Journal of Accounting Research
Vol. 33 Supplement 1995
Printed in U.S.A.

Divisional versus Company-Wide Focus: The Trade-Off between Allocation of Managerial Attention and Screening of Talent

MASAKO N. DARROUGH* AND NAHUM D. MELUMAD†

1. *Introduction*

In this paper, we analyze why managers are sometimes induced to maximize local objectives rather than adopt a company-wide perspective. American managers are sometimes criticized in the popular press for adopting a narrow focus, for example, for concentrating on maximizing divisional performance measures rather than pursuing company-wide objectives.[1] This concentration, however, is likely to be an optimal response

* University of California, Davis; †Columbia University. We thank an anonymous referee, Oren Fuerst, Steven Hansen, David Larcker, Michael Maher, Nandu Nagarajan, Paul Newman, Vincent Richman, Reed Smith, Guy Weyns, Mark Wolfson, Amir Ziv, and the participants of the seminars held at UC–Berkeley, Ohio State, Rutgers, Texas at Austin, Washington University, and the 1995 *JAR* Conference for helpful comments and suggestions.

[1] Cases in which division managers fail to internalize corporate objectives have been documented extensively by both the popular press and teaching cases. Examples include Chemical Bank (Bitetti and Merchant [1983]), Del Norte Company (Sahlman and Barrett [1976]), General Motors (Spaulding [1991]), General Motors versus Electronic Data Systems (*Wall Street Journal* [December 1, 1986]), Polaroid Corporation (Dauch [1992]), and Procter and Gamble (Bartlett [1983]). In the same vein, Ohmae [1989] is critical of American corporations for the lack of "companyism" and Kanter [1982] advocates the importance of collaborative innovation. The term "narrow focus" is also used to describe suboptimal management practices in other contexts, such as short-term orientation (Jacobs [1991] and Merchant [1989]), lack of global strategy (Hamel and Prahalad [1985]), and overemphasis on cutting costs for productivity improvement (Blaxil and Hout [1991]). We discuss an application of our model and its results to the problem of short-term perspective in section 5.

to existing reward structures. In fact, Fox [1979] reported that in nearly 30% of the companies surveyed by the Conference Board, the compensation of division managers was completely independent of corporate-wide performance measures. Even when compensation did depend on both divisional and corporate performance, a high percentage of managers' bonus usually was attributable to their divisional performance. A similar pattern has been reported more recently by Merchant [1989] and Sibson [1990; 1991].

A narrow managerial focus is especially perplexing in view of the argument that at top managerial levels, effort aversion is often unlikely to be a serious concern. Top-level managers are typically described as hard working, highly motivated, and driven by "intrinsic" factors such as the desire for achievement.[2] Why would such managers not simply be asked to do what is best for their companies? In particular, why are division managers often induced to focus on maximizing divisional performance measures?

In this paper, we focus on the trade-off between allocation of managerial attention and screening of talent and show that it alone can explain the above puzzle. We assume that top managers are not effort averse but differ in ability (referred to as "type"). When managerial type differs and is unobservable, the principal has to design a compensation scheme that possibly screens managerial types and induces optimal allocation of effort. Although effort is costless, managers care about its allocation, because the effort allocation alters their payoffs by changing the likelihood of different outcomes. We identify circumstances under which the principal finds it optimal to sacrifice desired effort allocation for better screening of managerial type.

In our model, a risk-averse, but not effort-averse, manager is hired to manage a division in a multidivisional organization. The principal cannot observe a manager's type. To differentiate our model from a traditional agency model that emphasizes costly effort, we refer to managerial input as "managerial attention," as in Geanakoplos and Milgrom [1991]. Managerial attention is potentially productive in both the manager's own division and in other divisions. Time spent in other divisions, effort spent to coordinate activities among divisions, and decisions made that take into account their impact on the overall organization are just a few examples of managerial attention allocated to other divisions. (Hereafter, we refer to the manager's own division and the rest of the organization as Divisions 1 and 2, respectively.) We focus on the hiring of one manager for one division so that we can isolate the trade-off between screening and allocation of managerial attention. As we argue in section

[2] See, for example, Hackman and Oldham [1980], Jennergren [1980], Scitovsky [1986], Holmström and Milgrom [1987], and Hellriegel, Socum, and Woodman [1989].

