The Value of Non-Value-Maximizing Managers*

Pierre Chaigneau†
Queen's and ECGI

Alex Edmans‡
LBS, CEPR, and ECGI

Nicolas Sahuguet§
HEC Montréal and CEPR

September 6, 2026

Abstract

Companies pursue multiple goals, such as financial returns and societal impact. Conventional wisdom suggests they should appoint managers who share investors' preferences across these goals. We show that, when performance measurement is asymmetric, appointing a manager who is biased towards the worse-measured activity ("impact") may instead be optimal. Doing so allows for stronger incentives on the better-measured activity ("profits") without distorting resource allocation. Strong profit incentives need not indicate that the manager lacks intrinsic motivation; instead, they counteract his intrinsic motivation for impact. Similarly, strong profit incentives need not imply that a mission-focused firm is greenwashing; it pursues its mission through a biased manager rather than impact incentives. We then study a market equilibrium with heterogeneous firms, characterizing equilibrium matching and the resulting dispersion in managers' rents. The framework offers new insights into executive incentives, corporate purpose, and organizational culture.

Keywords: executive compensation, incentives, sustainability, ESG, private benefits, managerial preferences.

*We thank Nicolas Inostroza and participants at the Canadian Sustainable Finance Network conference and the Society for Organizational and Institutional Economics conference for valuable comments. We gratefully acknowledge the financial support of the Research and Materials Development Fund (RAMD_Edmans_A_26/27_3222) at London Business School.

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

‡Address: Regent's Park, London NW1 4SA, United Kingdom. Email: aedmans@london.edu.

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

1 Introduction

Organizations routinely face decisions that require trading off multiple dimensions of performance. A board may care about both financial returns and environmental and social ("ES") impact, a hospital about treatment and preventive care, and a university about research output and teaching quality. These decisions may involve allocating resources such as money or time, setting the organization's priorities, or determining its strategy.

It might seem optimal to appoint an unbiased leader who shares the organization's overall objective and can be relied upon to make these decisions in pursuit of that objective. However, some companies select CEOs who appear to be biased towards one particular goal. Several executives, such as BlackRock's Larry Fink and former Disney boss Bob Chapek, have been criticized for prioritizing ES impact over financial returns – for example through efforts to combat climate change, promote diversity, equity, and inclusion ("DEI"), or support social justice initiatives. Such actions, and even the appointment of these CEOs, are sometimes dismissed as "woke capitalism." Other leaders, such as Tesla's Elon Musk or Meta's Mark Zuckerberg, are viewed as prioritizing technological innovation and ambitious scientific vision even at the expense of shareholder value.

This paper shows that such practices may be optimal. Central to our result is the observation that, across different dimensions, measured performance may diverge from actual output to varying degrees. For example, shareholder value can be proxied by the stock price in an efficient market, but many aspects of ES impact are qualitative and third-party quantitative ratings are inconsistent (Berg, Kölbel, and Rigobon, 2022). Focusing contractual incentives on the better-measured dimensions may lead the CEO to neglect the worse-measured ones (Holmström and Milgrom, 1991). Hiring managers whose intrinsic preferences are biased towards the latter allows the board to use high-powered incentives on the former, while relying on the manager's bias to prevent excessive neglect of the latter. Thus, appointing agents whose preferences differ from the principal's objective can help achieve her objective: non-value-maximizing managers can be value-maximizing.

We build a parsimonious principal-agent model in which a manager allocates a resource across two activities. For clarity, we will refer to the better-measured activity as "profits" and the worse-measured activity as "impact", although the model admits many other interpretations. Even under this terminology, "impact" encompasses several applications, such as ES impact, scientific innovation, or brand strength, and the model also allows "impact" to be better measured than "profits." The manager privately observes the productivity of each activity. Output of an activity is increasing in both productivity and the resources allocated to it. While output is non-contractible, an imperfect measure of performance is available and contractible for each activity. The principal's overall objective is the total output from both activities, minus the

manager's pay and an adjustment cost of deviating from a default allocation. Like output, this objective is also non-contractible. The manager maximizes his pay minus the adjustment cost of deviating from his default allocation; the latter reflects his preferences and may differ from the principal's default allocation.

Strong incentives on a given dimension induce the manager to respond to his private information about that dimension's productivity, rather than adhere to his default allocation. However, if measured performance diverges from true output, strong incentives also lead the manager to prioritize the former at the expense of the latter (Kerr (1975), Baker (1992)): to allocate resources towards an activity when doing so increases measured performance, even if productivity and thus the contribution to actual output is low. Because profits are more precisely measured, optimal incentive contracts place greater weight on profits and less weight on impact. This asymmetry, however, leads the manager to allocate too many resources to profits from the principal's perspective. Reducing incentives for profits is not a solution, as it would reduce the manager's responsiveness to his private information. Increasing incentives for impact is not a solution either, as it would lead him to improve measured impact rather than actual impact – to "hit the target, miss the point."

We show how this distortion can be mitigated through managerial selection. When impact is poorly measured, the principal optimally appoints a manager who values it more than she does. This bias allows the principal to offer high-powered incentives for profits, where they are most effective, without causing the manager to neglect impact. Note that bias is different from intrinsic motivation. In a single-activity model, it is well-known that an intrinsically motivated manager (e.g. one with a private benefit from effort or output) is valuable because it attenuates the agency problem (Akerlof and Kranton (2000, 2005), Besley and Ghatak (2005), Carlin and Gervais (2009)). In our setting, there are two activities that conflict with each other. Thus, while bias towards impact creates "intrinsic motivation" for impact, it is at the expense of profits. Indeed, in a first-best world where the principal can observe both productivity and resource allocation, bias is unambiguously costly. In addition, unlike in single-activity models where the activity (e.g. effort) is always beneficial to the principal, and so intrinsic motivation is always desirable, here the manager may pursue excessive impact.

The model provides implications for the link between bias and incentives. We find that increasing the manager's bias towards impact not only reduces the optimal incentives for impact, but also increases his incentives for profits. This result rationalizes some otherwise puzzling practices. Many companies state a purpose beyond shareholder value, yet the vast majority of executive incentives are tied to the stock price (Bebchuk and Tallarita, 2023), leading to criticism that purpose statements are "greenwashing". Vladasel et al. (2024) find that many social enterprises simultaneously offer a social mission and monetary rewards for (more easily measured) commercial tasks. Even non-profit organizations use financial incentives: hospitals tie

CEO compensation primarily to financial performance rather than quality (Jenkins, Short, and Ho, 2025), while churches use pay-for-performance incentives for clergy (Hartzell, Parsons, and Yermack, 2010). While this might seem counterintuitive, difficulties in measuring impact mean that the optimal way to pursue it may be to appoint an impact-conscious manager – in which case strong financial incentives are needed to ensure that resources are not excessively allocated to impact at the expense of profitability. Separately, strong financial incentives are often criticized on the grounds that intrinsic motivation should be sufficient to motivate managerial effort: if the CEO requires incentives to deliver exceptional performance, the board has hired the wrong CEO (see the survey of Edmans, Gosling, and Jenter (2023)). Our model shows that profit incentives are fully consistent with an intrinsically motivated CEO. Indeed, the CEO may be intrinsically motivated towards a dimension (impact) that creates value and enhances the principal's objective up to a point. Financial incentives are needed to ensure that he does not pursue impact beyond that point.

Our framework also yields predictions for the types of companies that should appoint biased managers. Bias is more valuable when performance measurement is particularly asymmetric, because financial incentives will be more distortionary. It is also greater when the manager's private information is more important, since financial incentives will be stronger and there is more asymmetry to correct. Bias is less valuable when the manager faces a high adjustment cost from deviating from his default allocation (as he is less responsive to incentives and so incentives are less distortionary), and when the firm faces a high adjustment cost from its default allocation (as a biased manager with a different default allocation is less desirable).

We next 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 biases. In a matching equilibrium, we show that firms with worse measurement of impact hire managers with stronger preferences for impact. Up to a point, managers who are more biased towards impact may earn greater rents, because this bias compensates for the ineffectiveness of impact incentives. Paradoxically, managers who do not share shareholders' objective may have greater potential to achieve this objective, given the endogenous response of financial incentives, and thus earn higher rents. However, when the bias is too high, the manager creates less value as his chosen allocation deviates too much from the principal's. Unlike in the single-firm model, a manager who is more biased towards profits can actually receive stronger incentives for profits. This is because he is matched with a firm that measures impact well and thus offers high impact incentives. The firm therefore offers high profit incentives to reduce the distortionary effect of high impact incentives. This "sorting effect" can outweigh the "direct effect" identified in the single-firm model, whereby a stronger preference for profits reduces profit incentives.

We then consider the case in which managers' preferences are unobservable but can be signaled through costly actions such as prior managerial decisions, charitable donations, and

volunteer activities. When the sorting effect dominates, signals are more dispersed than the actual preferences of managers: they exaggerate their difference from a pivot type who signals truthfully. When the direct effect dominates, signals are more concentrated around the pivot type: managers understate their differences. This provides a new theory of costly signaling of managerial preferences, including both "virtue signaling" and signaling of commercial discipline.

The manager's preference for a particular activity may stem from a number of sources. One is personal preferences, which may be shaped by his background and upbringing. A second is organizational factors such as corporate culture or social norms. For example, if a firm has issued a purpose statement committing to impact, has a historical reputation for delivering it, or employs colleagues who value it, the manager may face social or reputational pressure to deliver impact. Interestingly, our model predicts that improved measurement of impact (e.g. through superior ES reporting standards) may optimally shift corporate culture away from impact, as impact can instead be induced through financial incentives. Conversely, improvements in financial reporting, and thus the effectiveness of profit incentives, may lead boards to place greater emphasis on impact in public statements, to encourage managers to deliver it. The model thus rationalizes apparent overcommitment to impact (e.g. through mission statements), even if impact is poorly measured and may be detrimental to shareholder value. Such a commitment may reflect an optimal response to the superior measurement of profits.

This paper principally contributes to two literatures: one on performance measurement and a second on intrinsically motivated agents. Starting with the former, Holmström and Milgrom (1991) show that when some dimensions of performance are easier to measure than others, incentive pay distorts the allocation of effort toward the better-measured dimensions. They thus predict low-powered incentives across all tasks when tasks are substitutes; in our model, managerial bias towards one dimension enables high-powered incentives on the other dimension. Holmström and Milgrom (1991) instead propose job design as a solution: assigning easy-to-measure tasks to one agent, who is given strong incentives for all tasks, and difficult-to-measure tasks to another agent alongside weak incentives. Our model features a single resource-allocation decision across multiple activities, so job design is not a solution; it is thus particularly applicable to managers who are responsible for several functions. Baker (1992) considers a single performance measure and finds that optimal incentives are weaker when this measure is less closely aligned with the principal's objective.

Moving to the literature on intrinsically motivated agents, Akerlof and Kranton (2000, 2005) and Fischer and Huddart (2008) show that an agent who identifies with the principal's goals or follows norms that promote high effort does not need strong financial incentives. Besley and Ghatak (2005, 2006, 2018) show that, in a market equilibrium, firms hire agents whose preferences accord with their mission, allowing the agency problem to be mitigated at lower cost. We instead show that asymmetric measurement can make congruence suboptimal: the

principal may prefer a manager who is biased toward the worse-measured activity.

Our paper is closest to work showing that some bias can be valuable. Prendergast (2007) shows that appointing biased agents (those who place greater weight on a particular dimension than the principal) can increase effort in a model without financial incentives, so firms should choose agents whose biases are suited to the task at hand. Prendergast (2008) introduces financial incentives on a single aggregate performance measure, with symmetric mismeasurement across dimensions, and finds that incentives are lower for more biased (intrinsically motivated) agents. The firm hires a different agent to focus on each task, with a bias suited to that task. We consider a single resource allocation decision across two activities that are separately measured and incentivized, with measurement quality differing across dimensions. This asymmetry leads the principal to optimally appoint a manager biased toward the worse-measured dimension. Bias therefore plays an additional role: whereas in Prendergast (2008) it substitutes for monetary incentives weakened by imperfect performance measurement, here bias toward one dimension also enables stronger incentives on the other by counteracting the resource misallocation induced by asymmetric performance measurement.

Bénabou and Tirole (2016) study a setting in which one task is contractible and the other is not, and must therefore be induced through intrinsic motivation. When intrinsic motivation differs across agents and is privately observed, they show that agents with greater motivation for "impact" than other agents select weakly lower incentives for "profits" (using our terminology). Because effort across tasks is substitutable, stronger profit incentives draw effort away from the intrinsically valued impact task; more impact-motivated agents therefore value such incentives less. In contrast, we show that impact-motivated managers receive stronger profit incentives, because their impact motivation mitigates distortions induced by profit incentives.

The idea that a leader whose preferences differ from the principal's can create value also appears in the literature on managerial vision. Rotemberg and Saloner (2000) show that a "visionary" CEO biased toward a particular strategy encourages employees to invest in that strategy. In Van den Steen (2005), a manager with strong beliefs attracts like-minded employees, improving motivation and coordination. In these papers, bias creates value through its effect on employees and is directed toward the activity that requires encouragement. In our model, bias creates value through improving incentive design and is directed toward the worse-measured activity.

Finally, our paper is related to research on the consequences of managers focusing on narrow financial metrics. Aghion and Stein (2008) and Ben-David and Chinco (2026) study settings in which managerial decisions are shaped by metrics such as profit margins or short-term EPS, even though they are imperfect proxies for firm value. We show that tying incentive pay to these proxies may be efficient, despite their incompleteness, if managers have intrinsic preferences towards other performance dimensions.

2 Model

We consider a principal-agent model of resource allocation. A risk neutral principal hires a risk neutral agent (manager) who receives private information about the productivity of potential allocations of the firm's resources.[1] The principal may be a board acting on behalf of shareholders, an investor, or the firm itself; for ease of exposition, we often refer to the principal as "the firm." The firm produces unverifiable output V_i on two dimensions $i = 1, 2$. On each dimension, there is a contractible performance measure P_i. Output and performance are affected by a non-contractible allocation decision, a. This decision can reflect the allocation of time, money or other resources, or the choice of strategy. As in Baker (1992), output and performance are also affected by shocks $\boldsymbol{\epsilon} \equiv \{\epsilon_1^v, \epsilon_2^v, \epsilon_1^p, \epsilon_2^p\}$ which are privately observed by the manager:

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

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

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 normalized to $\mathbb{E}\left[\epsilon_i^j\right] = 1$. The correlation coefficient between output and measured performance on dimension i is: $\rho_i \equiv \frac{cov(\epsilon_i^v, \epsilon_i^p)}{\sigma_i^v \sigma_i^p} \geq 0$. Throughout the main analysis, we assume $\sigma_i^v > 0$ and $\sigma_i^p > 0$. Shocks to performance and output are independent across dimensions, so that:[2]

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

The manager makes a single resource allocation decision, a, which affects both outputs and performance measures.[3] This differs from the multitasking framework, in which a manager chooses separate effort levels for each task. While both frameworks allow activities to be substitutes, they differ in several ways. First, in a multitasking model, individual tasks can in principle be allocated to different agents (Holmström and Milgrom (1991); Prendergast (2007, 2008)). By contrast, the resource allocation model features a single decision involving a trade-off across multiple dimensions, making it particularly applicable to managers responsible for several activities. Second, multitasking models typically assume that the principal knows which actions are desirable (higher effort on each task) and intrinsic motivation helps induce them. Here, the principal does not know which allocation is optimal, as it depends on the agent's private infor-

[1] Private information may be interpreted literally; alternatively, it may refer to public information that the manager assesses using his unique expertise. For example, an experienced CEO may observe industry trends and use them to determine the optimal scale of investment.

