(a, \boldsymbol{\epsilon}) = S + b_1 P_1(a, \boldsymbol{\epsilon}) + b_2 P_2(a, \boldsymbol{\epsilon}) \tag{6}$$

Linear contracts are a standard assumption in this literature (Baker (1992), Besley and Ghatak (2005), Fischer and Huddart (2008), Bénabou and Tirole (2016)). As established in the Online Appendix, it can be microfounded by allowing the agent to costlessly manipulate the performance measure distributions in a mean-preserving spread or mean-preserving contraction sense. As shown by Barron, Georgiadis, and Swinkels (2020), any nonlinear contract can then be gamed by the agent, and only linear contracts are robust to such distributional manipulation. This ability to manipulate the performance measure is arguably all the more relevant for managers, who have many ways to increase or decrease risk.

Given a contract $\{S, b_1, b_2\}$, the agent's information $\boldsymbol{\epsilon}$, and resource allocation a, the expected utility of the agent is:

$$\mathbb{E}\left[S + b_1 P_1(a, \boldsymbol{\epsilon}) + b_2 P_2(a, \boldsymbol{\epsilon}) - \frac{c}{2}(a - \alpha)^2\right] \tag{7}$$

where $c > 0$ parameterizes the cost to the agent of deviating from her preferred resource

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

[9] This does not imply that performance will be independent across dimensions: a different resource allocation will improve one dimension of performance at the expense of the other one. For example, cost reductions often reduce quality. Instead, we are only assuming that shocks to performance are independent across dimensions.

allocation α. 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 (Holmström and Milgrom (1991)), 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)).[10] 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)).[11]

There are several interpretations for the cost of deviating from a baseline resource allocation $a^\star$ in the principal's objective function in equation (5). 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 "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

[10] 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.

[11] 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.

to Besley and Ghatak (2005, 2018)). The agent's objective function in equation (7) 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 α 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 (7). In practice, there is evidence that many economic agents have "professional motivations", i.e. preferences over actions (e.g. Hall (1996), Stevens et al. (2015), 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.

At the first-best, when the principal can offer a forcing contract after observing the state $\boldsymbol{\epsilon}$ and the agent's resource allocation decision a, he optimally hires an aligned agent, i.e. $\alpha = a^\star$. Intuitively, this allows to simultaneously align interests with respect to resource allocation while also minimizing the cost of the agent's compensation. Since the agent will need to be compensated for the deviation between the chosen resource allocation and her preferred resource allocation, hiring an aligned agent allows to implement the principal's preferred resource allocation at the lowest cost.

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 α that the principal prefers. We start by deriving several preliminary results.

Lemma 1 *For a given contract $\{S, b_1, b_2\}$, the agent optimally chooses the following resource allocation:*

$$a = \alpha + \frac{b_1 \epsilon_1^p - b_2 \epsilon_2^p}{c} \tag{8}$$

The optimally chosen resource allocation is additive in two terms. The first is the agent's baseline resource allocation α. 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.

Lemma 2 determines the optimal pay-performance sensitivity (PPS).

Lemma 2 *The optimal sensitivity of pay to performance on both dimensions is:*

$$b_1 = \frac{c\left(\sigma_2^{p2}\left(-k\left(\alpha-a^\star\right)\right)+\rho_2\sigma_2^p\sigma_2^v+\rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2}+1\right)\right)}{\left(c+k\right)\left(\sigma_1^{p2}\sigma_2^{p2}+\sigma_1^{p2}+\sigma_2^{p2}\right)} \tag{9}$$

$$b_2 = \frac{c\left(\sigma_1^{p2}k\left(\alpha-a^\star\right)+\rho_1\sigma_1^p\sigma_1^v+\rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2}+1\right)\right)}{\left(c+k\right)\left(\sigma_1^{p2}\sigma_2^{p2}+\sigma_1^{p2}+\sigma_2^{p2}\right)} \tag{10}$$

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

$$b_1 = \frac{1}{\sigma_1^{p2}+1}\left(\frac{\rho_1\sigma_1^v\sigma_1^p-k(\alpha-a^\star)}{1+\frac{k}{c}}+b_2\right) \tag{11}$$

$$b_2 = \frac{1}{\sigma_2^{p2}+1}\left(\frac{\rho_2\sigma_2^v\sigma_2^p+k(\alpha-a^\star)}{1+\frac{k}{c}}+b_1\right) \tag{12}$$

This shows that PPS on dimension i is additive in three terms.

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.[12] 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, α, 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,

[12]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)).

when the agent is biased toward dimension 2 relative to the principal, i.e. $\alpha < a^\star$, 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 2.

We now consider which preferred resource allocation α of the agent maximizes the principal's objective function. We refer to this value of α as the one that the principal "prefers".

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

$$\alpha = a^\star + \frac{\rho_2 \sigma_2^p \sigma_2^v \sigma_1^{p2} - \rho_1 \sigma_1^p \sigma_1^v \sigma_2^{p2}}{c\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right) + k\sigma_1^{p2}\sigma_2^{p2}} \tag{13}$$

Thus, the principal prefers to appoint an agent with $\alpha < a^\star$ if and only if

$$\rho_2 \frac{\sigma_2^v}{\sigma_2^p} < \rho_1 \frac{\sigma_1^v}{\sigma_1^p}.$$

This condition shows how the principal's preference for the agent's type depends on the performance measurement system. The principal will rather appoint a misaligned agent with $\alpha < a^\star$ 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.

For example, with normalized volatilities $\sigma_1^v = \sigma_1^p = \sigma_2^v = \sigma_2^p = 1$, performance is better measured on the first dimension when $\rho_1 > \rho_2$. In this case, the principal would rather appoint an agent whose bias is to allocate resources to the second dimension, i.e. $\alpha < a^\star$. 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 stronger incentives on resource allocation, the principal appoints an agent who is more biased in favor of the second dimension.

Example 1 *(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 (9)). 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. $\alpha < a^\star$. This further reduces the sensitivity of pay to ESG performance, b_2 (see equation (10)).

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. Corollary 1 makes these points formally.