6, hiring two managers for both divisions is a simple extension of the single-division case.

We show that although it is sometimes optimal to offer a contract that attracts only the most talented managers, in other cases, excessive screening costs preclude this practice. In such cases, the principal sometimes finds it optimal to hire any type of manager by offering a menu of contracts. In yet another scenario, better managers are simply too expensive; then only less talented managers are hired. We also show that in a second-best contract, better managers are always exposed to risk, even though the principal is risk neutral and managers are not effort averse, while the worst type of managers are never exposed to risk. We show that pay-to-performance sensitivity need not be high to accomplish efficient screening and managerial attention allocation. We then establish that under certain circumstances, the principal will induce *some* managers to devote all managerial attention to their own divisions.

We also examine alternative performance measures and identify their effect on the induced allocation of managerial attention. Because disaggregate company-wide measures are always (weakly) better than other measures (as long as the cost of more detailed information systems is ignored), the question is whether division-based or aggregate company-wide measures are ever *sufficient* for disaggregate measures. This question is of interest because using detailed performance measures often entails substantial transactions costs.

While most of our analysis examines divisional focus on the part of division managers, our model also provides a framework for studying other forms of goal incongruence problems. For example, the alleged short-sightedness of American managers has been remarked in the business press. By extending our analysis to a multiperiod framework, we address, in section 5, the question of whether managers should be induced to focus on the short term or encouraged to engage in long-term planning. Our model suggests that it might be too costly to induce managerial behavior that benefits the firm after a manager's departure.

In section 2, we present a brief review of the related literature. We then present our model and examine the problem of type screening in section 3. Section 4 discusses alternative performance measures and analyzes the issue of optimal assignment of managerial attention. Section 5 presents the implications of our model for a multiperiod framework, followed by a general discussion in section 6. Appendix A contains proofs.

2. *Relation to Earlier Literature*

The allocation of managerial input to different activities has been investigated by several recent papers. Gibbs [1990] studies job assignments versus promotion, Itoh [1991] studies specialization versus teamwork, and Amershi and Datar [1993] study routine versus thinking effort. Our paper is closest in spirit to Holmström and Milgrom [1991], who are also concerned with the costless allocation of managerial input to different tasks. The new dimension in our study is the interaction between hiring policy and the assignment of managerial attention. Holmström and Milgrom, who focus on the allocation of effort (or time) to different aspects of outputs (e.g., timely completion versus quality of work in household contracting), argue that if the principal focuses on only one aspect of performance (because it is easier to measure), the manager will be induced to misallocate effort. Therefore, when performance is difficult to measure, hence difficult to motivate, a fixed compensation scheme might be optimal to motivate proper effort allocation. In contrast, outcome-independent compensation is optimal in our framework only when the optimal hiring policy involves hiring only the worst-type managers. Otherwise, the optimal contract must involve outcome-contingent components.[3] Another distinguishing aspect of our analysis, relative to that of Holmström and Milgrom, is the explicit examination of the effect of the manager's reservation utility on the optimal hiring policy.

In our setting, the more talented managers receive only their reservation utility, while the less talented managers, if hired, expect to receive rents. A similar feature appears in other screening models (e.g., Stiglitz [1977] and Lazear [1986]) and agency models (e.g., Sappington [1983]).

Jensen and Murphy [1990] found a seemingly small empirical relation between *CEO* compensation and market (or accounting) performance measures. They interpret their results as being inconsistent with predictions from agency models of optimal contracting. Although we focus on division managers and not on *CEO*s, the issue of screening applies to all levels of managers. Our model offers a conjecture that the pay-to-performance connection is motivated by the desire to attract better managers and to compensate them according to their type. As we argue in connection with Observation 2, effective screening may be achievable through a seemingly small pay-to-performance relation. If so, Jensen and Murphy's observations are not inconsistent with optimal contracting theory.