[2] This does not imply that performance will be independent across dimensions: increasing a improves V_1 at the expense of V_2. Instead, we are only assuming that shocks to performance are independent across dimensions.

[3] For tractability, we do not restrict $a \in [0, 1]$. Note that $a < 0$ could be interpreted as the firm not spending any resources on impact, but instead having a harmful impact to increase profitability, whereas $a > 1$ could be interpreted as the firm raising financing to spend more than its existing resources on impact.

mation. As a result, rather than helping to implement the principal's preferred actions, bias can distort the agent's decision rule. The desirability of bias is thus a priori unclear. Third, with a single action, preferences over allocations can be summarized by a single parameter, making it straightforward to compare the principal's and manager's preferred allocations.

The firm's objective function is:

$$\mathbb{E}\left[V_1(a,\boldsymbol{\epsilon}) + V_2(a,\boldsymbol{\epsilon}) - W(a,\boldsymbol{\epsilon}) - \frac{k}{2}\left(a - \alpha_P\right)^2\right] \tag{3}$$

where $W(a,\boldsymbol{\epsilon})$ is the manager's payment, and $k > 0$. The last term captures the cost to the principal of deviating from a baseline resource allocation α_P, which may reflect adjustment costs (e.g., if α_P represents a historical or approved allocation), or board or investor preferences.

Because V_1 and V_2 are noncontractible, so is the firm's objective. One application is where the objective function is long-term shareholder value, with V_1 and V_2 capturing distinct contributors to it and P_1 and P_2 being interim proxies. For example, the dimensions may be innovation and profitability, with the proxies being patent counts and accounting earnings. Long-term value may not be fully reflected in contractible performance measures (such as the stock price) for many years – for example, the market may be unable to value innovation in the short term. Moreover, it may be infeasible or inefficient to defer all managerial compensation until that value is fully reflected in the stock price, because the manager values interim consumption.[4] The assumption that long-term shareholder value cannot be fully contracted upon is standard in models of managerial myopia, where the manager is instead evaluated using short-term performance measures (Narayanan, 1985; Stein, 1988, 1989; Bebchuk and Stole, 1993). In reality, executives are given equity with vesting periods of 3-5 years (Gopalan et al., 2014; Edmans, Fang, and Lewellen, 2017), so compensation is not conditioned entirely on the eventual realization of long-term shareholder value. Online Appendix A shows that the optimal incentives and bias are unchanged if there is a contractible but imperfect interim proxy of the principal's objective, such as the short-term stock price. If the principal's objective were contractible, the agency problem would be simple to address as the principal could contract directly on it. A second application is where V_1 and V_2 capture profits and impact, both of which shareholders value, as in models of responsible investing. Finally, the firm could be an organization with non-financial objectives, such as a hospital that values prevention and cure, or a university that values research and teaching.

As in Baker (1992), the firm offers a linear incentive contract $\{S, b_1, b_2\}$ such that the man-

[4]Dynamic agency models formalize this limit to deferral. DeMarzo and Fishman (2007) and DeMarzo and Sannikov (2006) assume that the manager has a higher discount rate than investors, making prolonged deferral of compensation costly.

ager's payment is equal to:

$$W(a, \boldsymbol{\epsilon}) = S + b_1 P_1(a, \boldsymbol{\epsilon}) + b_2 P_2(a, \boldsymbol{\epsilon}) \tag{4}$$

Linear contracts are a standard assumption in the literatures on performance measurement (Baker (1992)) and intrinsic motivation (Fischer and Huddart (2008), Prendergast (2008), Bénabou and Tirole (2016)). They allow incentives for a task to be captured by a single parameter, b_i, allowing for clear results on the link between incentives and bias (the focus of this paper), as well as straightforward comparative static analysis. The Online Appendix provides one potential microfoundation: if the manager can induce mean-preserving spreads or contractions of the performance measure distributions, then he can game any nonlinear contract. Only linear contracts are robust to such manipulation, as shown by Barron, Georgiadis, and Swinkels (2020).

The manager's reservation utility is $\overline{U}$. Given a contract $\{S, b_1, b_2\}$ and resource allocation rule $a(\boldsymbol{\epsilon})$, his expected utility is:

$$\mathbb{E}\left[S + b_1 P_1(a, \boldsymbol{\epsilon}) + b_2 P_2(a, \boldsymbol{\epsilon}) - \frac{c}{2}\left(a - \alpha_M\right)^2\right], \tag{5}$$

where $c > 0$ parameterizes the cost to the manager of deviating from his preferred resource allocation α_M. The manager has preferences over the resource allocation decision, similar to models of motivated agents (e.g., Akerlof and Kranton (2005); Besley and Ghatak (2005)). Empirically, there is a large body of evidence that agents have nonpecuniary preferences over actions, and that such preferences are heterogeneous across individuals (e.g., Cassar and Meier (2018), Fehr and Charness (2025)). There is a continuum of different managers that the firm can hire, each of whom has a different α_M, where $\alpha_M \in [\underline{\alpha_M}, \overline{\alpha_M}]$. If the firm and manager's baseline resource allocations are the same ($\alpha_M = \alpha_P$), we say that the manager is "congruent" or "unbiased". We assume that $\alpha_P \in [\underline{\alpha_M}, \overline{\alpha_M}]$, so that a congruent manager is available. We use "bias" to refer to the difference between α_M and α_P, e.g., if $\alpha_M > \alpha_P$, the manager is biased towards dimension 1. In the baseline model, managerial preferences α_M are observable to the firm, as in standard matching models (e.g. Gabaix and Landier (2008), Terviö (2008)). Section 5 considers the case in which preferences are unobservable and managers can signal them through costly actions.

The firm chooses the contract $\{S, b_1, b_2\}$ to maximize its objective function in equation (3), subject to the manager's incentive constraint (he chooses a to maximize (5)) and participation constraint (his expected utility is at least $\overline{U}$). The sequence of events is as follows. First, the firm chooses a manager with preferences α_M and offers him a contract $\{S, b_1, b_2\}$. Second, the manager accepts or rejects the contract. Third, if he accepts the contract, he privately observes the shocks $\boldsymbol{\epsilon}$, chooses the allocation a, and is paid according to the realized performance measures

P_1 and P_2.

3 Analysis

This section solves for both the type of manager that the firm hires and the compensation contract that it offers. We begin by characterizing the first-best, which arises when the resource allocation a is contractible and there is no asymmetric information (i.e. the principal observes the state $\boldsymbol{\epsilon}$). Then, the principal can dictate a state-contingent allocation $a(\boldsymbol{\epsilon})$ and pays a fixed wage w. She chooses the resource allocation $a(\boldsymbol{\epsilon})$ and the manager's type α_M to maximize her expected payoff subject to the manager's participation constraint:

$$\max_{a(\boldsymbol{\epsilon}),\alpha_M} \mathbb{E}\left[a\,\epsilon_1^v + (1-a)\epsilon_2^v - w - \frac{k}{2}(a-\alpha_P)^2\right] \tag{6}$$

$$\text{s.t.} \quad \mathbb{E}\left[w - \frac{c}{2}(a-\alpha_M)^2\right] \geq \overline{U}. \tag{7}$$

The participation constraint will bind at the optimum. Substituting for w from the binding participation constraint into the objective function, and optimizing pointwise over a for each realization of $\boldsymbol{\epsilon}$, yields the first-best state-contingent allocation:

$$a^{FB}(\boldsymbol{\epsilon}) = \frac{c\alpha_M + k\alpha_P + \epsilon_1^v - \epsilon_2^v}{c+k}. \tag{8}$$

Substituting back into (6) and optimizing over α_M shows that the principal hires a congruent manager, i.e. $\alpha_M = \alpha_P$. Intuitively, this minimizes the manager's expected disutility from deviating from his preferred resource allocation, which must be compensated to satisfy his participation constraint; there is no benefit to hiring a biased manager as the allocation a can be dictated anyway. Then, the first-best allocation is:

$$a^{FB}(\boldsymbol{\epsilon}) = \alpha_P + \frac{\epsilon_1^v - \epsilon_2^v}{c+k}. \tag{9}$$

The allocation depends on the output shocks ϵ_1^v and ϵ_2^v, not the performance shocks ϵ_1^p and ϵ_2^p. The wedge between output and measured performance, which is the main friction in the second-best, is irrelevant in the first-best when the principal observes $\boldsymbol{\epsilon}$ and directly controls a.

We now study the second-best outcome when the resource allocation decision is non-contractible. We solve the model backwards, starting with the manager's resource allocation decision, then deriving the optimal contract, and finally solving for the manager's bias α_M that the firm prefers. We start by deriving several preliminary results.

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

$$a = \alpha_M + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}.$$

The optimally chosen resource allocation is additive in two terms. The first is the manager's baseline resource allocation α_M, which he would choose in the absence of financial incentives. The second term captures the deviation from this baseline level due to his information and incentives, scaled by the deviation cost c. Stronger incentives on dimension 1 (b_1) have two effects. First, they increase the manager's expected allocation to dimension 1. Second, they make his allocation more responsive to his private information: the increase in a is greater when ϵ_1^p is higher. Incentives are thus valuable because they increase information responsiveness, but distortionary because they shift the expected allocation. Importantly, this responsiveness is to the effect of the allocation on measured performance ϵ_1^p rather than output ϵ_1^v. The distinction between ϵ_1^p and ϵ_1^v means that strong incentives for a poorly-measured dimension can be distortionary: the manager may increase a if ϵ_1^p is high, even if ϵ_1^v is low. For example, a bank manager may increase the number of loans even if the quality of such loans is poor, a hospital may conduct more tests even if they do not improve patient health, and a professor may design a course to improve teaching evaluations at the expense of learning quality. The effects are symmetric for dimension 2: stronger incentives b_2 reduce the expected allocation a and make it more responsive to ϵ_2^p.

Lemma 2 determines the optimal incentives for both dimensions of performance. We use "incentive gap" to refer to the difference between incentives on dimensions 2 and 1, $b_2 - b_1$. The fixed wage S is set to keep the manager at his reservation utility given the incentives derived in Lemma 2.

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

$$b_1 = \frac{c\left(\rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2}+1\right) + \rho_2\sigma_2^p\sigma_2^v + \sigma_2^{p2}k\left(\alpha_P - \alpha_M\right)\right)}{(c+k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)} \tag{10}$$

$$b_2 = \frac{c\left(\rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2}+1\right) + \rho_1\sigma_1^p\sigma_1^v + \sigma_1^{p2}k\left(\alpha_M - \alpha_P\right)\right)}{(c+k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)}. \tag{11}$$

The incentive gap is given by:

$$b_2 - b_1 = \frac{c\left[\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} - \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2} + k(\sigma_1^{p2} + \sigma_2^{p2})\left(\alpha_M - \alpha_P\right)\right]}{(c+k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)}. \tag{12}$$

Lemma 2 shows that incentives for dimension i are shaped by three forces. First, holding

α_M fixed, incentives b_i are increasing in $\rho_i\sigma_i^v$, similar to Baker (1992).[5] They are higher for a dimension that is well-measured (high ρ_i), and has greater uncertainty σ_i^v about the production technology, as the latter makes it particularly valuable to make the manager responsive to his private information.

Second, holding α_M fixed, incentives b_i are increasing in $\rho_j\sigma_j^v$, as this increases incentives for dimension j; incentives for dimension i therefore rise to avoid distortions. This is a standard effect when tasks are substitutes so that the balance of incentives matters. Thus, improvements in performance measurement on one dimension increase incentives for both dimensions – a spillover effect.

Third, incentives depend on the manager's bias, $\alpha_M - \alpha_P$. When the manager is biased towards dimension 1 ($\alpha_M > \alpha_P$), incentives decrease on dimension 1 and increase on dimension 2; the reverse is true if $\alpha_M < \alpha_P$. Intuitively, the manager's intrinsic preferences for a dimension substitute for explicit incentives.

Proposition 1 delivers the main result of the paper: the firm's optimal choice of the manager's preferences α_M.[6]

Proposition 1 *The firm appoints a manager with the following preferred resource allocation:*

$$\alpha_M = \begin{cases} \overline{\alpha_M} & \text{if } \alpha_M^\star > \overline{\alpha_M} \\ \alpha_M^\star & \text{if } \alpha_M^\star \in [\underline{\alpha_M}, \overline{\alpha_M}] \\ \underline{\alpha_M} & \text{if } \alpha_M^\star < \underline{\alpha_M} \end{cases} \quad \text{where} \quad \alpha_M^\star = \alpha_P + \frac{\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} - \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}}{c\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right) + k\sigma_1^{p2}\sigma_2^{p2}}.$$

For interior solutions ($\alpha_M = \alpha_M^\star$), the manager's preferred resource allocation α_M is increasing in ρ_2 and decreasing in ρ_1. If $\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} > \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}$ (i.e. $\alpha_M^\star > \alpha_P$), then α_M is decreasing in c and k.

For $\alpha_P \in (\underline{\alpha_M}, \overline{\alpha_M})$, the firm appoints a manager with $\alpha_M > \alpha_P$ if and only if:

$$\rho_2\frac{\sigma_2^v}{\sigma_2^p} > \rho_1\frac{\sigma_1^v}{\sigma_1^p}. \tag{13}$$

The type of manager hired depends on the firm's performance measurement system. The firm appoints a manager biased towards dimension 1 ($\alpha_M > \alpha_P$) if performance on dimension 2 is sufficiently more informative about output than performance on dimension 1, as captured by condition (13). A higher correlation coefficient ρ_i means that measured performance more closely tracks output, higher output volatility σ_i^v increases the value of the manager's private

[5] While the first term contains $\rho_i\sigma_i^p\sigma_i^v$, performance variability σ_i^p enters elsewhere in b_i, so its overall effect is ambiguous.

[6] In section 4, where we take into account the scarcity of any given managerial type, the firm does not choose α_M, so corner solutions do not arise.

information, while higher performance volatility σ_i^p reduces the informativeness of measured performance about output. (We henceforth refer to the full ratio $\rho_i \frac{\sigma_i^v}{\sigma_i^p}$ as the use "scaled quality" of performance measurement on dimension i.) Hiring a manager biased towards dimension 1 allows the firm to provide strong incentives on dimension 2, to make the manager more responsive to his private information, while relying on the manager's bias to mitigate excessive neglect of dimension 1. The optimal bias will be greater when output volatility on both dimensions is scaled up by the same amount, i.e., when the manager's private information matters more. The optimal bias is decreasing in c and k. When the manager's deviation cost c is higher, then he is less responsive to incentives, reducing the need for bias to correct the distortion caused by asymmetric measurement and thus asymmetric incentives. When the firm's deviation cost k is higher, it wants a less biased manager because bias shifts him away from the firm's preferred allocation.

In contrast, if $\rho_2 \frac{\sigma_2^v}{\sigma_2^p} = \rho_1 \frac{\sigma_1^v}{\sigma_1^p}$, then the firm optimally hires a congruent manager ($\alpha_M = \alpha_P$). Thus, it is asymmetry in the scaled quality of performance measurement, rather than imperfect measurement per se, that makes bias optimal. Performance measurement can be imperfect on both dimensions, but congruence remains optimal if the two dimensions are equally well measured.

One application of Proposition 1 is to the case in which the first dimension is impact and the second is profitability. Assume that profitability is well-measured due to accounting standards, and that impact is poorly measured because it is difficult to quantify and the relevant measures of impact are inconsistent across firms. If (13) is satisfied, then the firm optimally hires a manager biased toward impact, i.e. $\alpha_M > \alpha_P$.

A second application is to product design, where the two dimensions are high quality and low cost. If cost is measured more accurately than quality, then $\rho_2 \frac{\sigma_2^v}{\sigma_2^p}$ is high and $\rho_1 \frac{\sigma_1^v}{\sigma_1^p}$ is low. The firm then optimally hires a product manager biased towards quality.