Corollary 1 *The principal prefers to appoint an aligned agent in any of the following cases:*
(i) There is no asymmetric information between the principal and the agent about the optimal resource allocation, i.e. $\sigma_1^v = \sigma_2^v = 0$;
(ii) Performance measurement is unrelated to output, i.e. $\rho_1 = \rho_2 = 0$;
(iii) Performance measures perfectly reveals output, i.e. $\sigma_1^p = \sigma_1^v$, $\sigma_2^p = \sigma_2^v$, and $\rho_1 = \rho_2 = 1$;
(iv) Performance on some dimension reveals the resource allocation but is unrelated to output, i.e. $\sigma_1^p = 0$.

Case (i). 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. $\alpha = a^\star$, and not providing any explicit incentives, i.e. $b_1 = b_2 = 0$.

Case (ii). When $\rho_1 = \rho_2 = 0$, "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 $\alpha = a^\star$.

Case (iii). When output is contractible, 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 (iv). 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.

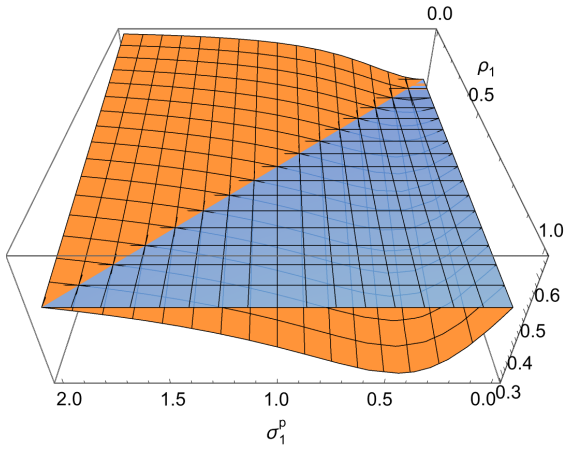


Figure 1: Orange surface: optimal α as a function of σ_1^p and ρ_1, for $a^\star = 0.5$, $\sigma_1^v = \sigma_2^v = \sigma_2^p = 1$, $\rho_2 = 0.5$, $c = k = 1$, and $\boldsymbol{\epsilon}$ follows a multivariate normal distribution. Translucent blue surface: $a^\star = 0.5$.

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

[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.

pay to performance on these two dimensions, the principal optimally appoints an agent who is biased in favor of the latter.

5 Matching in competitive markets

The baseline analysis shows that a firm with asymmetric performance measurement optimally hires a biased agent. We now derive the equilibrium matching and associated compensation packages in a competitive market with many firms and many managers.

Consider an economy with heterogeneous firms and heterogeneous managers. For example, firms differ in the quality of their ESG measurement, while managers can be either profit-motivated or ESG motivated. To make the exposition more concrete and improve readability, we will use this example throughout, without loss of generality.

We establish three main results. First, positive assortative matching (PAM): firms with better ESG measurement hire more profit-motivated managers. Intuitively, profit focus and good ESG measurement are complements – together they enable efficient allocation across both dimensions. Second, non-monotone rents: the highest-paid managers are not the most or least biased, but those whose bias best complements prevailing measurement technologies. Third, incentive substitution: managerial preferences substitute for explicit incentives. ESG-motivated managers receive stronger financial incentives and weaker ESG incentives, while profit-motivated managers receive stronger ESG incentives and weaker financial incentives. However, when sorting effects are strong (small k or steep matching), profit-motivated managers can also receive higher financial incentives due to matching with firms with better measurement systems.

These results generate testable predictions about the joint distribution of measurement quality, manager types, compensation levels, and incentive structures across firms.

5.1 Setup and equilibrium

We consider an economy populated by a continuum of agents and a continuum of principals. Otherwise identical principals have performance measurement systems with different ρ_2, which proxies for the quality of ESG measurement: the type of principal i, ρ_2^i, is distributed on $[\underline{\rho}, \bar{\rho}]$ with CDF F_ρ. Otherwise identical agents have different preferred resource allocations α: the type of agent i, α^i, is distributed on $[\underline{a}, \bar{a}]$ with CDF F_A. We simplify the model by analyzing the normalized-volatility case with $\sigma_1^v = \sigma_2^v = \sigma_1^p = \sigma_2^p = 1$.

We focus on the case where firms differ in ESG measurement quality (ρ_2) while having similar financial measurement quality (ρ_1). This is empirically motivated: financial performance

is measured relatively uniformly across firms through stock market valuation and audited accounts, whereas ESG measurement varies substantially depending on industry, disclosure practices, and the nature of ESG activities.

We consider a matching model with transferable utility. Let $Y(\rho_1, \rho_2; \alpha)$ denote the implementable joint surplus of a firm-agent match – the principal's objective function plus the agent's expected utility – which is not affected by transfers. This joint surplus is:

$$Y(\rho_1, \rho_2; \alpha) = \mathbb{E}\left[V_1 + V_2 - \frac{k}{2}(a - a^\star)^2 - \frac{c}{2}(a - \alpha)^2\right] \tag{14}$$

when the principal optimally chooses (b_1, b_2) and the agent optimally chooses the resource allocation. 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 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.

The matching pattern depends on the cross-derivative of the surplus with respect to the two matching variables (Legros and Newman (2007)).

5.2 Matching pattern

We first establish the key property of the surplus function that determines the matching pattern.

Lemma 3 *The surplus of a match $Y(\rho_1, \rho_2^i, \alpha^i)$ is supermodular in ρ_2^i and α^i:*

$$\frac{\partial^2 Y(\rho_1, \rho_2^i, \alpha^i)}{\partial \rho_2^i \partial \alpha^i} > 0 \tag{15}$$

The positive cross-derivative means that better ESG measurement (higher ρ_2) creates more surplus when matched with agents who are more profit-focused (higher α). Intuitively, a profit-focused agent naturally allocates toward the well-measured (financial) dimension. Good ESG measurement allows the firm to provide strong ESG incentives to rebalance allocation.

Supermodularity in a matching model with transferable utility leads to positive assortative matching:

Proposition 2 *The matching equilibrium involves positive assortative matching (PAM): principals with high ρ_2^i are matched with agents with high α^i. The matching function $\rho_2(\alpha)$, which gives the type ρ_2^i of the principal matched with an agent with type α^i, is strictly increasing.*

Firms vary in their ESG measurement quality ρ_2 while having similar financial measurement quality ρ_1. Agents vary in their profit orientation, with high α indicating a bias toward profit rather than ESG. Proposition 2 implies that firms with the best ESG measurement (highest ρ_2) hire the most profit-motivated managers (highest α) and that firms with the worst ESG measurement (lowest ρ_2) hire the most ESG-motivated managers (lowest α).