3. The Model

Consider a company consisting of Divisions 1 and 2. A risk-neutral principal offers a menu of contracts to hire a risk-averse manager mainly to

[3] Outcome-contingent compensation has also been studied in the agency literature. For example, Banker and Datar [1989] model a manager who expends effort that produces multiple signals, derive conditions under which linear aggregation is optimal, and identify the optimal weights. In Magee [1988], a risk- and effort-averse manager allocates his effort between two outputs. Although the model involves moral hazard, the principal and manager have no conflict about relative effort allocation. Thus, the manager's compensation depends on the aggregate output.

run Division 1.[4] There are two types of managers—"good" (type G) and "bad" (type B); type G managers are more productive than type B in their own division and at least as productive in the other division.[5] A fraction ρ, $0 < \rho < 1$, of the managerial population is ex ante of type G. The manager's type is known only by the manager.[6] While managerial attention is productive in both divisions, the manager must devote a given portion of his managerial attention to his own division, with the remainder spent in either division. We let $e_t \in \{0, 1\}$ be the portion of the remaining managerial attention a manager of type t spends in Division 2. When $e_t = 0$, type t devotes himself to Division 1 only and when $e_t = 1$, he works in two divisions.[7] An outcome in Division i can be high (H^i) or low (L^i), $i \in \{1,2\}$. Managers differ in their probability of attaining a high outcome in Division i. We denote $P_t^1(e_t)$ as the probability of H^1 and $P_t^2(e_t)$ as the probability of H^2 for type t, given e_t. A type G manager is more productive for any given allocation of managerial attention, i.e., $P_G^1(0) > P_B^1(0)$ and $P_G^1(1) > P_B^1(1)$ in Division 1, and $P_G^2(1) \geq P_B^2(1)$ in Division 2. Further, we assume $P_G^2(0) = P_B^2(0)$ to capture that the probability structure in the second division is independent of managerial type when no managerial attention is allocated to that division. When either manager allocates his attention to Division 2, the likelihood of a high outcome increases for Division 2 but decreases for Division 1, i.e, $P_t^1(0) > P_t^1(1)$ and $P_t^2(1) > P_t^2(0)$.

The trade-off between screening and allocation of managerial attention arises only when the first-best solution (i.e., the solution to the case of observable managerial attention and type) involves the assignment of some managerial attention to Division 2.[8] We therefore focus

---

[4] An alternative formulation, which we find less descriptive, would have the potential manager proposing a contract (to signal his type) to the principal. For a comparison of these two approaches, see Feltham and Hughes [1988] and Thakor [1991].

[5] The labels "good" and "bad" represent relative productivity, ignoring hiring costs. As we illustrate below, the principal sometimes prefers hiring "bad" managers to hiring "good" managers.

[6] We assume that managers possess private information about their type. Some private information is necessary for our results to hold. We note that having imperfect information would not simplify our model or change the essence of our results.

[7] While one would expect e_t to admit more values (possibly a continuum), we allow only two possible choices. This assumption is made to simplify the analysis. Although the assumption is restrictive, we expect the results to be qualitatively the same as long as the marginal benefit from allocating *MA* to Division 2 is positive in [0,1]. If marginal benefits decline, and screening costs increase over the domain, it is likely that optimal e_t will be below 1, which would amount to *not* fully exploiting externalities or synergy (i.e., "too much divisional focus").