Plugging the optimal managerial choice from Proposition 1 into the incentives of Lemma 2 gives the incentives as a function of exogenous parameters. They are given by Corollary 1.

Corollary 1 *Suppose that $\alpha_M^\star$ lies in $\left(\underline{\alpha_M}, \overline{\alpha_M}\right)$. When the firm appoints this optimal manager, incentives are:*

$$b_1 = c\frac{\sigma_2^{p2}\rho_1\sigma_1^p\sigma_1^v + \frac{c}{c+k}\left(\rho_1\sigma_1^p\sigma_1^v + \rho_2\sigma_2^p\sigma_2^v\right)}{c\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right) + k\sigma_1^{p2}\sigma_2^{p2}} \tag{14}$$

$$b_2 = c\frac{\sigma_1^{p2}\rho_2\sigma_2^p\sigma_2^v + \frac{c}{c+k}\left(\rho_1\sigma_1^p\sigma_1^v + \rho_2\sigma_2^p\sigma_2^v\right)}{c\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right) + k\sigma_1^{p2}\sigma_2^{p2}} \tag{15}$$

and the incentive gap is:

$$b_2 - b_1 = \frac{c\left(\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} - \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}\right)}{c\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right) + k\sigma_1^{p2}\sigma_2^{p2}} = c\left(\alpha_M^\star - \alpha_P\right). \tag{16}$$

Incentives b_i are increasing in measurement quality ρ_1 and ρ_2, output volatility σ_1^v and σ_2^v, the manager's deviation cost c, and decreasing in the firm's deviation cost k.

Corollary 1 provides comparative statics when the preferences of the manager α_M^* are optimally chosen, rather than holding α_M fixed as in Lemma 2. As in Lemma 2, incentives on each dimension increase with both measurement qualities ρ_1 and ρ_2 and both output volatilities σ_1^v and σ_2^v. They also increase with the manager's deviation cost c, because stronger incentives are required to induce a less responsive manager to use his private information. By contrast, they decrease with the firm's deviation cost k: when departures from the firm's preferred allocation are more costly, inducing responsiveness to private information becomes less valuable.

Equation (16) links managerial selection and incentive design. When dimension 2 is sufficiently better measured that (13) holds, the firm appoints a manager biased towards dimension 1 and offers stronger incentives on dimension 2. Comparing equation (16) with the incentive gap under a congruent manager (by setting $\alpha_M = \alpha_P$ in equation (12)) shows that appointing the optimal biased manager allows the firm to use a larger incentive gap. This gap shifts the manager's expected allocation towards dimension 2, while his bias shifts it towards dimension 1. For an interior solution, Lemma 1 shows that these two effects exactly offset in expectation, so that $\mathbb{E}[a] = \alpha_P$. Thus, managerial selection and incentive design are complements: by appointing a manager biased toward the worse-measured dimension, the firm can tailor incentives more strongly to differences in measurement quality while preserving its preferred allocation on average.

Proposition 2 studies the benefit to the principal from hiring the optimal manager rather than a congruent manager, i.e. the value of a non-value-maximizing manager. This is of particular interest to a firm with an incumbent congruent manager that faces a switching cost of replacing him with the optimal manager. We define $\Delta \equiv \alpha_M - \alpha_P$ as the manager's bias for dimension 1 and Δ^* as the optimal bias, i.e. Δ when α_M is as in Proposition 1.

Proposition 2 *Suppose the unconstrained optimum α_M^* in Proposition 1 lies in $(\underline{\alpha_M}, \overline{\alpha_M})$. The principal's gain from hiring the optimal manager rather than a congruent manager is given by:*

$$\Pi^* - \Pi_0 = \frac{k(\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} - \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2})^2}{2(c+k)(cD + k\sigma_1^{p2}\sigma_2^{p2})D} \geq 0. \tag{17}$$

where $D \equiv \sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}$. If $\rho_2\frac{\sigma_2^v}{\sigma_2^p} > \rho_1\frac{\sigma_1^v}{\sigma_1^p}$, this gain is increasing in ρ_2 and σ_2^v and decreasing

in ρ_1, σ_1^v, and c. The gain can also be expressed as:

$$\Pi^* - \Pi_0 = \frac{(c+k)(cD + k\sigma_1^{p2}\sigma_2^{p2})D}{2c^2 k\sigma_1^{p4}}(b_2(\Delta^*) - b_2(0))^2. \tag{18}$$

The gain is zero if and only if $\rho_1 \frac{\sigma_1^v}{\sigma_1^p} = \rho_2 \frac{\sigma_2^v}{\sigma_2^p}$, as then $\alpha_M = \alpha_P$ from Proposition 1: congruence is optimal. If $\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} > \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}$, then the firm optimally hires a manager biased towards dimension 1, and the gain is directly proportional to $(b_2(\Delta^*) - b_2(0))^2$, the square of the increase in incentives for dimension 2 from hiring the optimal manager rather than a congruent one. Equation (18) highlights the key mechanism of the paper: bias creates value by allowing the principal to use more asymmetric incentives.

The comparative statics are as follows. Without loss of generality, we consider the case in which dimension 2 is sufficiently better measured than dimension 1, i.e. $\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} > \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}$. The gain is increasing in ρ_2 and σ_2^v. When dimension 2 is better measured, or the manager's private information on dimension 2 is more valuable, the firm wishes to give stronger incentives for dimension 2. This increases the value of the manager's bias towards dimension 1 as a counterweight. In contrast, the gain is decreasing in ρ_1 and σ_1^v. When dimension 1 is better measured, or private information on dimension 1 is more valuable, the firm provides stronger incentives for this dimension. Since incentives already direct resources towards dimension 1, preferences for dimension 1 are less valuable. The gain is decreasing in the deviation cost c. A higher cost makes the manager less responsive to financial incentives. This reduces the distortion caused by asymmetric incentives and thus the need for correction through bias.

Our results imply that firms benefit from a broad managerial pipeline (high $\overline{\alpha_M} - \underline{\alpha_M}$) rather than only candidates with $\alpha_M = \alpha_P$. Performance measurement technologies and the value of private information may change over time, making different manager types optimal in different circumstances. This is relevant for a company's succession planning: it is valuable to develop a diversity of potential successors rather than only those in the same mould as the CEO. It is also relevant for business schools that train the next generation of managers: there is value to developing executives with a range of preferences, including preferences for impact rather than only shareholder value.

In summary, the precision of performance measurement and the value of the manager's private information on each dimension determine which manager is hired and the sensitivity of pay to performance. When performance is a good measure of output on dimension i, or private information on dimension i is valuable, the manager should be highly responsive to his signal, ϵ_i^p, and so incentives b_i should be high. To avoid excessive resource allocation to dimension i at the expense of dimension j, the principal optimally appoints a manager biased towards the latter.

4 Matching in Competitive Markets

We now extend the model to a continuum of firms that differ in their performance measurement, while retaining a continuum of managers with heterogeneous preferences. This analysis allows us to study the matching between firms and managers, as well as which managers receive the highest rents (in the single-firm model, the manager receives his reservation utility). We also show that, unlike the single-firm model, managers biased towards one dimension need not receive weaker incentives for that dimension, because they may be matched with a firm that measures the other dimension more precisely, which raises incentives on both dimensions through the spillover effect.

4.1 Setup

There is a continuum of managers who differ only in their preferred resource allocations α_M: the type of manager i, α_M^i, is distributed on $[\underline{\alpha}_M, \overline{\alpha}_M]$ with CDF F_{α_M} and PDF f_{α_M}. There is a continuum of firms that differ only in the quality of their measurement of dimension 1, ρ_1: the type of firm i, ρ_1^i, is distributed on $[\underline{\rho}, \bar{\rho}]$ with CDF F_ρ and PDF f_ρ. We assume that f_ρ and f_{α_M} are continuous and bounded above and below away from zero on their supports.

The quality of measurement of dimension 2, ρ_2, is common across firms. This allows firms to be indexed by a single parameter, ρ_1^i. One application is where dimension 2 is profit, which is measured relatively uniformly due to accounting standards, and dimension 1 is impact, the measurement of which varies substantially depending on the nature of impact activities. Importantly, we do not make any assumptions on ρ_2 compared to $\underline{\rho}$ and $\bar{\rho}$, i.e. we do not assume that dimension 2 is better or worse measured than dimension 1. Where there is no ambiguity, we will use "measurement" or "measurement quality" to refer to the measurement of dimension 1, ρ_1^i, since ρ_2 is common to all firms, and "preference" for the manager's preference for dimension 1. Since volatilities are common across firms, for tractability we additionally assume that all output and performance volatilities are equal, and normalize this common volatility to 1. Finally, to ensure that b_1 and b_2 are nonnegative for any firm-manager pair, we assume that $k(\overline{\alpha_M} - \underline{\alpha_M}) \leq \underline{\rho} + \rho_2$.

We consider a matching model with equal masses of firms and managers and transferable utility. Let $Y(\rho_1, \rho_2; \alpha_M)$ denote the total surplus from a firm-manager match – the firm's objective function plus the manager's expected utility. It is given by:

$$Y(\rho_1, \rho_2; \alpha_M) = \mathbb{E}\left[V_1 + V_2 - \frac{k}{2}(a - \alpha_P)^2 - \frac{c}{2}(a - \alpha_M)^2\right] \tag{19}$$

when the firm optimally chooses (b_1, b_2) and the manager optimally chooses a.

The firm maximizes its own payoff by maximizing joint surplus Y and then setting S to give

the manager his competitive equilibrium wage. As the fixed wage S is unrestricted, the model is a matching model with fully transferable utility. The equilibrium of the model is such that: (i) all firm-manager matches are stable, i.e., no firm and manager who are not matched with each other can both improve their payoffs by matching; (ii) all firms offer the contract (b_1, b_2) that maximizes the joint surplus $Y(\rho_1, \rho_2; \alpha_M)$, with the fixed wage S set to give the manager his competitive equilibrium payoff; (iii) all managers choose the optimal resource allocation given their contracts.

Stable matching determines rent differences but leaves the division of aggregate surplus between firms and managers indeterminate. We select the equilibrium in which the least-valued manager's participation constraint binds:

$$\min_{\alpha_M} U(\alpha_M) = 0.$$

This selection arises, for example, as the limit of a market with an arbitrarily small excess supply of managers.

4.2 Matching equilibrium

The matching assignment depends on the cross-derivative of the surplus with respect to the two matching variables (Becker (1973), Legros and Newman (2007)), which is given in Lemma 3.

Lemma 3 *The surplus of a match $Y(\rho_1, \rho_2, \alpha_M)$ is submodular in ρ_1 and α_M:*

$$\frac{\partial^2 Y(\rho_1, \rho_2, \alpha_M)}{\partial \rho_1 \partial \alpha_M} < 0. \tag{20}$$

The negative cross-derivative means that a manager with stronger preferences for dimension 1 (higher α_M) generates more surplus at a firm with weaker measurement of dimension 1 (lower ρ_1). Intuitively, strong preferences and weak measurement are complements, because the former compensate for the latter. Conversely, a firm with good measurement of dimension 1 does not need a manager biased towards it, because it can instead rely on financial incentives.

Submodularity in a matching model with transferable utility leads to negative assortative matching, as illustrated in Proposition 3.

Proposition 3 *The matching equilibrium involves negative assortative matching: firms with high ρ_1^i are matched with managers with low α_M^i. The matching function $\rho_1(\alpha_M)$, which gives the type ρ_1^i of the firm matched with a manager with type α_M^i, is strictly decreasing and given by:*

$$\rho_1(\alpha_M^i) = F_\rho^{-1}(1 - F_{\alpha_M}(\alpha_M^i)). \tag{21}$$

Firms with the best measurement of dimension 1 (highest ρ_1) hire the managers with weakest preferences for that dimension (lowest α_M), because of the complementarity shown in Lemma 3.

4.3 Equilibrium rents

We now characterize the rents earned by different managers in the matching equilibrium. The rent $U(\alpha_M)$ of a manager with type α_M is defined as:

$$U(\alpha_M) \equiv \mathbb{E}\Big[W(a,\epsilon) - \frac{c}{2}(a-\alpha_M)^2\Big|\alpha_M\Big] - \overline{U} \tag{22}$$

where a is optimally chosen.

Proposition 4 *The marginal rent of a manager with type α_M is:*

$$U'(\alpha_M) = \frac{k}{3(c+k)}\left(\rho_2 - \rho_1(\alpha_M) - (3c+k)(\alpha_M - \alpha_P)\right), \tag{23}$$

and the rent schedule is:

$$U(\alpha_M) = U(\underline{\alpha_M}) + \int_{\underline{\alpha_M}}^{\alpha_M}\left\{\frac{k(\rho_2-\rho_1(x))}{3(c+k)} - \frac{k(3c+k)}{3(c+k)}(x-\alpha_P)\right\}dx. \tag{24}$$

With equal masses and transferable utility, the conditions for stable matching determine each side's payoff only up to a common additive constant. We normalize payoffs by setting the rent of the least valued type to zero. In the integral above this pins down $U(\underline{\alpha_M})$. The marginal rent is higher if the firm has worse measurement for dimension 1 (lower ρ_1), which increases the value of managers with stronger preferences for dimension 1. It is also higher if dimension 2 is better measured (higher ρ_2): a manager with stronger preferences for dimension 1 is more valuable when higher incentives can be provided on dimension 2. Ceteris paribus, the marginal rent is lower if the manager is more biased, because of the higher deadweight loss from misallocation.

The rent function $U(\alpha_M)$ is generally non-monotonic: agents with extreme biases (very high or very low α_M) may earn lower rents than agents with intermediate biases, depending on the distribution of firm types and model parameters. We can derive a closed-form characterization when firms' measurement ρ_1 and managers' preferences α_M are uniformly distributed.

Corollary 2 *Assume $\rho_1 \sim U[\underline{\rho},\overline{\rho}]$, $\alpha_M \sim U[\underline{\alpha_M},\overline{\alpha_M}]$, $\alpha_P \in (\underline{\alpha_M},\overline{\alpha_M})$, $(3c+k)(\alpha_P - \underline{\alpha_M}) > \overline{\rho} - \rho_2$, $(3c+k)(\overline{\alpha_M} - \alpha_P) > \rho_2 - \underline{\rho}$, and that the slope of the matching function satisfies:*

$$m \equiv \frac{\overline{\rho} - \underline{\rho}}{\overline{\alpha_M} - \underline{\alpha_M}} < 3c + k$$

(i) The rent function $U(\alpha_M)$ is single-peaked. The highest-earning manager has type $\widehat{\alpha}_M$, the unique solution to:

$$\rho_2 - \rho_1(\widehat{\alpha}_M) = (3c + k)(\widehat{\alpha}_M - \alpha_P), \tag{25}$$

where $\rho_1(\alpha_M) = \overline{\rho} - m(\alpha_M - \underline{\alpha_M})$ is the matching function;
(ii) $\widehat{\alpha}_M > \alpha_P$ if and only if profit is better measured than impact at the firm matched with $\widehat{\alpha}_M$, i.e., $\rho_2 > \rho_1(\widehat{\alpha}_M)$.

Corollary 2 characterizes the manager who earns the highest rent. The condition on the slope of the matching function ensures that the rent function is strictly concave, while the two endpoint conditions ensure that its unique maximizer is interior. Equation (25) reflects two opposing forces. The left-hand side is the measurement advantage of profits over impact at the firm matched with the manager, which increases the value of preferences for impact. The right-hand side captures the cost of the manager's bias relative to the principal, which increases with the deviation costs c and k. Thus, managers with too little impact bias do not fully compensate for poor impact measurement, while those with too much bias create excessive resource-allocation distortions. The highest-earning manager balances these two forces.