This PAM pattern arises because profit-focused agents and good ESG measurement are complements. A profit-motivated manager at a firm with good ESG measurement creates high value because she naturally focuses on financial performance, so that the firm can provide strong ESG incentives to ensure adequate ESG attention. This helps achieve a more efficient allocation. By contrast, an ESG-motivated manager creates most value at firms with poor ESG measurement, where her preferences substitute for weak incentives. The matching is characterized by the quantile condition:

$$F_A(\alpha^i) = F_\rho(\rho_2(\alpha^i)) \tag{16}$$

Equivalently, the matching function can be written as:

$$\rho_2(\alpha^i) = F_\rho^{-1}(F_A(\alpha^i)) \tag{17}$$

5.3 Equilibrium rents

We now characterize the rents earned by different types of agents in the matching equilibrium. The rent $U(\alpha^i)$ of an agent with type α^i is defined as:

$$U(\alpha^i) \equiv \mathbb{E}[W(a,\epsilon)|\alpha = \alpha^i] - \frac{c}{2}(a-\alpha^i)^2 - H \tag{18}$$

where a is optimally chosen and H is the agent's reservation utility.

Proposition 3 *The marginal rent earned by an agent with type α is:*

$$U'(\alpha) = \frac{k}{3(c+k)}\left[\rho_2(\alpha) - \rho_1 - (3c+k)(\alpha - a^\star)\right] \tag{19}$$

The rent schedule is:

$$U(\alpha) = U(\underline{a}) + \int_{\underline{a}}^{\alpha} \left\{ \frac{k(\rho_2(x)-\rho_1)}{3(c+k)} - \frac{k(3c+k)}{3(c+k)}(x-a^\star) \right\} dx \tag{20}$$

An agent's marginal rent is: (i) increasing in $\rho_2(\alpha)$: it is higher if the matched firm has poor ESG measurement, which increases the value of ESG-biased agents; (ii) decreasing in ρ_1:

it is lower if financial measurement is good, which reduces the need for the complementary bias; (iii) decreasing in $(3c + k)(\alpha - a^\star)$: it is lower if the agent's bias differs greatly from principal's preference, because of the higher deadweight loss from misalignment.

The rent function is generally non-monotone: agents with extreme biases (very high or very low α) earn lower rents than agents with intermediate biases, depending on the distribution of firm types and the parameters.

We can derive a closed-form characterization under uniform distributions.

Proposition 4 *Assume $\rho_2 \sim \mathcal{U}[\underline{\rho}, \bar{\rho}]$ and $\alpha \sim \mathcal{U}[\underline{a}, \overline{a}]$.*
(i) If $a^\star \in (\underline{a}, \overline{a})$ and $\rho_1 \in (\underline{\rho}, \bar{\rho})$, the rent function $U(\alpha)$ is single-peaked, maximized at an interior value α^* satisfying:

$$\alpha^* = a^\star + \frac{(\underline{\rho} + \bar{\rho})/2 - \rho_1}{3c + k} \tag{21}$$

where $(\underline{\rho} + \bar{\rho})/2$ is the average ESG measurement quality.
(ii) If $\rho_1 > (\underline{\rho} + \bar{\rho})/2$, then $\alpha^* < a^\star$: agents who are more ESG-biased than the principals earn the highest rents.

Intuitively, agents who earn the highest rents are those whose bias complements the average measurement technology in the economy. When $\rho_1 > (\underline{\rho} + \bar{\rho})/2$, which means that financial measurement is better than average ESG measurement across firms, the agents who earn the highest rents are those who are more ESG-biased than principals.

Example 2 *We compute numerically the rent schedule in two cases. In the first case, we assume: $\rho_2 \sim \mathcal{U}[0.3, 0.7]$, $\alpha \sim \mathcal{U}[0.3, 0.7]$, $a^\star = 0.5$, $\rho_1 = 0.9$, $c = k = 1$. The highest rent is at $\alpha^* = 0.4$, which is below $a^\star = 0.5$: the agent is ESG-biased relative to the principal. Thus, ESG-motivated managers earn the highest rents when financial measurement is much better than ESG measurement. In the second case, ESG measurement is worse: $\rho_2 \sim \mathcal{U}[0.1, 0.4]$ but we still have $\rho_1 = 0.9$. The highest rent is now at $\alpha^* = 0.34$: the manager with the highest rent is even more ESG-biased than in the first case.*

Figure 2 depicts the two rent schedules. It shows that worse ESG measurement increases demand for more ESG-motivated managers. This increases rents overall: the red dashed curve is higher than the blue curve for all agent types. Intuitively, competition for ESG-motivated managers intensifies when ESG is harder to measure. Accordingly, the decline in rents is steeper in the second case.