[8] We are interested in studying settings for which a first-best solution involves a manager contributing to both divisions due to externalities and synergies. Naturally, there are situations in which this is not the case. However, in those situations, there is not much to discuss in terms of the trade-off between screening and allocation of managerial attention. Investigating the link between the compensation of business-unit managers and aggregate

our attention on settings where the marginal productivity of managerial attention is strictly higher in Division 2. Formally:

$$P_t^2(1) - P_t^2(0) > P_t^1(0) - P_t^1(1) \quad \forall \quad t \in \{G,B\}. \qquad (A1)$$

Let $\underset{\sim}{y} \equiv \{y_1, y_2, y_3, y_4\}$ be the vector of the four possible outcome configurations: L^1L^2, L^1H^2, H^1L^2, and H^1H^2. As will become apparent, the values of high and low outcomes play no role in the ensuing analysis. It is only the difference between H^i and L^i that matters. We further assume:

$$H^2 - L^2 = H^1 - L^1 \equiv \Delta. \qquad (A2)$$

Assumption $(A2)$ implies overlapping outcomes in the sense that the realization of aggregate outcome cannot always reveal individual division outcome. We make this assumption, first, to distinguish between aggregate and disaggregate performance measures and, second, to approximate cases of continuous output where, naturally, all outcome combinations are overlapping.

3.1 THE PRINCIPAL'S PROBLEM

The principal maximizes expected net profit by offering a menu of contracts (or possibly a single contract). We denote the contract chosen by type t in equilibrium by $\underset{\sim}{s}^t \equiv \{s_1^t, s_2^t, s_3^t, s_4^t\}$, where s_i^t is the compensation when y_i is realized. The compensation is said to be *outcome independent*, if $s_1^t = s_2^t = s_3^t = s_4^t$, and *outcome contingent* otherwise. The additional amount required to compensate a manager for the risk associated with the contingent compensation is referred to as the *risk premium.*

The manager maximizes his expected utility, $EU_t(e_t; \underset{\sim}{s}^t)$, by choosing an allocation of managerial attention. (Recall that the manager is assumed not to be effort averse.) The principal's maximization problem can be presented as a choice among the following three submaximization programs:

($P1$): Maximize expected net profit by hiring only B.
($P2$): Maximize expected net profit by hiring either B or G.
($P3$): Maximize expected net profit by hiring only G.

Clearly, randomization among programs cannot be beneficial. Let θ denote the probability of hiring a manager of type G. Then, $\theta = 0$, ρ, or 1, corresponding to ($P1$), ($P2$), and ($P3$), respectively. The principal's maximization problem then amounts to:

$$\operatorname{Max}_{\underset{\sim}{s}^t, e_t, \theta} \sum_{i=1}^{4} \left\{ \theta P_G(y_i | e_G) \left\{ y_i - s_i^G \right\} + (1-\theta) P_B(y_i | e_B) \left\{ y_i - s_i^B \right\} \right\} \qquad (1)$$

$$\text{subject to:} \quad EU_t(e_t; \underset{\sim}{s}^t) \geq \bar{U}_t, \qquad (2)$$

---

performance measures, Bushman, Indjejikian, and Smith [1995] find that the association increases with the degree of interdependencies.

$$EU_t(e_t;\, \underset{\sim}{s}^t) \geq EU_t(e'_t;\, \underset{\sim}{s}^{t'}),\ \forall e', t', \tag{3}$$

$$\theta \in \{0, \rho, 1\}, \tag{4}$$

where $EU_t(e_t; \underset{\sim}{s}^t)$ is the expected utility of a type t manager over his probability distribution, given the chosen e_t and $\underset{\sim}{s}^t$, and expectation is taken over his probability distribution.[9] $P_t(y_i|e_t)$ denotes the probability of outcome y_i for type t, given e_t; for example, $P_t(y_1|e_t) \equiv (1 - P_t^1(e_t))(1 - P_t^2(e_t))$. Note that $\theta = 0$ implies $t = B$, $\theta = \rho$ implies $t \in \{B, G\}$ and $\theta = 1$ implies $t = G$.