Part (ii) determines the direction of this bias: the highest-earning manager is biased towards impact ($\widehat{\alpha}_M > \alpha_P$) if and only if profits are better measured than impact at his matched firm. A sufficient condition is

$$\rho_2 > \rho_1(\alpha_P) = \bar{\rho} - m\,(\alpha_P - \underline{\alpha_M}). \tag{26}$$

This condition compares profit measurement with impact measurement at the firm that would be matched with a congruent manager. If profits are better measured at this benchmark firm, managers with stronger preferences for impact than the principal earn the highest rents.

We now consider two examples that allow us to solve numerically for the manager's rent as a function of his preferences.[7]

Example 1 *Assume $\rho_1 \sim \mathcal{U}[0.3, 0.7]$, $\rho_2 = 0.9$, $\alpha_M \sim \mathcal{U}[0.3, 0.7]$, $\alpha_P = 0.5$, $c = k = 1$. The highest rent is earned by $\widehat{\alpha}_M = 0.63$, which is above $\alpha_P = 0.5$.*

Example 2 *Assume $\rho_1 \sim \mathcal{U}[0.1, 0.4]$ while all other parameters are unchanged. Measurement of dimension 1 is worse than in Example 1. The highest rent is now earned by $\widehat{\alpha}_M = 0.7$, a manager with even stronger preferences for dimension 1 than in Example 1.*

Figure 1 depicts the two rent schedules. In Example 1, dimension 1 is measured more poorly, and so the highest rent is earned by a manager biased towards dimension 1. However, this bias is interior: $\widehat{\alpha}_M < \overline{\alpha_M}$. In Example 2, measurement of dimension 1 is even poorer, increasing the

[7]Example 1 satisfies the conditions of Corollary 2, whereas Example 2 illustrates a boundary maximizer.

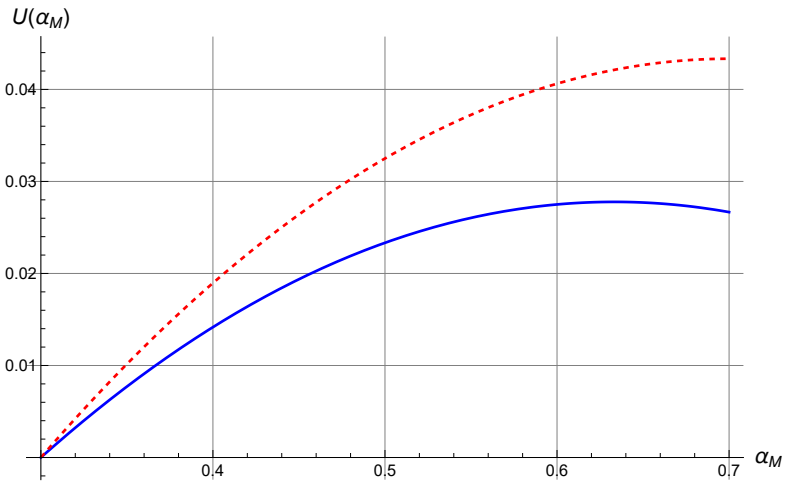

Figure 1: Comparison of rent schedules for managers with different preferences α_M under different measurement quality for dimension 1. The solid blue line is the agency rent in Example 1 with $\rho_1 \in [0.3, 0.7]$. The dashed red line is the agency rent in Example 2 with $\rho_1 \in [0.1, 0.4]$.

demand for managers with stronger preferences for dimension 1. This raises rents overall, as the least valuable manager type must still be willing to participate, with a rent of zero. Now the highest rent accrues to the manager with the most extreme bias: $\widehat{\alpha}_M = \overline{\alpha_M}$.

4.4 Equilibrium incentives

Having solved for the firm a manager works for and the level of rents he receives, we finally study the incentives he is given. From Lemma 2, incentives are given by:

$$b_1 = \frac{c\left(2\rho_1 + \rho_2 - k\Delta\right)}{3(c+k)} \tag{27}$$

$$b_2 = \frac{c\left(\rho_1 + 2\rho_2 + k\Delta\right)}{3(c+k)}. \tag{28}$$

To study how the manager's type affects incentives, we take the total derivative with respect to α_M, which yields:

$$\frac{db_1}{d\alpha_M} = \frac{c}{3(c+k)}\left[-k + 2\rho_1'\left(\alpha_M\right)\right]$$

$$\frac{db_2}{d\alpha_M} = \frac{c}{3(c+k)}\left[k + \rho_1'\left(\alpha_M\right)\right].$$

The effect of the manager's type on incentives can thus be decomposed into two components. The first is the *direct effect* holding the matched firm constant, which is given by the first term in square brackets. As in the single-firm model, a manager biased towards dimension 1 receives weaker incentives on dimension 1 and stronger incentives on dimension 2. The second

is the *sorting effect*: the manager will be matched with a firm that measures dimension 1 less accurately: $\rho_1'(\alpha_M) < 0$ due to negative assortative matching. Such a firm provides weaker incentives on dimension 1, and also on dimension 2 due to the spillover effect in Lemma 2. This reinforces the direct effect for b_1 (both terms in square brackets are negative) but counteracts the direct effect for b_2. The overall effect is given in Proposition 5:

Proposition 5 *At any α_M, the equilibrium incentive $b_1(\alpha_M)$ is strictly decreasing in α_M. The equilibrium incentive $b_2(\alpha_M)$ is locally increasing in α_M if $k > -\rho_1'(\alpha_M)$, and locally decreasing if $k < -\rho_1'(\alpha_M)$.*

Consider two managers with types $\alpha_M^H > \alpha_M^L$. Manager H always receives weaker incentives for dimension 1 than manager L. He receives stronger incentives for dimension 2 if $k > -\rho_1'(\alpha_M)$ holds for all $\alpha_M \in [\alpha_M^L, \alpha_M^H]$, and weaker incentives if $k < -\rho_1'(\alpha_M)$ holds for all $\alpha_M \in [\alpha_M^L, \alpha_M^H]$.

The direct effect is increasing in the firm's adjustment cost k. A biased manager (high Δ) allocates too many resources to dimension 1. This is more costly to the firm when k is high, so it responds by providing strong incentives for dimension 2: in equation (28), b_2 is increasing in $k\Delta$. The sorting effect is increasing in the slope of the matching function $-\rho_1'(\alpha_M)$. From Proposition 3, we have

$$-\rho_1'(\alpha_M) = \frac{f_{\alpha_M}(\alpha_M)}{f_\rho(\rho_1(\alpha_M))}.$$

This is large when variation in firm measurement is high relative to variation in manager preferences. When ρ_1 and α_M are both uniformly distributed, this becomes:

$$-\rho_1'(\alpha_M) = \frac{\overline{\rho} - \underline{\rho}}{\overline{\alpha_M} - \underline{\alpha_M}}.$$

The slope is then constant, so the comparison between any two types reduces to $k \gtrless \frac{\bar{\rho}-\underline{\rho}}{\overline{\alpha_M}-\underline{\alpha_M}}$. If $\overline{\rho} - \underline{\rho}$ is large and $\overline{\alpha_M} - \underline{\alpha_M}$ is small, managers with only slightly different preferences will be matched to firms with substantially different measurement, leading to a large sorting effect.

Proposition 5 shows how endogenous matching can either preserve or reverse the single-firm relation between managerial bias and incentives in Lemma 2. If the direct effect dominates, managers with stronger preferences for dimension 1 receive higher incentives for dimension 2, as in the single-firm model – even in a market equilibrium with heterogeneous firms and managers. If the sorting effect dominates, the relationship in Lemma 2 reverses. Managers with stronger preferences for dimension 1 are matched with firms that measure dimension 1 poorly, which weakens incentives on dimension 1 and, through the spillover effect in Lemma 2, also weakens incentives on dimension 2. This prediction of weak incentives on both dimensions is generated by matching rather than by the standard multitasking channel. Conversely, managers with weaker

preferences for dimension 1 are matched with firms that measure dimension 1 well and therefore receive stronger incentives on both dimensions.

Although matching can reverse the relationship between managerial bias and the level of incentives on the better-measured dimension derived in the single-firm model, it cannot reverse the relationship between bias and the incentive gap. Subtracting equation (27) from (28) yields:

$$b_2 - b_1 = \frac{c}{3(c+k)} \left[\rho_2 - \rho_1(\alpha_M) + 2k(\alpha_M - \alpha_P)\right].$$

Since $\rho_1'(\alpha_M) < 0$ from Proposition 3,

$$\frac{d(b_2 - b_1)}{d\alpha_M} = \frac{c}{3(c+k)} \left[2k - \rho_1'(\alpha_M)\right] > 0.$$

Thus, managers with stronger preferences for dimension 1 always receive relatively stronger incentives on dimension 2.

Figure 2 illustrates these two regimes using the same parameter values as in Example 1, except for k. In Panel A, with $k = 2$, the direct effect dominates: managers with stronger preferences for dimension 1 receive higher incentives for dimension 2 and weaker incentives for dimension 1. In Panel B, with $k = 0.5$, the sorting effect dominates: they receive weaker incentives on both dimensions.

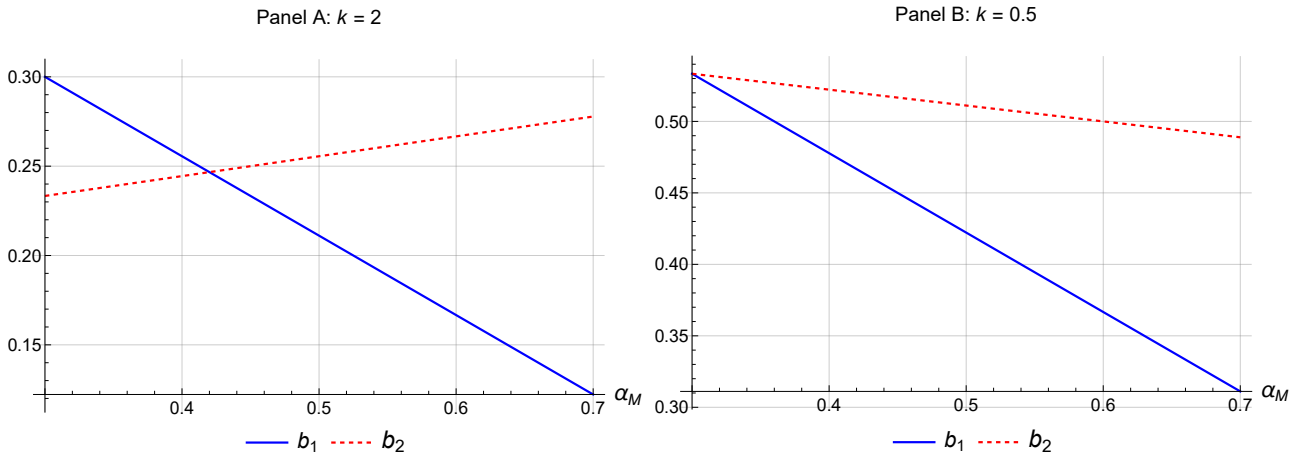

Figure 2: Equilibrium incentives b_1 and b_2 as a function of managerial preferences α_M.

5 Unobservable Managerial Types

This section considers the case in which managers' types are unobservable but can instead be signaled through a costly action. We characterize the conditions under which the signal is credible, in which case the matching outcome derived in Section 4 continues to obtain.

Formally, we now assume that α_M is the manager's private information. In addition, prior to matching, each manager publicly chooses an observable activity $x \in \mathbb{R}$. Examples include prior managerial decisions, charitable donations, and volunteer activities. The disutility of this activity is captured by the same quadratic cost deviation used in the specification of individual preferences in equation (5):

$$C(x, \alpha_M) = \frac{\gamma}{2}\big(x - \alpha_M\big)^2, \tag{29}$$

where $\gamma > 0$ can be different from c. Thus, it is costly for the manager to send a signal that differs from his own type in either direction. The signaling cost satisfies the single-crossing property: $\frac{\partial^2 C}{\partial x \partial \alpha_M} = -\gamma$.

Firms observe the activity x, develop a belief $\beta(x)$ about the manager's type, and then enter the matching market of section 4. We restrict attention to fully separating Perfect Bayesian Equilibria ("PBE") with equilibrium beliefs such that $\beta(x(\alpha_M)) = \alpha_M$ and the divinity refinement of Banks and Sobel (1987), which for signaling games with a continuum of types selects the least-cost separating schedule (Ramey, 1996). Each manager decides whether to enter the market before incurring the signaling cost, so his participation constraint is that his equilibrium utility net of the signaling cost be at least his reservation utility $\overline{U}$. As in Section 4, we continue to assume normalized volatilities: $\sigma_1^v = \sigma_1^p = \sigma_2^v = \sigma_2^p = 1$. The densities f_ρ and f_{α_M} are continuous and bounded above and below away from zero, and so the matching slope $-\rho_1'(\alpha_M) = \frac{f_{\alpha_M}(\alpha_M)}{f_\rho(\rho_1(\alpha_M))}$ satisfies $0 < -\rho_1'(\alpha_M) < \infty$. A type-$\alpha_M$ manager who sends the signal intended for type t is matched with firm with performance measurement $\rho_1(t)$ and offered the type-t contract. Before subtracting his signaling cost, his payoff is $U(t) + \overline{U} + (b_1(t) - b_2(t))(\alpha_M - t)$. We call an agent's rent his overall utility (net of the signaling cost) above reservation utility $\overline{U}$, and let $U(\cdot)$ be the rent schedule of Proposition 4.

Proposition 6 *There exists $\bar{\gamma} > 0$ such that, for all $\gamma > \bar{\gamma}$, there is a fully separating PBE in which:*
(i) The signaling schedule $x(\alpha_M)$ is strictly increasing and satisfies:

$$\gamma\left(x(\alpha_M) - \alpha_M\right) x'(\alpha_M) = \frac{1}{3}\left(\rho_2 - \rho_1(\alpha_M) - k(\alpha_M - \alpha_P)\right), \tag{30}$$

for every interior type α_M;
(ii) Firms infer each manager's type α_M correctly from his signal x, so the matching assignment and incentives b_1 and b_2 are the same as in Section 4. Fixed payments adjust to account for signaling costs;
(iii) The equilibrium signaling cost is $C(x(\alpha_M), \alpha_M)$, and each type's rent is equal to:

$$U(\alpha_M) - C(x(\alpha_M), \alpha_M) - \min_{\alpha' \in [\underline{\alpha_M}, \overline{\alpha_M}]} \left\{U(\alpha') - C(x(\alpha'), \alpha')\right\}, \tag{31}$$

Proposition 6 shows that if the cost of misrepresenting one's type is sufficiently high, there exists a fully separating equilibrium in which each manager's signal uniquely identifies his type. Firms therefore infer managers' types correctly, so the same matching assignment and incentives b_1 and b_2 obtain as in Section 4; fixed payments adjust to account for signaling costs.

In equilibrium, if the right-hand side ("RHS") of (30) is zero for some type within the support of managerial types, there is a pivot type whose signal is undistorted, i.e. who chooses $x = \alpha_M$. This is the type with no first-order gain from being perceived as a marginally different type (without considering signaling costs). Managers for whom the RHS is positive would prefer to be perceived as a higher type, while those for whom it is negative would prefer to be perceived as a lower type, since doing so would give them a match and contract that they value more. If instead the RHS is positive for all types, the signal is undistorted for the lowest type, because there is no lower type. The case with a negative RHS for all types is symmetric. Thus, the separating equilibrium requires misrepresentation to be sufficiently costly. As γ approaches infinity, signals converge to true types and signaling costs vanish in equilibrium, so the observable-type matching outcome of Section 4 is recovered.