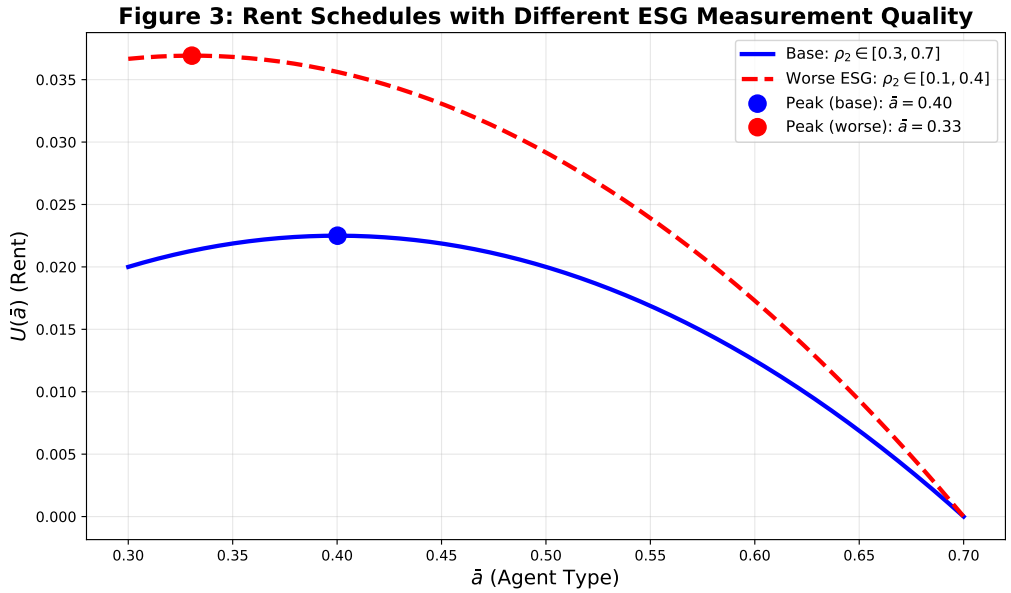


Figure 2: Comparison of rent schedules under different ESG measurement quality in Example 2. The blue solid line is the agency rent in the first case with $\rho_2 \in [0.3, 0.7]$. The red dashed line is the agency rent in the second case with $\rho_2 \in [0.1, 0.4]$.

5.4 Equilibrium incentives

How do incentive structures vary across agents in the matching equilibrium? We decompose the effect of agent type on incentives into two components: a direct effect (holding the matched firm constant) and an indirect sorting effect (through the change in matched firm type).

Proposition 5 *Consider two agents with types $\alpha_H > \alpha_L$ (agent H is more profit-motivated, less ESG-motivated). There are two effects:*

(i) Direct effect: Holding firm type constant, agent H receives lower-powered incentives on financial performance (dimension 1) and higher-powered incentives on ESG (dimension 2).

(ii) Sorting effect: Under PAM, agent H is matched to a firm with better ESG measurement: $\rho_2(\alpha_H) > \rho_2(\alpha_L)$. This increases the power of incentives on both dimensions.

The overall effect on ESG incentives (dimension 2) is positive. The overall effect on financial incentives (dimension 1) is positive if k is sufficiently large, and negative otherwise.

Both the direct effect and the sorting effect are positive for ESG incentives, so profit-motivated agents always receive higher-powered ESG incentives. However, these two effects are opposite for financial incentives. When k is sufficiently large relative to the matching slope $\rho_2'(\alpha)$ (which can be expressed as a function of model primitives via equation 17), the direct effect dominates, so that ESG-motivated agents receive higher-powered financial incentives. Otherwise, when $\rho_2'(\alpha)$ is sufficiently large relative to k, the sorting effect dominates, so that

ESG-motivated agents receive lower-powered financial incentives. Intuitively, when there is either a lot of variation in firm measurement or not much variation in agents' types, agents with only slightly different types will be matched to firms with substantially different performance measurement, so that changes in firm types matter more than changes in agents' types.

To illustrate these two regimes, Figure 3 uses the same parameter values as in Example 2 but varies k to show when the direct effect or the sorting effect dominates. In Panel A, with $k = 2$, the direct effect dominates: more ESG-motivated agents get stronger financial incentives and weaker ESG incentives. In Panel B, with $k = 0.5$, the sorting effect dominates: more ESG-motivated agents get stronger incentives in both dimensions.

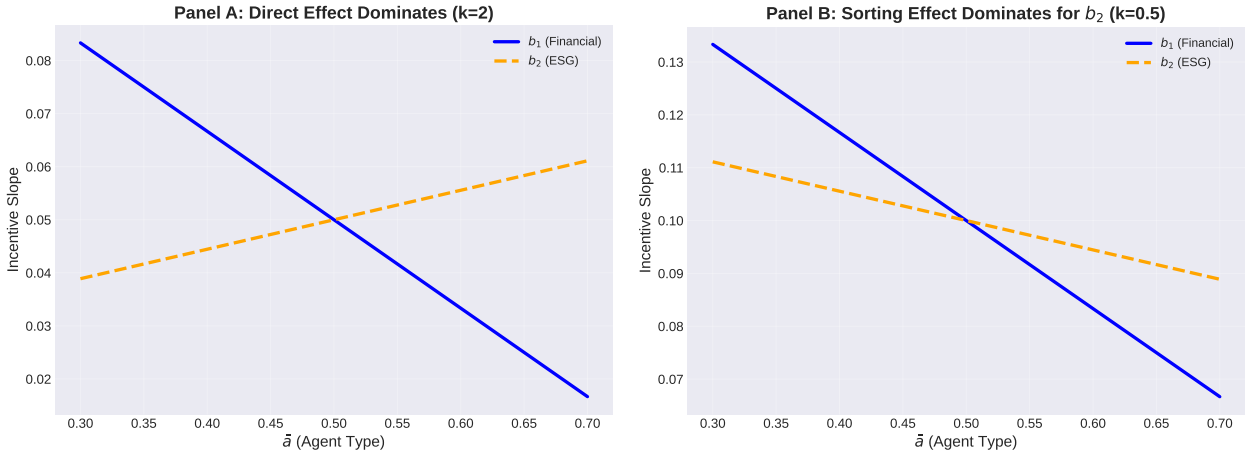


Figure 3: Equilibrium incentive slopes as a function of agent type.

When principals care enough their preferred resource allocation (high k), more ESG-motivated agents (low α) receive higher-powered financial incentives and lower ESG incentives. Their ESG motivation substitutes for explicit ESG incentives, allowing firms to provide high-powered incentives on the well-measured financial dimension. This explains why ESG-motivated CEOs are mostly compensated based on financial metrics despite their sustainability focus.

5.5 Changes in distributions

In our matching model with transferable utility, the matching equilibrium is constrained efficient when matches are independent and measurement technologies are exogenous. We now consider changes in the distribution of agents' types and in measurement technologies.

Define total welfare as the total surpluses across all principal-agent pairs:

$$\mathcal{W} = \int_{\underline{a}}^{\overline{a}} Y(\rho_1, \rho_2(\alpha); \alpha)\, dF_A(\alpha) \tag{22}$$

where $Y(\rho_1, \rho_2; \alpha)$ is the joint surplus from a match between a firm with measurement quality (ρ_1, ρ_2) and an agent with preferred allocation α, and $\rho_2(\alpha)$ is the equilibrium matching function. With a unit measure of agents and firms, welfare is the expected surplus over the distribution of agent types. This section explores two policy interventions that affect welfare: (i) improvements in measurement technologies, and (ii) shifts in the distribution of agent types.