Equation (2) represents the manager's individual rationality constraint, where $\bar{U}_t$ is the reservation utility of a manager of type t, $t \in \{B, G\}$. We assume $\bar{U}_G \geq \bar{U}_B$. This inequality is likely to be strict, for example, when talent is transferable and managers' alternatives are self-employment. When the alternative is employment with other firms, one equilibrium when talent is transferable would involve other firms offering contingent contracts, again resulting in different reservation utilities for the different types.[10] The next two constraints represent incentive compatibility with respect to the allocation of managerial attention and compensation selection. If the compensation scheme is type specific in equilibrium, we say that a separating menu of contracts is offered. Thus, in ($P2$), where either type manager might be hired, the contracts offered could be a menu rather than a single (pooling) contract.[11] In ($P1$), where only type B is hired, the principal screens out type G by offering a single contract $\underset{\sim}{s}^B$ such that $EU_G(e_G; \underset{\sim}{s}^B) \leq \bar{U}_G\ \forall\ e_G$; whereas in ($P3$), where only type G is hired, he screens out type B by a single contract $\underset{\sim}{s}^G$ such that $EU_B(e_B; \underset{\sim}{s}^G) \leq \bar{U}_B\ \forall\ e_B$.[12] Depending on the combination of parameter values, the principal might find it optimal to follow either ($P1$), ($P2$), or ($P3$). (See Observation 1 below.)

In all programs, the principal weighs the benefits against the costs of hiring a manager. The benefits depend on both the manager's type and the allocation of managerial attention. The cost depends on the level of reservation utility as well as the screening cost. It is convenient to define *residual expected output* for each type as $REY_t(e_t) \equiv EY_t(e_t) - W_t$, $t \in \{B, G\}$, where $EY_t(e_t) = \sum P_t(y_i|e_t)y_i$, and W_t is the outcome-independent compensation that provides the reservation level of utility, i.e., $W_t \equiv U^{-1}(\bar{U}_t)$. Recall that $EY_G(1) > EY_B(1)$ and $EY_G(0) > EY_B(0)$ by definition.

---

[9] Note $\underset{\sim}{s}^G$ and $\underset{\sim}{s}^B$ need not be identical in ($P1$), ($P2$), and ($P3$).

[10] Assuming different reservation utilities for different managers is common in the literature. See, for example, Calvo and Wellisz [1979], Weiss [1980], Ricart i Costa [1988], and Lewis and Sappington [1993].

[11] An implicit assumption made here is that the matching process of the principal with a manager is such that the posterior-type probability is equal to the prior probability.

[12] We assume that an indifferent manager will adopt the decision preferred by the principal. Alternatively, at an arbitrarily small cost, the principal can induce the manager to do so.

W_t represents the minimum cost (certainty equivalent) of hiring a manager of type t when type is observable, and therefore no risk is imposed. Let REY_t^* denote the value of $REY_t(e_t)$ at the second-best optimal e_t. We assume $REY_t^* \geq 0$. We now define:

DEFINITION. The first-best setting is one in which managerial type is observable. In this case, the principal hires the manager with a higher $REY_t(1)$, designates $e_t = 1$, and compensates the manager with a fixed wage of W_t.[13]

Note that in the first-best case, only one type is hired. In addition, for any type hired, $EU_t(e_t;\, \underline{s}^t) = \bar{U}_t$ and $\underline{s}^t = W_t$. In the second-best case, any of these conditions may not hold. If, in equilibrium, $EU_t(e_t;\, \underline{s}^t) > \bar{U}_t$, then the type t manager is said to receive an *informational rent.* If the second-best compensation is outcome contingent, the principal pays the manager a *risk premium.*

Before we can discuss the optimality of the alternative hiring policy, it is useful to determine whether the principal would ever offer a pooling contract in ($P2$). Proposition 1 establishes that he would not.

PROPOSITION 1. In ($P2$), where either type is hired, the optimal contract is a menu consisting of an outcome-independent contract for a type B manager and an outcome-contingent contract for a type G manager. The former receives an informational rent and the latter receives his reservation utility. In ($P3$), where only type G is hired, the optimal contract is outcome contingent, and the type G manager receives his reservation utility.