With uniform distributions as in Corollary 2, the RHS of (30) is affine in α_M and the ordinary differential equation in equation (30) has a closed-form solution. This is given in Corollary 3.

Corollary 3 *With $\rho_1 \sim \mathcal{U}[\underline{\rho}, \bar{\rho}]$, $\alpha_M \sim \mathcal{U}[\underline{\alpha_M}, \overline{\alpha_M}]$, matching slope $m = \frac{\bar{\rho}-\underline{\rho}}{\overline{\alpha_M}-\underline{\alpha_M}}$, and if there exists $\alpha_0 \in [\underline{\alpha_M}, \overline{\alpha_M}]$ such that*

$$\rho_2 - \rho_1(\alpha_0) - k(\alpha_0 - \alpha_P) = 0, \tag{32}$$

the threshold $\bar{\gamma}$ of Proposition 6 is explicit: a fully separating PBE exists whenever

$$\gamma > \max\left\{\frac{2c(m+2k)}{3(c+k)}, \frac{4(k-m)}{3}\right\}, \tag{33}$$

and the signal schedule is affine,

$$x(\alpha_M) - \alpha_M = \frac{1}{3\gamma(1+\lambda)}\big[\rho_2 - \rho_1(\alpha_M) - k(\alpha_M - \alpha_P)\big], \qquad \lambda = \frac{-1 + \sqrt{1 + \frac{4(m-k)}{3\gamma}}}{2}. \tag{34}$$

The direction of the distortion depends on $m - k$ exactly as in the matching analysis of Section 4. With uniform distributions, the slope of the matching function is constant and equal to m, which captures the strength of the sorting effect: a higher m means that a given change in managerial type leads to a larger change in the matched firm's measurement quality. From Corollary 3, when the sorting effect dominates ($m > k$), then $\lambda > 0$ and $x'(\alpha_M) = 1 + \lambda > 1$. Managers below the pivot type signal below their true types, while managers above the pivot

signal above them, amplifying differences in stated preferences. When the direct effect dominates ($k > m$), then $\lambda < 0$ and $x'(\alpha_M) = 1 + \lambda < 1$. Managers below the pivot signal above their true types, while managers above the pivot signal below them, compressing signals toward the pivot. Intuitively, when the firm's adjustment cost k is high relative to the sorting benefit m, firms prefer managers whose preferences are closer to their own. Managers therefore signal more moderate preferences by choosing signals closer to the pivot type.

Example 3 *Assume* $k = 2$, $\rho_2 = 0.5$, $\rho_1 \sim \mathcal{U}[0.3, 0.7]$, $\alpha_M \sim \mathcal{U}[0.3, 0.7]$, $\alpha_P = 0.5$, *and* $c = 1$. *Then* $m = 1 < k$, *so the direct effect dominates the sorting effect. The pivot type is* 0.5. *With* $\gamma = 2$, *condition 33 holds and* $\lambda \approx -0.211$, *so* (34) *gives* $x(\alpha_M) \approx 0.789\,\alpha_M + 0.106$. *The signal slope is below one and* $x(0.3) \approx 0.343$ *and* $x(0.7) \approx 0.658$*: managers on both sides of the pivot signal more moderate preferences.*

Example 4 *Assume* $k = 0.5$ *while all other parameters are unchanged. Then* $m = 1 > k$, *so the sorting effect dominates the direct effect. We have* $\lambda \approx 0.077$, *and* $x(\alpha_M) \approx 1.077\,\alpha_M - 0.039$. *The signal slope is above one and* $x(0.3) \approx 0.285$ *and* $x(0.7) \approx 0.715$*: managers signal more extreme preferences.*

Examples 3 and 4 hold measurement heterogeneity fixed and vary only the firm's adjustment cost k. When k is high, the direct effect dominates and signals are compressed toward the pivot; managers with stronger preferences for dimension 1 receive weaker incentives for dimension 1 but stronger incentives for dimension 2. When k is low, the sorting effect dominates and signals are spread away from the pivot; managers with stronger preferences for dimension 1 receive weaker incentives on both dimensions.

In sum, the signaling model yields implications for observable managerial behaviour. When firms value managers with stronger preferences for a particular activity because it is difficult to measure, candidates have incentives to demonstrate these preferences through costly actions. If the relevant activity is societal impact, such signals may include charitable donations, volunteering, accepting positions in mission-driven organizations, or pursuing careers with lower financial rewards. If the relevant activity is profitability or commercial discipline, they may instead include leading corporate turnarounds or taking difficult decisions such as closing plants or reducing employment to improve long-run profitability despite their personal unpopularity. Thus, our model can account for "profit signaling" as well as "virtue signaling". Conversely, when firms place greater value on moderation, candidates have incentives to present themselves as more balanced by avoiding actions that reveal extreme preferences.

6 Implications

The model yields implications for the joint distribution of measurement quality, managerial type, incentive contracts, and compensation. This section discusses these implications.

Implication 1 *Managerial preferences should be matched to measurement technology. Within a firm, the manager should be biased toward the worse-measured dimension; across firms, those with worse measurement of a given dimension should hire managers with stronger preferences for that dimension.*

Implication 1 concerns managerial selection. Empirically, a stronger preference for impact might be reflected in donations, affiliations, and voting records (Christensen et al. (2017), Di Giuli and Kostovetsky (2014)). Section 5 provides a signaling microfoundation for these proxies: they are costly actions whose costs depend on managerial type and therefore shape firms' beliefs about managers' preferences. Exogenous shocks to impact preferences can provide a complementary source of identification (e.g. Cronqvist and Yu (2017)). Better impact measurement would be reflected in a higher correlation between impact ratings or scores (Berg, Kölbel, and Rigobon (2022)), ES reporting standards, third-party verification, or anti-greenwashing enforcement.

Implication 2 *A manager who is more biased toward the poorly-measured dimension should receive weaker incentives on that dimension.*

Implication 2 concerns incentives. For example, more impact-motivated executives should have weaker impact incentives. This prediction holds in both the single-firm model and the matching equilibrium.

Implication 3 *A manager who is more biased toward the poorly-measured dimension should receive stronger incentives on the well-measured dimension if firms face large adjustment costs or if there is little cross-firm variation in scaled quality of the poorly-measured dimension relative to the dispersion of managerial preferences. Otherwise, he receives weaker incentives on the well-measured dimension.*

Implication 3 emphasizes that the effects of performance measurement and managerial selection on incentives for the well-measured dimension act in opposite directions. The direct effect identified in the single-firm model of section 3 is that a manager more biased towards impact will receive stronger profit incentives. The matching model of section 4 identifies a potentially offsetting effect, which will dominate if there is enough variation in the scaled quality of impact measurement across firms, or firms' adjustment costs are low.

Implication 4 *Managers with intermediate biases earn the highest rents when the measurement asymmetry at their matched firm is sufficiently high (low) relative to their bias for the least (most) biased managers. When one of these two forces dominates for all manager types, a manager with an extreme type earns the highest rent.*

As highlighted in Corollary 2, the rent can be either monotonic or nonmonotonic in the manager's type. A manager's bias is valuable because it compensates for poor measurement, but costly because it distorts resource allocation. When firms' and managers' deviation costs are low, the first effect dominates over a wide range and the most biased manager available earns the highest rent. When they are high, the deadweight loss from bias grows more quickly and the highest rent accrues to a manager with an intermediate bias.[8]

Implication 5 *Following a market-wide improvement in the quality of performance measurement on one dimension that leaves the ranking of firms unchanged, managers receive stronger incentives on both dimensions, with the larger increase on the improved dimension, and the marginal rent of a manager with a stronger preference for this dimension is lower.*

If impact performance becomes better measured following mandatory disclosure regulation, each firm would prefer a manager with a more moderate impact bias, since impact can now be induced through explicit incentives (they may still prefer some impact bias, if impact remains worse-measured than profits). However, because the improvement is market-wide and leaves the ranking of firms unchanged, the equilibrium assignment does not change: all firms' preferences shift together, and no firm-manager pair can profitably re-match. What adjusts is the price of bias: the marginal rent of managers with stronger impact preferences falls. In addition, incentives become higher-powered on the impact dimension, so that managers are more responsive to their information, and also on the profit dimension, through the spillover effect of Lemma 2.

7 Applications

We now consider several applications of our model, and discuss the empirical implications in each context.

7.1 Executive compensation

In some firms, shareholders have a purely financial objective: they seek to maximize long-term shareholder value. However, long-term shareholder value is non-contractible because the

[8]The rent-maximizing type satisfies $\widehat{\alpha}_M = \alpha_P + \frac{\rho_2 - \rho_1(\widehat{\alpha}_M)}{3c+k}$, so it lies further from α_P when measurement is more asymmetric and closer to α_P when c and k are larger. When it falls outside $[\underline{\alpha_M}, \overline{\alpha_M}]$, the highest rent is earned by the most extreme available type. A steep matching function also pushes towards a corner, since the rent schedule is concave only if $m < 3c + k$.

manager's consumption needs limit the extent to which he can be given restricted equity. Thus, he will be paid at least in part on the short- or medium-term stock price. Some business activities may improve long-term shareholder value but not be fully reflected in the stock price, such as customer satisfaction (Fornell et al., 2006), employee satisfaction (Edmans, 2011), and other intangible assets. This is in part because they are poorly measured, which also makes them difficult to incentivize directly. Our model suggests that this could be addressed by hiring managers with a preference for these activities, and incentivizing them for financial performance using equity incentives. Our single-firm model predicts that profit incentives will be stronger for managers with greater preferences for impact.[9]

Another application is to venture capital. Early-stage firms often have measurable short-term metrics such as user growth or revenue, but their long-term value depends on poorly measured dimensions such as product quality, innovation, and organizational culture. Our model suggests that investors may optimally back founders who are intrinsically biased toward these harder-to-measure dimensions, while using strong financial incentives tied to observable metrics. This helps prevent excessive focus on short-term performance while preserving responsiveness to measurable outcomes.

While shareholders in the above examples care exclusively about long-term shareholder value, a third application concerns firms whose shareholders have non-financial objectives, such as societal impact (Hart and Zingales, 2017). Some proxy advisors assess a company's executive compensation scheme in part on the relative strength of profit and ES incentives, and commentators often interpret strong ES (profit) incentives as an unambiguously positive (negative) commitment to a societal mission. Our model cautions against such interpretations. ES incentives may be an ineffective way to pursue impact given measurement difficulties and the associated agency costs (Jung, 2025); instead, hiring a manager with intrinsic preferences for impact may be more effective. Indeed, ES incentives should be optimally low if such a manager is hired. Similarly, an increase in ES incentives is only warranted following improvements in impact measurement, such as improved disclosure standards or third-party verification such as sustainability audits, or following changes in manager type, adjustment or deviation costs, or profit measurement. Otherwise, such an increase would increase "hitting the target, missing the point" – improving measured rather than actual impact.

Finally, the model identifies the settings in which diversity in managerial preferences is valuable: when the relative quality of performance measurement varies across firms. In such environments, firms differ in the extent to which they can rely on incentives: firms with weak measurement of impact rely more on bias and thus prefer managers who place greater weight on impact, whereas firms with strong measurement rely more on incentives and prefer managers

[9]In addition, as in Baker (1992), these incentives will be stronger if the stock price is a better measure of firm value, for example due to improvements in stock market efficiency.

with weaker impact preferences. As a result, a heterogeneous pool of managerial types improves matching efficiency. This has implications for policies that homogenize or diversify the pool of available managers, such as MBA admissions and curricula, and for the value of cognitive diversity within organizations.

7.2 Universities

A university employs professors who engage in a range of activities. For simplicity, we will focus on just two (research and teaching), although the implications extend to other activities such as internal contribution and external visibility. These professors must decide how to allocate their time between teaching and research.

Some institutions place greater weight on the quantity of research output, which is relatively easy to incentivize – for example, through bonuses or pay rises tied to the number of publications. If they primarily teach undergraduate students, who may be less able to assess teaching quality (for example, because these students prioritize test-taking skills over critical thinking or real-world relevance), common measures such as student evaluations are noisy and easily gamed. In such settings, explicit incentives for teaching may backfire. Instead, these institutions rely on selection and culture, hiring faculty with intrinsic motivation for teaching and fostering a "teaching culture" to sustain output on this dimension.

Other institutions emphasize research quality and societal impact, which are difficult to measure, particularly in the short-term. If they primarily teach more experienced students, such as MBAs, Masters, and PhDs, who can better evaluate teaching, performance on the teaching dimension becomes more informative. These institutions can therefore rely more on explicit incentives for teaching, such as overload payments or executive education opportunities, while selecting faculty with strong intrinsic motivation for research.

In summary, the model predicts that universities will rely more on intrinsic motivation and culture in dimensions that are difficult to measure and prone to gaming, and use explicit incentives for better-measured dimensions.[10]

7.3 Healthcare

Healthcare offers two distinct applications of the model: individual doctors and hospital CEOs. Starting with doctors, some activities ("cure"), such as performing procedures, ordering tests, and prescribing drugs, generate measurable outputs and are often linked to compensation through fee-for-service payments or productivity targets. Others ("prevention"), such as giving

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

advice and better understanding individual cases to avoid unnecessary procedures, affect longer-term outcomes that are difficult to measure. Since doctors are best placed to assess which activities are appropriate for a given patient, this creates a classic resource-allocation problem with private information.

Our model suggests that hospitals should hire doctors with an intrinsic preference for prevention, and complement this with incentives tied to cure. Concerns about overuse of revenue-generating procedures and underinvestment in prevention (e.g. Gawande (2009)) may reflect not the use of incentives per se, but a mismatch between incentives and the preferences of the doctors to whom they are applied. When doctors lack intrinsic motivation for patient-centered care, incentives tied to measurable outputs can lead to distortions. In such cases, the solution is not to reduce incentives but to hire doctors with strong preferences for prevention and create a patient-first culture. Only if this is not possible (e.g. because doctor preferences are relatively homogeneous or culture is difficult to change) should they reduce incentives for measured performance.

For hospital CEOs, the allocation is between readily measured financial performance and harder-to-measure patient quality and community benefit. Our model predicts that hospitals should hire impact-motivated CEOs and incentivize them primarily on financial performance. Consistent with physician status proxying for stronger intrinsic preferences for patient care, Goodall (2011) finds that hospital quality is positively associated with having a physician rather than a non-physician CEO. Jenkins, Short and Ho (2025) find that although hospital boards report using quality to assess organizational performance, nonprofit hospital CEO compensation is increasingly associated with profits and organizational size rather than quality.

Rather than being paradoxical, these findings are complementary in our model. A physician CEO's intrinsic motivation substitutes for explicit quality incentives, allowing compensation to focus on the well-measured financial dimension. This yields a novel empirical prediction: the gap between financial and quality incentives should be larger for physician CEOs than for non-physician CEOs.

8 Conclusion

This paper develops a new theory of managerial selection and incentive provision. We study a setting in which a manager with private information allocates resources across competing activities. When performance is imperfectly and asymmetrically measured, the principal optimally appoints a biased manager – one whose preferences differ from her own. In particular, the principal prefers a manager biased toward the more poorly measured dimension, as this bias mitigates distortions in resource allocation induced by incentive provision. As a result, changes in the measurement technology not only affect optimal contracts, but also the type of managerial

preferences that are most valuable.

The model explains the coexistence of high-powered incentives on well-measured dimensions and low-powered incentives on poorly measured ones, even when these activities are substitutes. By selecting a biased manager, the principal can rely on the manager's preferences to ensure resource allocation to poorly measured activities, while using strong incentives where they are most effective. This selection channel amplifies asymmetries in pay-for-performance across dimensions and highlights the importance of jointly determining compensation and hiring policies.