To understand when shifts in these distribution increase or decrease welfare, we decompose welfare into two components:

$$\mathcal{W} = \underbrace{\frac{c}{6(c+k)} \int (\rho_1 + \rho_2(\alpha))^2 \, dF_A(\alpha)}_{\text{Measurement benefit}} - \underbrace{\frac{k(3c+k)}{6(c+k)} \int (\alpha - a^\star)^2 \, dF_A(\alpha)}_{\text{Misalignment cost}} \tag{23}$$

Improvements in the quality of performance measurement always improve welfare, regardless of whether they apply to all firms or not.[14]

We now consider changes in the agent type distribution. The preference misalignment between a principal and a given agent, captured by $(\alpha - a^\star)$, affects welfare because it determines what contracts the principal must offer to induce some actions. Greater misalignment requires larger incentive distortions, reducing surplus. Define the second moment around $a^\star$ as:

$$M(F_A) = \int (\alpha - a^\star)^2 \, dF_A(\alpha) \tag{24}$$

With uniform distributions, the welfare effect of a change in the agent type distribution only depends on the effect of the misalignment cost term in equation (23), so that $M(F_A)$ is a sufficient statistic for welfare comparisons.[15] Specifically, a distribution shift from F_A to F'_A increases welfare if and only if:

$$M(F'_A) < M(F_A) \tag{25}$$

The shift reduces welfare if it increases the average squared distance from $a^\star$. For uniform distributions $\alpha \sim \mathcal{U}[\underline{a}, \overline{a}]$:

$$M(F_A) = \frac{(\overline{a} - \underline{a})^2}{12} + \left(\frac{\underline{a} + \overline{a}}{2} - a^\star \right)^2 = \text{Variance} + \text{Bias}^2 \tag{26}$$

[14] Whereas uniform improvements in measurement quality unambiguously improve welfare because they preserve the matching pattern, equation (23) shows that non-uniform improvements in measurement also improve welfare. Intuitively, the direct effect of better performance measurement is positive given existing matches, and the indirect effect via better matches is also positive.

[15] With the equilibrium matching ($F_A(\alpha) = F_\rho(\rho_2(\alpha))$), uniform distributions imply $\mathbb{E}[\rho_2(\alpha)] = \mathbb{E}[\rho_2]$, which is independent of the agent type distribution. When both the performance measurement distribution F_ρ and the agent type distribution F_A are uniform, the measurement benefit term is constant (equilibrium matching with uniform distributions preserves all moments of $\rho_2(\alpha)$).

By symmetry, when the agent type distribution F_A is fixed and the performance measurement distribution F_ρ changes (e.g., ESG measurement improves), the misalignment cost remains constant and welfare changes come entirely from the measurement benefit. For uniform distributions with $\rho_2 \sim \mathcal{U}[\underline{\rho}, \overline{\rho}]$, improvements that increase $\mathbb{E}[\rho_2]$ and $\mathbb{E}[\rho_2^2]$ (i.e., shift the distribution rightward) always increase welfare. This contrasts with shifts in agent type distribution, where welfare can decrease if the shift increases misalignment.

This leads to the following implications: (i) A symmetric compression around $a^\star$ that reduces the variance while leaving the bias unchanged always increases welfare. (ii) A horizontal translation of the whole agent type distribution increase welfare if and only if the new mean is closer to $a^\star$. (iii) Asymmetric shifts have ambiguous effects: welfare depends on whether variance reduction outweighs bias increase.

Example 3 *Consider the first case in Example 2, but now suppose that there is a leftward shift in the agent type distribution from* $\alpha \sim \mathcal{U}[0.3, 0.7]$ *to* $\alpha \sim \mathcal{U}[0.2, 0.5]$ *while performance measurement remains distributed as* $\rho_2 \sim \mathcal{U}[0.3, 0.7]$*. This means that the pool of available managers becomes more ESG-motivated, for example because of societal cultural changes. The shift reduces variance but increases the bias*^{2} *term more, so* M *increases and welfare decreases by 3.45%. Figure 4 illustrates the welfare effects: 71% of firms experience surplus losses while only the top 29% gain.*

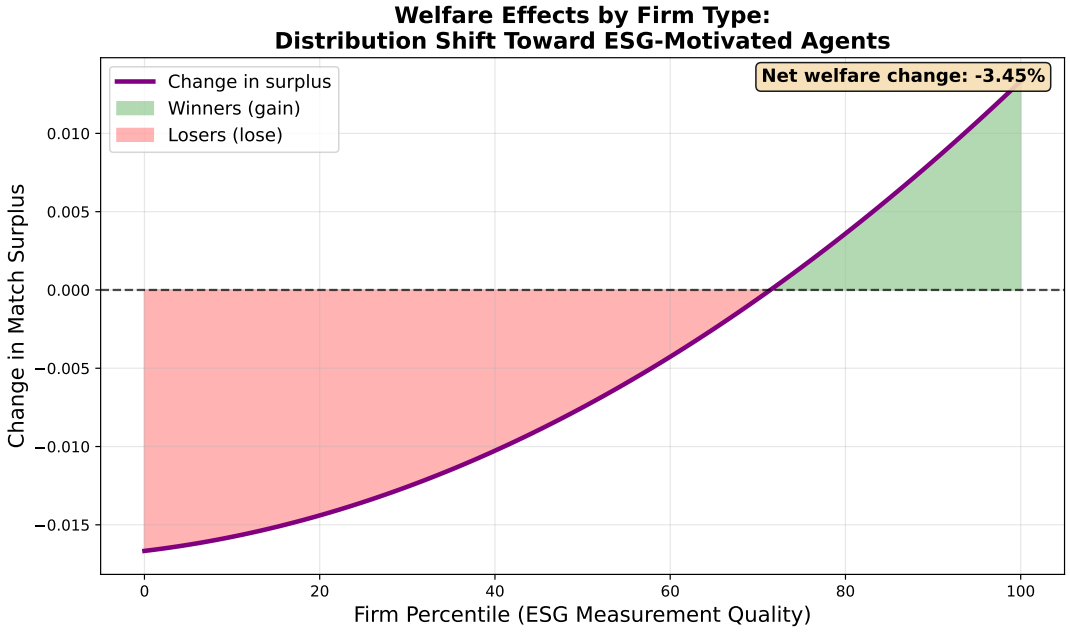


Figure 4: Welfare effects of distribution shift toward ESG-motivated agents

Under PAM with quantile matching, the kth percentile agent matches to the kth percentile firm both before and after. However, firm types remain fixed while agent identities change. For