Proof. Proofs are presented in Appendix A.

Proposition 1 resembles the results of Stiglitz [1977] for the insurance market. There, a high-risk purchaser obtains complete insurance and achieves a utility level higher than his reservation, while a low-risk purchaser gets partial insurance and achieves utility equal to his reservation level. The insurance model is parallel to the case of a single division, so Proposition 1 generalizes Stiglitz's results (specifically, Properties 1–3) to the case of two divisions, with the added concern of inducing optimal allocation of managerial attention.

Since a type B manager receives an informational rent in ($P2$) and a type G manager is paid a risk premium in either ($P2$) or ($P3$), the principal's hiring policy depends on the expected output from each type *net* of the expected cost of compensation. The following observation provides sufficient (but not necessary) conditions in terms of exogenous parameters for each of the three hiring policies to become optimal in the second-best setting.

OBSERVATION 1. (1) It is optimal to hire only a type B manager if $REY_B(1) > REY_G(1)$. (2) It is optimal to hire either type if $REY_B(1) < REY_G(0)$, $REY_G(1) - REY_B(1)$ is not too large, and ρ is sufficiently large. (3) It is optimal to hire only a type G manager if $REY_G(0)$ is sufficiently large relative to $REY_B(1)$.

---

[13] Recall that assumption ($A1$) implies that first-best allocation of managerial attention is $e_t = 1$ for both types.

This observation is based on the following arguments. Recall that the principal can always instruct $e_B = 1$, since a type B manager is paid a fixed salary. When $REY_B(1) > \max_{e_G}\{REY_G(e_G)\}$, the principal cannot benefit from hiring type G. Note that this happens only if $\bar{U}_G > \bar{U}_B$, since $EY_G(1) > EY_B(1)$ and $EY_G(0) > EY_B(0)$. The principal can always discriminate against type G by offering a single outcome-independent contract W_B. When the productivity difference is not too large, hiring either type ($P2$) is optimal. The principal's trade-off is between productivity and screening, which includes both the risk premium (paid to type G) and the informational rent (paid to B). For the appropriate parameter values, it is not worthwhile for the principal to screen out either type. If $\min_{e_G}\{REY_G(e_G)\}$ is sufficiently larger than $REY_B(1)$ to cover the necessary cost of screening (i.e., the risk premium paid to type G), then ($P3$) is optimal. This risk is necessary to make the contract unattractive to type B. Note that $\min_{e_G}\{REY_G(e_G)\} = REY_G(0)$, $\max_{e_G}\{REY_G(e_G)\} = REY_G(1)$, and $REY_G(0) < REY_G(1)$.

Next, we consider the sensitivity of managerial pay to performance. Denote the optimal menu in ($P2$) by $\{w, \underset{\sim}{s}^G\}$. Then the expected risk premium paid to a type G manager is $\sum P_G(y_i|e_G)s_i^G - W_G$, where the vector $\underset{\sim}{s}^G$ is such that $\sum P_G(y_i|e_G)U(s_i^G) = \bar{U}_G$, while the expected informational rent paid to type B is $U^{-1}(EU_B(e_B; \underset{\sim}{s}^G)) - W_B$. Proposition 1 implies that, regardless of the proportion, ρ, of type G managers, the principal never finds it optimal to offer an outcome-independent contract that will be chosen by type G. However, as type G becomes more prevalent, it is preferable to reduce the performance sensitivity of the outcome-contingent contract. This is because as ρ increases, the principal can reduce expected compensation costs by reducing the risk premium even though the informational rent is increased. Clearly, other factors such as risk aversion affect the sensitivity of pay to performance. Similarly, larger differences in talent should require smaller risks to achieve screening. Basically, by reducing the contract's performance sensitivity, the principal trades off expected risk premium against expected information rent. In either ($P1$) or ($P3$), on the other hand, there is no informational rent, although in ($P3$), the type G manager receives a relatively large risk premium. The above discussion is summarized in the following observation.