In a competitive market, firms with weaker measurement of a dimension match with managers who place greater weight on it, generating differences in managerial rents. Endogenous sorting can also reverse the within-firm relation between preferences and incentives. When managerial preferences are unobservable, signals that are sufficiently costly to mimic can sustain the same assignment and incentive sensitivities.

Our results build on the idea that an important part of education is to endow agents destined for careers in some field with preferences appropriate for this field (Akerlof and Kranton (2005), Van den Steen (2010)). Our contribution is to show that fostering a diversity of preferences can improve matching between workers and firms, and that the most valuable preferences to endow depend on the nature of performance measurement. More broadly, incentive contracts cannot be interpreted independently of managerial selection. Strong incentives on one dimension may be optimal precisely because the organization has selected a manager intrinsically committed to another.

9 References

Aghion, P. and Stein, J.C., 2008. Growth versus margins: Destabilizing consequences of giving the stock market what it wants. Journal of Finance 63, 1025-1058.

Aghion, P. and Tirole, J., 1997. Formal and real authority in organizations. Journal of Political Economy 105, 1-29.

Akerlof, G.A. and Kranton, R.E., 2005. Identity and the economics of organizations. Journal of Economic Perspectives 19, 9-32.

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

Alonso, R. and Matouschek, N., 2008. Optimal delegation. Review of Economic Studies 75, 259-293.

Athey, S. and Roberts, J., 2001. Organizational design: Decision rights and incentive contracts. American Economic Review 91, 200-205.

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

Banks, J.S. and Sobel, J., 1987. Equilibrium selection in signaling games. Econometrica 55, 647-661.

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

Bebchuk, L.A. and Stole, L.A., 1993. Do short-term managerial objectives lead to under- or overinvestment in long-term projects? Journal of Finance 48, 719-729.

Bebchuk, L.A. and Tallarita, R., 2023. Will corporations deliver value to all stakeholders? Vanderbilt Law Review 75, 1031-1091.

Becker, G. S. 1973. A theory of marriage: Part I. Journal of Political Economy, 81, 813-846.

Ben-David, I. and Chinco, A., 2026. The max EPS paradigm for corporate finance. Working paper, National Bureau of Economic Research.

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, 1315-1344.

Besley, T. and Ghatak, M., 2005. Competition and incentives with motivated agents. American Economic Review 95, 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, 404-414.

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

Carlin, B.I. and Gervais, S., 2009. Work ethic, employment contracts, and firm value. Journal of Finance 64, 785-821.

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

Christensen, D.M., Mikhail, M., Walther, B. and Wellman, L.A., 2017. From K Street to Wall Street: Politically connected analysts and stock recommendations. The Accounting Review 92, 1-28.

Cronqvist, H. and Yu, F., 2017. Shaped by their daughters: Executives, female socialization, and corporate social responsibility. Journal of Financial Economics 126, 543-562.

DeMarzo, P.M. and Fishman, M.J., 2007. Optimal long-term financial contracting. Review of Financial Studies 20, 2079-2128.

DeMarzo, P.M. and Sannikov, Y., 2006. Optimal security design and dynamic capital structure in a continuous-time agency model. Journal of Finance 61, 2681-2724.

Dessein, W., 2002. Authority and communication in organizations. Review of Economic Studies 69, 811-838.

Di Giuli, A. and Kostovetsky, L., 2014. Are red or blue companies more likely to go green? Politics and corporate social responsibility. Journal of Financial Economics 111, 158-180.

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

Edmans, A., Fang, V.W. and Lewellen, K.A., 2017. Equity vesting and investment. Review of Financial Studies 30, 2229-2271.

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

Fehr, E. and Charness, G., 2025. Social preferences: fundamental characteristics and economic consequences. Journal of Economic Literature 63, 440-514.

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

Fornell, C., Mithas, S., Morgeson III, F.V. and Krishnan, M.S., 2006. Customer satisfaction and stock prices: High returns, low risk. Journal of Marketing, 70, 3-14.

Gabaix, X. and Landier, A., 2008. Why has CEO pay increased so much? Quarterly Journal of Economics 123, 49-100.

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

Goodall, A.H., 2011, Physician-leaders and hospital performance: Is there an association?, Social Science & Medicine 73, 535-539.

Gopalan, R., Milbourn, T., Song, F. and Thakor, A.V., 2014. Duration of executive compensation. Journal of Finance 69, 2777-2817.

Hart, O. and Zingales, L., 2017. Companies should maximize shareholder welfare not market value. Journal of Law, Finance, and Accounting, 2, 247-275.

Hartzell, J.C., Parsons, C.A. and Yermack, D.L., 2010. Is a higher calling enough? Incentive compensation in the church. Journal of Labor Economics 28, 509-539.

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.

Jenkins, D., Short, M. and Ho, V., 2025. Nonprofit hospital CEO compensation: Does quality matter? Medical Care 63, 787-793.

Jung, K., 2025. Green moral hazard: Estimating the financial and the real implications of CEO incentives. Working Paper, New York University.

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

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

Mailath, G.J., 1987. Incentive compatibility in signaling games with a continuum of types. Econometrica 55, 1349-1365.

Narayanan, M.P., 1985. Managerial incentives for short-term results. Journal of Finance 40, 1469-1484.

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

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

Ramey, G., 1996. D1 signaling equilibria with multiple signals and a continuum of types. Journal of Economic Theory 69, 508-531.

Rotemberg, J.J. and Saloner, G., 2000. Visionaries, managers, and strategic direction. RAND Journal of Economics 31, 693-716.

Stein, J.C., 1988. Takeover threats and managerial myopia. Journal of Political Economy 96, 61-80.

Stein, J.C., 1989. Efficient capital markets, inefficient firms: A model of myopic corporate behavior. Quarterly Journal of Economics 104, 655-669.

Terviö , M., 2008. The Difference That CEOs Make: An Assignment Model Approach. American Economic Review 98, 642-668.

Van den Steen, E., 2005. Organizational beliefs and managerial vision. Journal of Law, Economics, and Organization 21, 256-283.

Van den Steen, E., 2010. On the origin of shared beliefs (and corporate culture). RAND Journal of Economics 41, 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, 630-651.

10 Proofs

Proof of Lemma 1: The agent chooses a to maximize his objective function in equation (5). At the time of choosing his action, the agent observes ϵ_1^p and ϵ_2^p, so that his objective function can be rewritten as:

$$S + b_1 a\epsilon_1^p + b_2(1-a)\epsilon_2^p - \frac{c}{2}\left(a - \alpha_M\right)^2$$

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

$$b_1\epsilon_1^p - b_2\epsilon_2^p - c\left(a - \alpha_M\right) = 0$$

Rearranging gives the expression in Lemma 1. ■

Proof of Lemma 2: Plugging equation (4) into equation (3), the objective function of the principal is:

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

The participation constraint and incentive constraint are respectively:

$$\mathbb{E}\left[S + b_1 P_1(a,\epsilon) + b_2 P_2(a,\epsilon) - \frac{c}{2}\left(a - \alpha_M\right)^2\right] \geq \overline{U} \tag{35}$$

$$b_1\epsilon_1^p - b_2\epsilon_2^p - c\left(a - \alpha_M\right) = 0 \quad \Leftrightarrow \quad a = \alpha_M + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c} \tag{36}$$

Substituting in equation (3) for the binding participation constraint in equation (35), the principal's optimization program rewrites as:

$$\max_{b} \mathbb{E}\left[V_1(a,\epsilon) + V_2(a,\epsilon) - \frac{c}{2}\left(a - \alpha_M\right)^2 - \overline{U} - \frac{k}{2}\left(a - \alpha_P\right)^2\right]$$

This objective function can be rewritten as:

$$\mathbb{E}\Bigg[\left(\alpha_M + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\epsilon_1^v + \left(1 - \alpha_M - \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\epsilon_2^v - \frac{c}{2}\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)^2 - \overline{U}$$
$$- \frac{k}{2}\left(\alpha_M - \alpha_P + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)^2\Bigg]$$

The FOC with respect to b_1 is:

$$\mathbb{E}\left[\frac{\epsilon_1^p}{c}\epsilon_1^v - \frac{\epsilon_1^p}{c}\epsilon_2^v - c\epsilon_1^p\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c^2}\right) - k\frac{\epsilon_1^p}{c}\left(\alpha_M - \alpha_P + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\right] = 0$$

$$\Leftrightarrow \mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right] - \mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right] - b_1\mathbb{E}\left[\epsilon_1^{p2}\right] + b_2\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right] - k\left((\alpha_M - \alpha_P)\mathbb{E}\left[\epsilon_1^p\right] + \frac{b_1}{c}\mathbb{E}\left[\epsilon_1^{p2}\right] - \frac{b_2}{c}\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right]\right) = 0$$

$$\Leftrightarrow b_1 = \frac{\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right] - \mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right] - k(\alpha_M - \alpha_P)\mathbb{E}\left[\epsilon_1^p\right] + b_2\left(1 + \frac{k}{c}\right)\mathbb{E}\left[\epsilon_1^p\epsilon_2^p\right]}{\left(1 + \frac{k}{c}\right)\mathbb{E}\left[\epsilon_1^{p2}\right]}$$

$$\Leftrightarrow b_1 = \frac{\rho_1\sigma_1^v\sigma_1^p - k(\alpha_M - \alpha_P)}{\left(1 + \frac{k}{c}\right)\left(\sigma_1^{p2} + 1\right)} + b_2\frac{1}{\sigma_1^{p2} + 1} \tag{37}$$

Likewise, the FOC with respect to b_2 is:

$$\mathbb{E}\left[-\frac{\epsilon_2^p}{c}\epsilon_1^v + \frac{\epsilon_2^p}{c}\epsilon_2^v + c\epsilon_2^p\left(\frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c^2}\right) + k\frac{\epsilon_2^p}{c}\left(\alpha_M - \alpha_P + \frac{b_1\epsilon_1^p - b_2\epsilon_2^p}{c}\right)\right] = 0$$

$$\Leftrightarrow\ b_2 = \frac{\rho_2\sigma_2^v\sigma_2^p + k(\alpha_M - \alpha_P)}{\left(1 + \frac{k}{c}\right)\left(\sigma_2^{p2} + 1\right)} + b_1\frac{1}{\sigma_2^{p2} + 1} \tag{38}$$

Solving equations (37) and (38) for b_1 and b_2 yields:

$$b_1 = \frac{c\left(\sigma_2^{p2}\left(-k\left(\alpha_M - \alpha_P\right)\right) + \rho_2\sigma_2^p\sigma_2^v + \rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2} + 1\right)\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)} \tag{39}$$

$$b_2 = \frac{c\left(\sigma_1^{p2}k\left(\alpha_M - \alpha_P\right) + \rho_1\sigma_1^p\sigma_1^v + \rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2} + 1\right)\right)}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)} \tag{40}$$

The optimal pay-performance sensitivities for a given Δ are:

$$b_1(\Delta) = \frac{c\left[\rho_1\sigma_1^p\sigma_1^v\left(\sigma_2^{p2} + 1\right) + \rho_2\sigma_2^p\sigma_2^v - k\sigma_2^{p2}\Delta\right]}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)}, \tag{41}$$

$$b_2(\Delta) = \frac{c\left[\rho_2\sigma_2^p\sigma_2^v\left(\sigma_1^{p2} + 1\right) + \rho_1\sigma_1^p\sigma_1^v + k\sigma_1^{p2}\Delta\right]}{(c + k)\left(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}\right)}. \tag{42}$$

Since both expressions share the same denominator, taking the difference $b_2(\Delta) - b_1(\Delta)$ reduces

to expanding the numerator:

$$\begin{aligned}
&\rho_2\sigma_2^p\sigma_2^v\Big(\sigma_1^{p2}+1\Big)+\rho_1\sigma_1^p\sigma_1^v+k\sigma_1^{p2}\Delta\\
&\quad-\rho_1\sigma_1^p\sigma_1^v\Big(\sigma_2^{p2}+1\Big)-\rho_2\sigma_2^p\sigma_2^v+k\sigma_2^{p2}\Delta\\
&=\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2}-\rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}+k\Big(\sigma_1^{p2}+\sigma_2^{p2}\Big)\Delta, \qquad (43)
\end{aligned}$$

where the level terms $\rho_1\sigma_1^p\sigma_1^v$ and $\rho_2\sigma_2^p\sigma_2^v$ cancel. This yields the incentive gap stated in Lemma 2:

$$b_2(\Delta)-b_1(\Delta)=\frac{c\Big[\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2}-\rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}+k\Big(\sigma_1^{p2}+\sigma_2^{p2}\Big)\Delta\Big]}{(c+k)\Big(\sigma_1^{p2}\sigma_2^{p2}+\sigma_1^{p2}+\sigma_2^{p2}\Big)}.$$

The gap is increasing in Δ since the coefficient $ck(\sigma_1^{p2}+\sigma_2^{p2})/[(c+k)(\sigma_1^{p2}\sigma_2^{p2}+\sigma_1^{p2}+\sigma_2^{p2})]$ is strictly positive. ■

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

$$\begin{aligned}
&\mathbb{E}\Bigg[\left(\alpha_M+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)\epsilon_1^v+\left(1-\alpha_M-\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)\epsilon_2^v\\
&-\frac{c}{2}\left(\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2-\overline{U}-\frac{k}{2}\left(\alpha_M-\alpha_P+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\Bigg]\\
&=\left(\alpha_M\mathbb{E}\left[\epsilon_1^v\right]+\frac{b_1\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right]-b_2\mathbb{E}\left[\epsilon_2^p\epsilon_1^v\right]}{c}\right)+\left((1-\alpha_M)\mathbb{E}\left[\epsilon_2^v\right]-\frac{b_1\mathbb{E}\left[\epsilon_1^p\epsilon_2^v\right]-b_2\mathbb{E}\left[\epsilon_2^p\epsilon_2^v\right]}{c}\right)\\
&-\frac{1}{2c}\mathbb{E}\left[(b_1\epsilon_1^p-b_2\epsilon_2^p)^2\right]-\overline{U}-\frac{k}{2}\mathbb{E}\left[\left(\alpha_M-\alpha_P+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\right]\\
&=1+\frac{1}{c}\left(b_1\left(\mathbb{E}\left[\epsilon_1^p\epsilon_1^v\right]-1\right)+b_2\left(\mathbb{E}\left[\epsilon_2^p\epsilon_2^v\right]-1\right)\right)\\
&-\frac{1}{2c}\left(b_1^2\mathbb{E}\left[\epsilon_1^{p2}\right]+b_2^2\mathbb{E}\left[\epsilon_2^{p2}\right]-2b_1b_2\right)-\overline{U}-\frac{k}{2}\mathbb{E}\left[\left(\alpha_M-\alpha_P+\frac{b_1\epsilon_1^p-b_2\epsilon_2^p}{c}\right)^2\right] \qquad (44)
\end{aligned}$$

where b_1 and b_2 are described in Lemma 2. Since b_1 and b_2 are chosen optimally by the principal and are interior solutions, we use the envelope theorem that states that the total derivative of the objective function with respect to α_M is equal to the partial derivative holding b_1 and b_2 constant. This partial derivative is:

$$k\left(\alpha_P-\alpha_M+\frac{-b_1+b_2}{c}\right) \qquad (45)$$

Since the objective function of the principal is concave in α_M, setting this derivative to zero

gives:

$$\alpha_M = \alpha_P + \frac{b_2 - b_1}{c} \tag{46}$$

Combining with the optimal values of b_1 and b_2 in Lemma 2 and reorganizing completes the proof.