example, the 100th percentile firm ($\rho_2 = 0.7$) matched to an agent with $\alpha = 0.7$ before, but now matches to an agent with $\alpha = 0.5$ (the most profit-motivated available). The surplus function includes a convex misalignment cost $k(3c + k)(\alpha - a^\star)^2/(6(c + k))$. The shift is asymmetric around $a^\star = 0.5$: before, agents ranged from 0.2 below $a^\star$ to 0.2 above (symmetric); after, agents range from 0.3 below $a^\star$ to 0 above (asymmetric, all at or below $a^\star$). Most firms (0-71st percentile) match with agents further from $a^\star$, increasing squared misalignment. Only top firms (72-100th percentile) get agents closer to $a^\star$. By convexity, losses exceed gains.

This provides a surprising policy implication. Policies that increase the supply of ESG-motivated managers (e.g., sustainability education, prestige campaigns) can decrease aggregate welfare even when ESG-motivation is valued by firms due to poor ESG measurement. The welfare loss arises from increased misalignment for most firms, not from loss of complementarity. This occurs despite ESG being poorly measured: the shift moves agents too far from efficient allocation even accounting for measurement asymmetries. The remaining profit-motivated agents become scarce, earning higher rents but unable to offset the aggregate welfare loss.

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.

[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)).

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 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]

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 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.[18]

6.3 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

[17]This can be achieved indirectly, including in hiring and promotion decisions, but also by celebrating 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]Alternatively, if all measures are easy to game, then incentives will be low-powered on all dimensions, as also predicted by other models.

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.[19] 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 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. This highlights a new reason for the importance of jointly determining compensation policy and hiring policy.

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

[19] 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.

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 multifaceted and 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., 2008. Identity, supervision, and work groups. *American Economic Review*, 98(2), 212-217.

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

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

Barron, D., Georgiadis, G. and Swinkels, J., 2020. Optimal contracts with a risk-taking agent. *Theoretical Economics*, 15(2), 715-761.

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.

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.

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, 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., 2025. 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.

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

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., 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.

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.

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.

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.

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.

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, M., 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.

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 1: The agent chooses a to maximize her objective function in equation (7). 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:

$$S + b_1 a\epsilon_1^p + b_2(1-a)\epsilon_2^p - \frac{c}{2}(a-\alpha)^2 \tag{27}$$

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

$$b_1\epsilon_1^p - b_2\epsilon_2^p - c(a-\alpha) = 0 \tag{28}$$

Rearranging gives the expression in Lemma 1. ■

Proof of Lemma 2: Plugging equation (6) into equation (5), the objective function of the principal is:

$$\max_{b,S} \mathbb{E}\left[V_1(a,\epsilon) + V_2(a,\epsilon) - S - b_1 P_1(a,\epsilon) - b_2 P_2(a,\epsilon) - \frac{k}{2}(a-a^\star)^2\right] \tag{29}$$

The participation constraint and incentive constraint are respectively:

$$S + b_1 P_1(a,\epsilon) + b_2 P_2(a,\epsilon) - \frac{c}{2}(a-\alpha)^2 \geq H \tag{30}$$

$$b_1\epsilon_1^p - b_2\epsilon_2^p - c(a-\alpha) = 0 \quad \Leftrightarrow \quad a = \alpha + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c} \tag{31}$$

Substituting in equation (29) for the binding participation constraint in equation (30), the principal's optimization program rewrites as:

$$\max_{b} \mathbb{E}\left[V_1(a,\epsilon) + V_2(a,\epsilon) - \frac{c}{2}(a-\alpha)^2 - H - \frac{k}{2}(a-a^\star)^2\right] \tag{32}$$

This objective function can be rewritten as:

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

The first-order condition with respect to b_1 is:

$$\begin{aligned}
& \mathbb{E}\left[\frac{\epsilon_1^p}{c}\epsilon_1^v - \frac{\epsilon_1^p}{c}\epsilon_2^v - c\epsilon_1^p\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c^2}\right) - k\frac{\epsilon_1^p}{c}\left(\alpha - a^\star + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\right] = 0 \\
\Leftrightarrow \quad & \mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right] - \mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right] - b_1\mathbb{E}\left[\epsilon_1^{p2}\right] + b_2\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right] - k\left((\alpha - a^\star)\mathbb{E}\left[\epsilon_1^p\right] + \frac{b_1}{c}\mathbb{E}\left[\epsilon_1^{p2}\right] - \frac{b_2}{c}\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right]\right) = 0 \\
\Leftrightarrow \quad & b_1 = \frac{\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right] - \mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right] - k(\alpha - a^\star)\mathbb{E}\left[\epsilon_1^p\right] + b_2\left(1 + \frac{k}{c}\right)\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right]}{\left(1 + \frac{k}{c}\right)\mathbb{E}\left[\epsilon_1^{p2}\right]} \\
\Leftrightarrow \quad & b_1 = \frac{\rho_1\sigma_1^v\sigma_1^p - k(\alpha - a^\star)}{\left(1 + \frac{k}{c}\right)\left(\sigma_1^{p2} + 1\right)} + b_2\frac{1}{\sigma_1^{p2} + 1} \qquad (33)
\end{aligned}$$

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

$$\begin{aligned}
& \mathbb{E}\left[-\frac{\epsilon_2^p}{c}\epsilon_1^v + \frac{\epsilon_2^p}{c}\epsilon_2^v + c\epsilon_2^p\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c^2}\right) + k\frac{\epsilon_2^p}{c}\left(\alpha - a^\star + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\right] = 0 \\
\Leftrightarrow \quad & b_2 = \frac{\rho_2\sigma_2^v\sigma_2^p + k(\alpha - a^\star)}{\left(1 + \frac{k}{c}\right)\left(\sigma_2^{p2} + 1\right)} + b_1\frac{1}{\sigma_2^{p2} + 1} \qquad (34)
\end{aligned}$$

Solving equations (33) and (34) for b_1 and b_2 yields:

$$\begin{aligned}
b_1 &= \frac{c\left(\sigma_2^{p2}\left(-k\left(\alpha - a^\star\right)\right) + \rho_2\sigma_2^p\sigma_2^v + \rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2} + 1\right)\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)} \qquad (35) \\
b_2 &= \frac{c\left(\sigma_1^{p2}k\left(\alpha - a^\star\right) + \rho_1\sigma_1^p\sigma_1^v + \rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2} + 1\right)\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)} \qquad (36)
\end{aligned}$$

■

Proof of Proposition 1: The principal's objective function is:

$$\begin{aligned}
& \mathbb{E}\Bigg[\left(\alpha+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)\epsilon_1^v+\left(1-\alpha-\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)\epsilon_2^v \\
& -\frac{c}{2}\left(\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2-H-\frac{k}{2}\left(\alpha-a^\star+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\Bigg] \\
= & \left(\alpha\mathbb{E}\left[\epsilon_1^v\right]+\frac{b_1\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right]-b_2\mathbb{E}\left[\epsilon_2^p\epsilon_1^v\right]}{c}\right)+\left((1-\alpha)\mathbb{E}\left[\epsilon_2^v\right]-\frac{b_1\mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right]-b_2\mathbb{E}\left[\epsilon_2^p\epsilon_2^v\right]}{c}\right) \\
& -\frac{1}{2c}\mathbb{E}\left[(b_1\epsilon_1^p-b_2\epsilon_2^p)^2\right]-H-\frac{k}{2}\mathbb{E}\left[\left(\alpha-a^\star+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\right] \\
= & 1+\frac{1}{c}\left(b_1\left(\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right]-1\right)+b_2\left(\mathbb{E}\left[\epsilon_2^p\epsilon_2^v\right]-1\right)\right) \\
& -\frac{1}{2c}\left(b_1^2\mathbb{E}\left[\epsilon_1^{p2}\right]+b_2^2\mathbb{E}\left[\epsilon_2^{p2}\right]-2b_1b_2\right)-H-\frac{k}{2}\mathbb{E}\left[\left(\alpha-a^\star+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\right] \qquad (37)
\end{aligned}$$

where b_1 and b_2 are described in Lemma 2. Differentiating this objective function with respect to α holding b_1 and b_2 constant gives:

$$k\left(a^\star-\alpha+\frac{-b_1+b_2}{c}\right) \qquad (38)$$

Since the objective function of the principal is concave in α, setting this derivative to zero gives:

$$\alpha=a^\star+\frac{b_2-b_1}{c} \qquad (39)$$

Combining with the optimal values of b_1 and b_2 in Lemma 2 and reorganizing gives equation (13). ■

Proof of Corollary 1: The proofs of parts (i)-(iii) are special cases of the proof of Proposition 1.

The proof of part (iv) builds on the proof of Lemma 1. Differentiating the principal's objective function with respect to α for $\sigma_1^p=0$ gives:

$$\frac{ck}{c+k}\left(a^\star-\alpha\right) \qquad (40)$$

This expression has the same sign as $a^\star-\alpha$. Thus, if $\alpha<a^\star$, the principal would increase its objective function by increasing α; if $\alpha>a^\star$, the principal would increase its objective function by decreasing α. Therefore, the principal's objective function is maximized at $\alpha=a^\star$. ■

Proof of Lemma 3: With normalized volatilities $\sigma_1^v = \sigma_2^v = \sigma_1^p = \sigma_2^p = 1$ and ρ_1 held constant, the surplus when the contract is chosen optimally is:

$$Y(\rho_1, \rho_2; \alpha) = 1 + \frac{b_1\rho_1 + b_2\rho_2}{c} - \frac{k\Delta^2}{2} - \Delta\frac{k(b_1 - b_2)}{c} - (b_1^2 + b_2^2 - b_1 b_2)\frac{c+k}{c \cdot 2} \tag{41}$$

where $\Delta \equiv \alpha - a^\star$, and with optimal incentive slopes:

$$b_1(\rho_1, \rho_2, \alpha) = \frac{c(-k\Delta + 2\rho_1 + \rho_2)}{3(c+k)} \tag{42}$$

$$b_2(\rho_1, \rho_2, \alpha) = \frac{c(k\Delta + \rho_1 + 2\rho_2)}{3(c+k)} \tag{43}$$

Taking the derivative of Y with respect to α (using $\frac{\partial\Delta}{\partial\alpha} = 1$ and the chain rule):

$$\frac{\partial Y}{\partial \alpha} = \frac{k(\rho_2 - \rho_1)}{3(c+k)} - k\Delta \tag{44}$$

Now taking the derivative with respect to ρ_2:

$$\frac{\partial^2 Y}{\partial\rho_2\partial\alpha} = \frac{k}{3(c+k)} > 0 \tag{45}$$

This establishes supermodularity. ■

Proof of Proposition 5: From the optimal contract formulas:

$$b_1 = \frac{c(-k(\alpha - a^\star) + \rho_1 + \rho_2)}{6(c+k)} \tag{46}$$

$$b_2 = \frac{c(k(\alpha - a^\star) + \rho_1 + \rho_2)}{6(c+k)} \tag{47}$$

Direct effects: Taking partial derivatives with respect to α holding ρ_2 constant:

$$\frac{\partial b_1}{\partial\alpha} = -\frac{ck}{6(c+k)} < 0 \tag{48}$$

$$\frac{\partial b_2}{\partial\alpha} = +\frac{ck}{6(c+k)} > 0 \tag{49}$$

Sorting effects: Under PAM with uniform distributions, the matching function is $\rho_2(\alpha) =$

$\rho_2^{\min} + \frac{\rho_2^{\max} - \rho_2^{\min}}{\overline{a} - \underline{a}}(\alpha - \underline{a})$. Taking derivatives:

$$\frac{d\rho_2}{d\alpha} = \frac{\rho_2^{\max} - \rho_2^{\min}}{\overline{a} - \underline{a}} \equiv s > 0 \tag{50}$$