The principal's objective function is strictly concave in α_M, so that it is maximized at a unique point on $[\underline{\alpha_M}, \overline{\alpha_M}]$, which is described by the FOC above if an interior solution exists in $[\underline{\alpha_M}, \overline{\alpha_M}]$, and by a corner solution otherwise. ■

Proof of Corollary 1: Substituting $\alpha_M^\star$ from Proposition 1 into equations (10) and (11) gives equations (14) and (15). Comparative statics are immediate. ■

Proof of Proposition 2: As derived in the proof of Proposition 1, the principal's payoff after substituting the manager's allocation from Lemma 1 and the binding participation constraint is:

$$\Pi(\alpha_M) = 1 - \bar{U} + \frac{b_1 Q_1 + b_2 Q_2}{c} - \frac{(c+k)\Sigma}{2c^2} - \frac{k\Delta^2}{2} - \frac{k\Delta(b_1 - b_2)}{c}, \tag{47}$$

where $Q_i \equiv \rho_i \sigma_i^p \sigma_i^v$, $\Sigma \equiv b_1^2(\sigma_1^{p2}+1) + b_2^2(\sigma_2^{p2}+1) - 2b_1 b_2$, and $b_1(\Delta)$, $b_2(\Delta)$ are as in Lemma 2.

Since $b_1(\Delta)$ and $b_2(\Delta)$ maximize (47) for each Δ, the FOCs $\frac{\partial \Pi}{\partial b_i} = 0$ hold. Using the envelope theorem and $\frac{\partial \Delta}{\partial \alpha_M} = 1$:

$$\frac{d\Pi}{d\alpha_M} = \left.\frac{\partial \Pi}{\partial \Delta}\right|_{b_1(\Delta),\, b_2(\Delta)} = -k\Delta - \frac{k(b_1(\Delta) - b_2(\Delta))}{c}. \tag{48}$$

Substituting the incentive gap from Lemma 2 into (48):

$$\begin{aligned}
\frac{d\Pi}{d\alpha_M} &= -k\Delta - \frac{k\left[\rho_1 \sigma_1^p \sigma_1^v \sigma_2^{p2} - \rho_2 \sigma_2^p \sigma_2^v \sigma_1^{p2} - k(\sigma_1^{p2} + \sigma_2^{p2})\Delta\right]}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})} \\
&= -\frac{k}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})} \\
&\quad \times \left[\Big((c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}) - k(\sigma_1^{p2} + \sigma_2^{p2})\Big)\Delta + \rho_1 \sigma_1^p \sigma_1^v \sigma_2^{p2} - \rho_2 \sigma_2^p \sigma_2^v \sigma_1^{p2}\right].
\end{aligned} \tag{49}$$

The coefficient of Δ in the bracket simplifies:

$$\begin{aligned}
&(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2}) - k(\sigma_1^{p2} + \sigma_2^{p2}) \\
&\quad = (c+k)\sigma_1^{p2}\sigma_2^{p2} + c(\sigma_1^{p2} + \sigma_2^{p2}) = G,
\end{aligned}$$

where the last equality uses the definition $G \equiv (c+k)\sigma_1^{p2}\sigma_2^{p2} + c(\sigma_1^{p2} + \sigma_2^{p2})$. Letting $\mathcal{A} \equiv \rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} - \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}$, equation (49) becomes:

$$\frac{d\Pi}{d\alpha_M} = -\frac{k(G\Delta - \mathcal{A})}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})}. \tag{50}$$

This is affine and strictly decreasing in Δ, confirming that Π is strictly concave in α_M. Setting (50) to zero gives $\Delta^* = \mathcal{A}/G$, recovering Proposition 1.

Integrating (50) from $\alpha_M^0 = \alpha_P$ (corresponding to $\Delta = 0$) to α_M^* ($\Delta = \Delta^* = \mathcal{A}/G$):

$$\begin{aligned}\Pi^* - \Pi_0 &= \int_0^{\Delta^*} \frac{-k(G\tilde{\Delta} - \mathcal{A})}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})} d\tilde{\Delta} \\ &= \frac{k}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})}\left[\mathcal{A}\Delta^* - \frac{G(\Delta^*)^2}{2}\right]. \end{aligned} \tag{51}$$

Substituting $\Delta^* = \mathcal{A}/G$:

$$\mathcal{A}\Delta^* - \frac{G(\Delta^*)^2}{2} = \frac{\mathcal{A}^2}{G} - \frac{\mathcal{A}^2}{2G} = \frac{\mathcal{A}^2}{2G}.$$

Substituting into (51) and expanding $\mathcal{A}$:

$$\Pi^* - \Pi_0 = \frac{k\left(\rho_2\sigma_2^p\sigma_2^v\sigma_1^{p2} - \rho_1\sigma_1^p\sigma_1^v\sigma_2^{p2}\right)^2}{2(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})\,G} \;\geq\; 0,$$

which is equation (17). The gain is zero if and only if $\mathcal{A} = 0$, i.e. $\rho_1\sigma_1^v/\sigma_1^p = \rho_2\sigma_2^v/\sigma_2^p$, the condition for $\alpha_M^* = \alpha_P$ in Proposition 1.

From Lemma 2, the incentive rate on dimension 2 at the aligned benchmark ($\Delta = 0$) is:

$$b_2(0) = \frac{c\left[\rho_2\sigma_2^p\sigma_2^v(\sigma_1^{p2} + 1) + \rho_1\sigma_1^p\sigma_1^v\right]}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})}.$$

At the optimum ($\Delta = \Delta^* = \mathcal{A}/G$):

$$b_2(\Delta^*) = b_2(0) + \frac{ck\sigma_1^{p2}\Delta^*}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})} = b_2(0) + \frac{ck\sigma_1^{p2}\mathcal{A}}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})\,G}.$$

Therefore:

$$b_2(\Delta^*) - b_2(0) = \frac{ck\,\sigma_1^{p2}\,\mathcal{A}}{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})\,G}. \tag{52}$$

Solving (52) for $\mathcal{A}$ and substituting into equation (17):

$$\begin{aligned}\Pi^* - \Pi_0 &= \frac{k}{2(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})\,G} \cdot \frac{(c+k)^2(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})^2 G^2}{c^2 k^2 \sigma_1^{p4}} \left(b_2(\Delta^*) - b_2(0)\right)^2 \\ &= \frac{(c+k)(\sigma_1^{p2}\sigma_2^{p2} + \sigma_1^{p2} + \sigma_2^{p2})\,G}{2c^2 k\,\sigma_1^{p4}} \left(b_2(\Delta^*) - b_2(0)\right)^2,\end{aligned}$$

which is equation (18). ■

Proof of Lemma 3: The surplus when the contract is chosen optimally is:

$$Y(\rho_1, \rho_2; \alpha_M) = 1 + \frac{b_1\rho_1 + b_2\rho_2}{c} - \frac{k\Delta^2}{2} - \Delta\frac{k(b_1 - b_2)}{c} - (b_1^2 + b_2^2 - b_1 b_2)\frac{c+k}{c^2} \tag{53}$$

where $\Delta \equiv \alpha_M - \alpha_P$, and with optimal incentive slopes:

$$b_1(\rho_1, \rho_2, \alpha_M) = \frac{c(-k\Delta + 2\rho_1 + \rho_2)}{3(c+k)} \tag{54}$$

$$b_2(\rho_1, \rho_2, \alpha_M) = \frac{c(k\Delta + \rho_1 + 2\rho_2)}{3(c+k)} \tag{55}$$

Substituting the optimal b_1, b_2 into equation (53) and simplifying yields the surplus in the form $Y = 1 + \frac{\frac{1}{3}(\rho_1^2+\rho_1\rho_2+\rho_2^2)+\frac{k}{3}(\rho_2-\rho_1)\Delta-\frac{k(3c+k)}{6}\Delta^2}{c+k}$. Taking the derivative with respect to α_M (using $\frac{\partial\Delta}{\partial\alpha_M} = 1$):

$$\frac{\partial Y}{\partial \alpha_M} = \frac{k(\rho_2 - \rho_1)}{3(c+k)} - \frac{k(3c+k)}{3(c+k)}(\alpha_M - \alpha_P). \tag{56}$$

Now taking the derivative with respect to ρ_1:

$$\frac{\partial^2 Y}{\partial\rho_1\partial\alpha_M} = -\frac{k}{3(c+k)} < 0 \tag{57}$$

This establishes submodularity. ■

Proof of Proposition 4: The optimal incentive slopes are in equations (54) and (55), with $\Delta \equiv \alpha_M - \alpha_P$, and the joint surplus of a match is in equation (53). Substituting the optimal b_1, b_2 and simplifying yields:

$$Y(\rho_1, \rho_2; \alpha_M) = 1 + \frac{\frac{1}{3}(\rho_1^2 + \rho_1\rho_2 + \rho_2^2) + \frac{k}{3}(\rho_2 - \rho_1)\Delta - \frac{k(3c+k)}{6}\Delta^2}{c+k}.$$

Differentiating with respect to α_M (with $\frac{\partial \Delta}{\partial \alpha_M} = 1$):

$$\frac{\partial Y}{\partial \alpha_M} = \frac{k(\rho_2 - \rho_1)}{3(c+k)} - \frac{k(3c+k)}{3(c+k)}(\alpha_M - \alpha_P). \tag{58}$$

In a transferable-utility matching equilibrium, the marginal rent equals the marginal surplus generated by the agent (Terviö (2008)): when agent type α_M is matched with firm type $\rho_1(\alpha_M)$, we have $U'(\alpha_M) = \frac{\partial Y(\rho_1(\alpha_M), \rho_2; \alpha_M)}{\partial \alpha_M}$ (the partial with respect to agent type, evaluated at the matched firm). So:

$$U'(\alpha_M) = \frac{k}{3(c+k)} \left[\rho_2 - \rho_1(\alpha_M) - (3c+k)(\alpha_M - \alpha_P)\right]. \tag{59}$$

The rent schedule follows from the fundamental theorem of calculus: $U(\alpha_M) = U(\underline{\alpha_M}) + \int_{\underline{\alpha_M}}^{\alpha_M} U'(x)dx$. ■

Proof of Corollary 2: $U'(\alpha_M)$ is given in equation (59). Under uniform distributions, $F_{\alpha_M}(\alpha_M) = \frac{\alpha_M - \underline{\alpha}_M}{\overline{\alpha}_M - \underline{\alpha}_M}$ and $F_\rho^{-1}(q) = \underline{\rho} + q(\bar{\rho} - \underline{\rho})$, so that:

$$\rho_1(\alpha_M) = \overline{\rho} - \frac{\alpha_M - \underline{\alpha_M}}{\overline{\alpha_M} - \underline{\alpha_M}}(\bar{\rho} - \underline{\rho}), \qquad \rho_1'(\alpha_M) = -\frac{\bar{\rho} - \underline{\rho}}{\overline{\alpha_M} - \underline{\alpha_M}} \equiv -m < 0.$$

Therefore:

$$U''(\alpha_M) = \frac{k}{3(c+k)}\big[-\rho_1'(\alpha_M) - (3c+k)\big] = \frac{k}{3(c+k)}\big(m - (3c+k)\big).$$

If $m < 3c + k$, then $U''(\alpha_M) < 0$ for all α_M, so $U(\alpha_M)$ is strictly concave and single-peaked. In addition, the two endpoint conditions imply:

$$U'(\underline{\alpha_M}) > 0 \qquad \text{and} \qquad U'(\overline{\alpha_M}) < 0.$$

The type $\widehat{\alpha}_M$ is therefore interior and given by the FOC $U'(\widehat{\alpha}_M) = 0$:

$$\rho_2 - \rho_1(\widehat{\alpha}_M) - (3c+k)(\widehat{\alpha}_M - \alpha_P) = 0 \quad \Leftrightarrow \quad \widehat{\alpha}_M = \alpha_P + \frac{\rho_2 - \rho_1(\widehat{\alpha}_M)}{3c+k}.$$

Substituting $\rho_1(\widehat{\alpha}_M) = \overline{\rho} - m(\widehat{\alpha}_M - \underline{\alpha_M})$ yields :

$$\widehat{\alpha}_M = \alpha_P + \frac{\rho_2 - \overline{\rho} + m(\alpha_P - \underline{\alpha_M})}{(3c+k) - m}.$$

So part (i) holds.

For part (ii), $\widehat{\alpha}_M > \alpha_P$ if and only if $\frac{\rho_2 - \rho_1(\widehat{\alpha}_M)}{3c+k} > 0$, i.e., $\rho_2 > \rho_1(\widehat{\alpha}_M)$. ■

Proof of Proposition 6: Let the payoff to an agent of type α_M who induces belief $\hat{\alpha}$ by sending signal $x(\hat{\alpha})$ be:

$$W(\hat{\alpha}; \alpha_M) \equiv U(\hat{\alpha}) + \overline{U} + (b_1(\hat{\alpha}) - b_2(\hat{\alpha}))(\alpha_M - \hat{\alpha}) - C(x(\hat{\alpha}), \alpha_M) \tag{60}$$

Using Proposition 4 and Lemma 2,

$$U'(\alpha_M) - (b_1(\alpha_M) - b_2(\alpha_M)) = \frac{1}{3}\left(\rho_2 - \rho_1(\alpha_M) - k(\alpha_M - \alpha_P)\right) \tag{61}$$

The FOC $\frac{\partial W}{\partial \hat{\alpha}} = 0$ evaluated at $\hat{\alpha} = \alpha_M$ is:

$$\gamma(x(\alpha_M) - \alpha_M)x'(\alpha_M) = U'(\alpha_M) - (b_1(\alpha_M) - b_2(\alpha_M)) \tag{62}$$

Combining equations (61) and (62) gives (30).

Write $z(\alpha_M) \equiv x(\alpha_M) - \alpha_M$ and $R(\alpha_M) \equiv \frac{1}{3}\left(\rho_2 - \rho_1(\alpha_M) - k(\alpha_M - \alpha_P)\right)$, so that (30) reads:

$$\gamma\, z(\alpha_M)\left(1 + z'(\alpha_M)\right) = R(\alpha_M), \tag{63}$$

with $R'(\alpha_M) = \frac{1}{3}\left(m(\alpha_M) - k\right)$ and $m(\alpha_M) \equiv -\rho_1'(\alpha_M) > 0$. Equation (63) is regular wherever $z \neq 0$ and singular at a pivot type α_0 where $R(\alpha_0) = 0$. At this type, we impose the condition for an undistorted signal: $z(\alpha_0) = 0$.

A continuously differentiable solution has slope $\zeta \equiv z'(\alpha_0)$ at the pivot type, which solves:

$$\gamma\,\zeta(1 + \zeta) = R'(\alpha_0). \tag{64}$$

The roots are real if $\gamma > \frac{4}{3}\left(k - m(\alpha_0)\right)$, which is satisfied for γ sufficiently large. The larger root,

$$\zeta = \frac{-1 + \sqrt{1 + \frac{4R'(\alpha_0)}{\gamma}}}{2}, \qquad x'(\alpha_0) = 1 + \zeta = \frac{1 + \sqrt{1 + \frac{4R'(\alpha_0)}{\gamma}}}{2} > \frac{1}{2},$$

is the relevant root. If $m(\alpha_0) > k$, the other root gives $x' < 0$, so the larger root gives the only increasing schedule. If instead $k > m(\alpha_0)$, both roots give increasing schedules, and the divinity refinement selects the larger root, which involves less distortion.