Since $\frac{\partial b_1}{\partial \rho_2} = \frac{\partial b_2}{\partial \rho_2} = \frac{c}{6(c+k)}$, the sorting effects are:

$$\frac{\partial b_1}{\partial \rho_2} \cdot \frac{d\rho_2}{d\alpha} = \frac{c}{6(c+k)} \cdot s > 0 \tag{51}$$

$$\frac{\partial b_2}{\partial \rho_2} \cdot \frac{d\rho_2}{d\alpha} = \frac{c}{6(c+k)} \cdot s > 0 \tag{52}$$

Combining direct and sorting effects:

$$\frac{db_1}{d\alpha} = \frac{c}{6(c+k)}[-k + s] \tag{53}$$

$$\frac{db_2}{d\alpha} = \frac{c}{6(c+k)}[k + s] > 0 \tag{54}$$

For b_1: the sign depends on whether $s > k$ (sorting dominates) or $s < k$ (direct dominates). For b_2: both terms are positive, so $\frac{db_2}{d\alpha} > 0$ always. ■

Online Appendix
Misaligned Agents

Microfoundation for linear contracts

This Appendix adapts an argument from Barron, Georgiadis, and Swinkels (2020) to our setting. Suppose that, conditional on a, the agent can implement arbitrary mean-preserving spreads and mean-preserving contractions of each performance measure, independently across dimensions. Formally:

Assumption 1 *After choosing a resource allocation a, the agent can costlessly replace each performance measure P_i by any random variable $\tilde{P}_i$ such that:*

$$\mathbb{E}[\tilde{P}_i|a] = \mathbb{E}[P_i|a], \qquad i = 1, 2.$$

A contract is defined as robust if every feasible joint distribution $(\tilde{P}_1, \tilde{P}_2)$ satisfying the mean constraints in Assumption 1 yields the same expected payment. Now consider a general contract $W(P_1, P_2)$ which can be a nonlinear function of performance measures.

Claim 4 *A contract $W(P_1, P_2) = S + \phi(P_1, P_2)$ is robust to the manipulation described in Assumption 1 if and only if:*

$$\phi(P_1, P_2) = b_0 + b_1 P_1 + b_2 P_2,$$

for some constants b_0, b_1, b_2.

Proof of Claim 4: Given a resource allocation a, define $\mu_i \equiv \mathbb{E}[P_i|a]$ for $i = 1, 2$. Consider an arbitrary measurable contract $\phi : \mathbb{R}^2 \to \mathbb{R}$. The agent's problem at the manipulation stage is:

$$\sup_{(\tilde{P}_1, \tilde{P}_2)} \mathbb{E}\big[\phi(\tilde{P}_1, \tilde{P}_2)|a\big] \quad \text{s.t.} \quad \mathbb{E}[\tilde{P}_i|a] = \mu_i \ \text{ for } i = 1, 2.$$

Fix an arbitrary value $p_2 \in \mathbb{R}$ and define the univariate function of p_1:

$$g_{p_2}(p_1) \equiv \phi(p_1, p_2).$$

Suppose that the function g_{p_2} is not affine in p_1. Then there exist x, y, and $\lambda \in (0, 1)$ such that:

$$g_{p_2}(\lambda x + (1 - \lambda)y) \ \neq \ \lambda g_{p_2}(x) + (1 - \lambda) g_{p_2}(y).$$

There are two cases. First, if:

$$g_{p_2}(\lambda x + (1-\lambda)y) \quad < \quad \lambda g_{p_2}(x) + (1-\lambda) g_{p_2}(y), \tag{55}$$

then let $\tilde{P}_1$ take values x and y with probabilities λ and $1-\lambda$, respectively, and set $P_1 = \lambda x + (1-\lambda)y$ (i.e. P_1 is a constant). Then $\mathbb{E}[\tilde{P}_1|a] = \mathbb{E}[P_1|a]$, and

$$g_{p_2}\big(\mathbb{E}[\tilde{P}_1|a]\big) \quad < \quad \mathbb{E}\big[g_{p_2}(\tilde{P}_1)|a\big], \tag{56}$$

so the agent strictly increases her expected compensation with a mean-preserving spread. Second, if the inequality in equation (55) is reversed, a mean-preserving contraction yields the same conclusion.

Thus robustness requires that g_{p_2} be affine for every p_2, i.e.

$$\phi(p_1, p_2) = \xi(p_2) + \beta(p_2) p_1$$

for some functions ξ and β. By symmetry, fixing p_1 and applying the same argument to $\phi(p_1, \cdot)$, robustness requires:

$$\phi(p_1, p_2) = \gamma(p_1) + \delta(p_1) p_2.$$

for some functions γ and δ. In sum, robustness requires:

$$\phi(p_1, p_2) = A + Bp_1 + Cp_2 + Dp_1 p_2$$

for some constants A, B, C, D.

We now show by contradiction that a robust contract is such that $D = 0$. Suppose $D \neq 0$. Consider random variables $(\tilde{P}_1, \tilde{P}_2)$ taking values (x, y) and (y, x) with equal probability $1/2$, where $x \neq y$, and satisfying

$$\mathbb{E}[\tilde{P}_1|a] = \mathbb{E}[\tilde{P}_2|a] = \tfrac{x+y}{2}.$$

Then

$$\mathbb{E}[\tilde{P}_1 \tilde{P}_2|a] = \tfrac{1}{2}(xy + yx) = xy \neq \left(\tfrac{x+y}{2}\right)^2 = \mathbb{E}[\tilde{P}_1|a]\mathbb{E}[\tilde{P}_2|a].$$

This shows that the agent can change $\mathbb{E}[\phi(\tilde{P}_1, \tilde{P}_2)|a]$ without changing the means of the performance measures, contradicting robustness. Therefore $D = 0$.

If ϕ is linear, then for any admissible $(\tilde{P}_1, \tilde{P}_2)$, we can write:

$$\mathbb{E}[\phi(\tilde{P}_1, \tilde{P}_2)|a] = b_0 + b_1 \mathbb{E}[\tilde{P}_1|a] + b_2 \mathbb{E}[\tilde{P}_2|a],$$

which depends only on the means of the performance measures. Thus, no distributional manipulation can change the agent's expected pay. ■