Away from the pivot type, $z \neq 0$, so $z' = \frac{R(\alpha_M)}{\gamma z} - 1$ is Lipschitz and the solution that satisfies $z(\alpha_0) = 0$ extends uniquely to $[\underline{\alpha_M}, \overline{\alpha_M}]$ (Mailath (1987)). For γ large enough, $z = O(\frac{1}{\gamma})$ and $x'(\alpha_M) = 1 + O(\frac{1}{\gamma}) > \frac{1}{2}$ for all types, so $x(\cdot)$ is strictly increasing and beliefs $\beta(\hat{\alpha}) = x^{-1}(\hat{\alpha})$ are well-defined. If R has no zero on $[\underline{\alpha_M}, \overline{\alpha_M}]$, R has a constant sign, and the condition $z = 0$ is

imposed at $\underline{\alpha_M}$ if $R > 0$ and at $\overline{\alpha_M}$ if $R < 0$, with one-sided incentive compatibility at this type.

The cross-derivative of W is:

$$\frac{\partial^2 W}{\partial \hat{\alpha} \partial \alpha_M} = b_1'(\hat{\alpha}) - b_2'(\hat{\alpha}) + \gamma x'(\hat{\alpha}).$$

On the RHS, $b_1'(\hat{\alpha}) - b_2'(\hat{\alpha}) = -\frac{c\left(2k - \rho_1'(\hat{\alpha})\right)}{3(c+k)} < 0$ depends only on $\hat{\alpha}$. For γ large enough, the sum on the RHS is strictly positive, since $x'(\hat{\alpha}) = 1 + O(\frac{1}{\gamma}) > \frac{1}{2}$ and γ is higher than twice the maximum of $b_2'(\alpha_M) - b_1'(\alpha_M)$. The FOC $W_{\hat{\alpha}}(s;s) = 0$ holds for every interior type s by (30), so strict supermodularity gives $W_{\hat{\alpha}}(\hat{\alpha}; \alpha_M) < W_{\hat{\alpha}}(\hat{\alpha}; \hat{\alpha}) = 0$ for $\hat{\alpha} > \alpha_M$ and $W_{\hat{\alpha}}(\hat{\alpha}; \alpha_M) > 0$ for $\hat{\alpha} < \alpha_M$. It follows that $W(\cdot\,; \alpha_M)$ is single-peaked at $\hat{\alpha} = \alpha_M$, which establishes global incentive compatibility. The second-order condition holds, since $W_{\hat{\alpha}\hat{\alpha}}(s;s) = -W_{\hat{\alpha}\alpha_M}(s;s) < 0$.

For any out-of-equilibrium signal x, the divinity criterion places belief $\beta(x)$ on the type(s) for which the set of firm responses that would make sending x profitable is largest. Since the signaling cost $C(x, \alpha_M)$ satisfies single crossing ($\frac{\partial^2 C}{\partial x \partial \alpha_M} = -\gamma < 0$), this set is ordered monotonically in type, so $\beta(x)$ is monotone and the refinement selects the least-cost (Riley) separating schedule $x(\alpha_M)$, ruling out separating schedules with greater distortion (Ramey (1996)). Equation (31) follows by subtracting the signaling cost from the rent schedule of Proposition 4 and adjusting fixed wages so the participation constraint of the type with the lowest net utility binds. Since the signaling cost is sunk at the matching stage, this change involves a uniform shift in fixed wages that does not affect incentives. ■

Proof of Corollary 3: Under uniform distributions $\rho_1(\alpha_M)$ is affine with constant slope $-m$, so the RHS of (30) is affine with slope $\frac{m-k}{3}$. The distortion $x(\alpha_M) - \alpha_M$ and this RHS are both affine. Using (32) and the boundary condition of Proposition 6, both vanish at α_0. Thus the distortion is a constant multiple of the RHS:

$$x(\alpha_M) - \alpha_M = \frac{1}{3\gamma(1+\lambda)}\big[\rho_2 - \rho_1(\alpha_M) - k(\alpha_M - \alpha_P)\big],$$

with slope λ, so $x'(\alpha_M) = 1 + \lambda$.

Substituting into (30) gives:

$$\gamma\lambda(1+\lambda) = \frac{m-k}{3}.$$

The root associated with the least costly signaling equilibrium is the λ in the Corollary.[11] A real root requires $\gamma \geq \frac{4(k-m)}{3}$, the second term in (33).

Since b_1, b_2 are affine in t, $b_2'(\alpha_M) - b_1'(\alpha_M) = \frac{c(m+2k)}{3(c+k)}$ is constant, and the supermodularity

[11] The other root, $\lambda_- = \frac{-1-\sqrt{1+\frac{4(m-k)}{3\gamma}}}{2} \leq -\frac{1}{2}$, gives $x' = 1 + \lambda_- \leq \frac{1}{2}$: it is either decreasing (when $m > k$) and not separating, or increasing (when $k > m$) but with strictly greater signaling cost, and is eliminated by the divinity refinement.

condition $\gamma(1+\lambda) > b_2'(\alpha_M) - b_1'(\alpha_M)$ is implied by the first term of (33), since $1+\lambda > \frac{1}{2}$ under (33). ■

Online Appendix
The Value of Non-Value-Maximizing Managers

A A Contractible Aggregate Proxy for Output

Our analysis assumes that total output is noncontractible. In some applications, a noisy contractible measure of output may be available, such as the stock price for a CEO. This Appendix extends the model by assuming that the principal can also contract on an aggregate measure $P_0(a, \boldsymbol{\epsilon}) = a\epsilon_1^0 + (1-a)\epsilon_2^0$, where $\mathbb{E}[\epsilon_i^0] = 1$, $var(\epsilon_i^0) = \sigma_i^{0\,2}$, ϵ_i^0 has a correlation $\rho_0 \in [0,1]$ with the output shock ϵ_i^v on the same dimension. As in the baseline model, shocks are independent across dimensions, i.e., $cov(\epsilon_1^0, \epsilon_2^0) = 0$ and $cov(\epsilon_i^0, \epsilon_j^v) = cov(\epsilon_i^0, \epsilon_j^p) = 0$ for $i \neq j$. Since P_0 weights the two dimensions symmetrically, it is a noisy measure of total output $V_1 + V_2$. The compensation contract is now: $W = S + b_0 P_0 + b_1 P_1 + b_2 P_2$.

Proposition 7 *(i) If the aggregate measure is perfect ($P_0 = V_1 + V_2$) and a congruent manager is available, the principal reaches the first-best outcome by appointing this congruent manager ($\alpha_M = \alpha_P$) with $b_0 = \frac{c}{c+k}$ and $b_1 = b_2 = 0$.*
(ii) If the aggregate measure is imperfect, with shocks uncorrelated with those of the dimension-specific measures, the optimal incentives b_1 and b_2 and managerial bias α_M are as in Lemma 2 and Proposition 1.

Since the aggregate measure incorporates the two dimensions symmetrically, on average it does not shift the allocation that the manager chooses: for any correlation structure between P_0 and the dimension-specific measures, with $\alpha_M = \alpha_M^\star$ we have $\alpha_M - \alpha_P = \frac{b_2 - b_1}{c}$, so a measure P_0 of the overall objective is not a substitute for a biased manager.[12] This shows that our results are robust to a setting in which the principal's objective is directly measured, as long as this measure is imperfect.

Proof of Proposition 7: The manager's FOC gives $a = \alpha_M + \frac{u}{c}$ with $u \equiv b_0(\epsilon_1^0 - \epsilon_2^0) + b_1\epsilon_1^p - b_2\epsilon_2^p$. Since $\mathbb{E}[\epsilon_i^0] = 1$, b_0 cancels from the mean of u: $\mathbb{E}[u] = b_1 - b_2$. Following the same steps as in the proof of Proposition 1, the principal's objective is:

$$\Pi = 1 - \bar{U} + \frac{cov(u, \epsilon_1^v - \epsilon_2^v)}{c} - \frac{k}{2}\left(\Delta + \frac{\mathbb{E}[u]}{c}\right)^2 - \frac{\mathbb{E}[u]^2}{2c} - \frac{(c+k)var(u)}{2c^2}, \tag{A.1}$$

[12] Since $\mathbb{E}[\epsilon_i^0] = 1$, the aggregate measure does not change the manager's expected allocation, which is $\mathbb{E}[a] = \alpha_M + \frac{b_1 - b_2}{c}$, so the principal's objective depends on α_M only via $\frac{k}{2}\left(\Delta + \frac{b_1 - b_2}{c}\right)^2$. An interior optimum for the manager's bias is therefore such that $\Delta = \frac{b_2 - b_1}{c}$, as in Corollary 1.

which reduces to (47) when $b_0 = 0$.

We first establish part (ii). With $cov(\epsilon_i^0, \epsilon_i^p) = 0$, a condition that requires $\rho_0^2 + \rho_i^2 \leq 1$ for existence of the joint distribution, we have:

$$cov(u, \epsilon_1^v - \epsilon_2^v) = b_0\rho_0(\sigma_1^0\sigma_1^v + \sigma_2^0\sigma_2^v) + b_1Q_1 + b_2Q_2, \qquad var(u) = b_0^2(\sigma_1^{0^2} + \sigma_2^{0^2}) + b_1^2\sigma_1^{p2} + b_2^2\sigma_2^{p2}.$$

Thus, b_0 enters (A.1) only via the additively separable term

$$\frac{b_0\rho_0(\sigma_1^0\sigma_1^v + \sigma_2^0\sigma_2^v)}{c} - \frac{(c+k)(\sigma_1^{0^2} + \sigma_2^{0^2})}{2c^2}b_0^2,$$

while the remaining terms are those of (47). Maximizing each part separately gives the b_1, b_2, α_M of Lemma 2 and Proposition 1, and $b_0 = \frac{c\rho_0(\sigma_1^0\sigma_1^v + \sigma_2^0\sigma_2^v)}{(c+k)(\sigma_1^{0^2} + \sigma_2^{0^2})} > 0$.

We now establish part (i). If $P_0 = V_1 + V_2$ then $\epsilon_i^0 = \epsilon_i^v$. With $b_1 = b_2 = 0$ and $b_0 = \frac{c}{c+k}$, the manager chooses $a = \alpha_M + \frac{\epsilon_1^v - \epsilon_2^v}{c+k}$. The first-best allocation, which maximizes total surplus state by state, is:

$$a^{FB} = \frac{c\alpha_M + k\alpha_P}{c+k} + \frac{\epsilon_1^v - \epsilon_2^v}{c+k}.$$

The first-best surplus depends on the manager's type only via the deviation cost $-\frac{ck\Delta^2}{2(c+k)}$, which is maximized at $\Delta = 0$. Thus, with $\alpha_M = \alpha_P$, this contract implements a^{FB} in every state and reaches the first-best outcome, and any bias strictly reduces this surplus. ■

B Microfoundation for linear contracts

This Appendix adapts an argument from Barron, Georgiadis, and Swinkels (2020) to our setting.

We extend the model to account for the ex-post riskiness of the performance measure and the manager's ability to affect it. Suppose that the performance measure is as defined in the main text but is also noisy ex-post, so that $P_1(a, \boldsymbol{\epsilon}) = a\epsilon_1^p + \varepsilon_1$ and $P_2(a, \boldsymbol{\epsilon}) = (1-a)\epsilon_2^p + \varepsilon_2$ where ε_1 and ε_2 are random variables with a mean of zero and a variance of one independent from other random variables that are unobservable.[13] Suppose also that, conditional on a, the agent can implement arbitrary mean-preserving spreads and mean-preserving contractions of each performance measure distribution. Formally:

Assumption 1 *After choosing a resource allocation a, the agent can costlessly replace (P_1, P_2) by any jointly distributed $(\tilde{P}_1, \tilde{P}_2)$ such that, for $i = 1, 2$:*

$$\mathbb{E}[\tilde{P}_i|a] = \mathbb{E}[P_i|a].$$

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

Claim 1 *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 1: 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:

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

Since the point mass at (μ_1, μ_2) is itself an admissible manipulation, robustness requires

$$\mathbb{E}[\phi(\tilde{P}_1, \tilde{P}_2)|a] = \phi(\mu_1, \mu_2)$$

for every admissible $(\tilde{P}_1, \tilde{P}_2)$. Consider manipulations that hold the second measure at its mean, $\tilde{P}_2 \equiv \mu_2$, and let $\tilde{P}_1$ take values x, y with probabilities $\lambda, 1-\lambda$ chosen so that $\lambda x + (1-\lambda)y = \mu_1$.

[13] With universal risk neutrality, the presence of this pure risk does not change any of the relevant outcomes of the model.

Such a manipulation is admissible, so robustness gives

$$\lambda\,\phi(x,\mu_2)+(1-\lambda)\,\phi(y,\mu_2)=\phi(\mu_1,\mu_2)\quad\text{for all } x,y \text{ and } \lambda\in(0,1) \text{ with } \lambda x+(1-\lambda)y=\mu_1.$$

Equivalently, for all $x<\mu_1<y$ the points $(x,\phi(x,\mu_2))$, $(\mu_1,\phi(\mu_1,\mu_2))$ and $(y,\phi(y,\mu_2))$ are collinear, so that $p_1\mapsto\phi(p_1,\mu_2)$ is affine. Since $\mathbb{E}[\epsilon_i^p]=1$ gives $\mu_1=a$ and $\mu_2=1-a$, letting a vary shows that $\phi(\cdot,p_2)$ is affine for every p_2 in an interval; by the symmetric argument $\phi(p_1,\cdot)$ is affine. A function that is affine in each argument can be written as:

$$\phi(p_1,p_2)=A+Bp_1+Cp_2+Dp_1p_2$$

for some constants A,B,C,D.

We now show by contradiction that a robust contract is such that $D=0$. Suppose that $D\neq 0$. Fix the resource allocation a, and recall that $\mu_1=\mathbb{E}[P_1|a]$ and $\mu_2=\mathbb{E}[P_2|a]$. Pick any $\delta_1,\delta_2>0$ and consider two manipulations: (a) $(\tilde{P}_1,\tilde{P}_2)$ takes the values $(\mu_1+\delta_1,\mu_2+\delta_2)$ and $(\mu_1-\delta_1,\mu_2-\delta_2)$, each with probability $\frac{1}{2}$; (b) $(\tilde{P}_1,\tilde{P}_2)$ takes the values $(\mu_1+\delta_1,\mu_2-\delta_2)$ and $(\mu_1-\delta_1,\mu_2+\delta_2)$, each with probability $\frac{1}{2}$.

Both have the same marginal distributions: $\tilde{P}_i\in\{\mu_i-\delta_i,\mu_i+\delta_i\}$ each with probability $\frac{1}{2}$, so $\mathbb{E}[\tilde{P}_i|a]=\mu_i$ for $i=1,2$ and both satisfy the constraints in Assumption 1. However:

$$\mathbb{E}[\tilde{P}_1\tilde{P}_2|a]=\mu_1\mu_2+\delta_1\delta_2\ \text{ under (a)},\qquad \mathbb{E}[\tilde{P}_1\tilde{P}_2|a]=\mu_1\mu_2-\delta_1\delta_2\ \text{ under (b)}.$$

Since $\mathbb{E}[\phi(\tilde{P}_1,\tilde{P}_2)|a]=A+B\mu_1+C\mu_2+D\mathbb{E}[\tilde{P}_1\tilde{P}_2|a]$, the expected payment differs across (a) and (b) by $2D\delta_1\delta_2$, which is nonzero with $D\neq 0$, and these manipulations do not change the marginal means. This contradicts robustness, so $D=0$.

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. ■

References

Banks, J. S. and Sobel, J. (1987). Equilibrium selection in signaling games. *Econometrica* 55(3), 647–661.

Mailath, G.J. (1987), Incentive Compatibility in Signaling Games with a Continuum of

Types. *Econometrica* 55(6), 1349–1365

Ramey, G. (1996). D1 signaling equilibria with multiple signals and a continuum of types. *Journal of Economic Theory* 69(2), 508–531.