OBSERVATION 2. In either ($P2$) or ($P3$), the sensitivity of type G managers' compensation to performance is inversely related to the degree of risk aversion and the difference in talent, but is positively related to the difference in reservation utility levels. Furthermore, in ($P2$), the sensitivity is inversely related to the prevalence of type G managers in the managerial population.

Although we allow $\bar{U}_G \geq \bar{U}_B$, the case of equal reservation utility levels is of limited interest, because:

OBSERVATION 3. If $\bar{U}_B = \bar{U}_G$, the principal can achieve performance arbitrarily close to that of the first-best setting. An optimal solution, however, does not exist.

In this case, the principal hires only the more productive type G ($P3$). To screen out type B, the contract cannot be outcome independent, but the risk imposed can be arbitrarily small. Our discussion henceforth will focus on the case of different reservation utility levels.

4. Alternative Performance Measures

In this section, we compare alternative managerial performance measures, based on their effectiveness in screening managerial talent and inducing allocation of managerial attention. Although not explicitly addressed in the last section, we now examine how different performance measures affect the choice of allocation of managerial attention. We consider three performance measures: division-based, aggregate company-wide, and disaggregate company-wide performance measures.[14] We identify the circumstances under which each measure is optimal. These measures lead to the following three types of compensation structure.[15]

DB: Division-based compensation: $s_1 = s_2$ and $s_3 = s_4$.
ACW: Aggregate company-wide compensation: $s_2 = s_3$, since $L^1 + H^2 = H^1 + L^2$ by assumption ($A2$).
DCW: Disaggregate company-wide compensation.

By definition, a disaggregate company-wide measure utilizes all the information contained in the outcomes of both divisions, while an aggregate company-wide measure cannot differentiate L^1H^2 from H^1L^2. Division-based and aggregate company-wide measures are special cases of disaggregate company-wide measures. For comparison purposes, we say that a contract is disaggregate company-wide only if it is neither division-based nor aggregate company-wide.

Part of the motivation for comparing these measures is based on an assumption of differential costs. We assume that an aggregate company-wide measure is the least expensive, since it is required for financial reporting and it avoids the problems associated with cost allocations, transfer pricing, etc. Thus, an aggregate measure would be preferred as long as it does not lead to a significant performance loss. The comparison of division-based and disaggregate company-wide measures is more subtle. In a multidivision firm, a division-based measure might be less costly to implement than a disaggregate company-wide measure because the latter requires detailed reporting and verification of other divisions' figures to each division manager. Thus, a division-based measure would

---

[14] In Sibson's [1990] survey, more than 80% of stand-alone business units and 50% of corporate-managed units measured performance solely on the results of their own unit. The remainder used a combination of unit and corporate results. See also *Employee Benefit Plan Review* [1992] and Sibson [1991].

[15] For notational simplicity, we omit the superscripts indicating type. As is clear from the previous analysis, a contingent contract is relevant only for type G.

be preferred to a disaggregate company-wide measure when the resulting performance loss is small.[16]

We consider first the case in which allocation of managerial attention is observable. By Assumption ($A1$), the optimal assignment of managerial attention is $e_t = 1$ for both types. The principal will instruct the manager to allocate his attention to Division 2, regardless of his type, and the manager will do so. Since allocation of managerial attention is costlessly enforceable, the only concern is efficient screening, which can be achieved by focusing on an outcome that is more informative about type difference. Recall that by assumption, $P_G^1(1) > P_B^1(1)$. Thus, Division 1's outcome is informative about type. Division 2's outcome, however, might not be informative, since we allow $P_G^2(1) = P_B^2(1)$. It is useful to define the following:

DEFINITION. Division 2's outcome is informative with respect to managerial type if $P_G^2(1) > P_B^2(1)$.[17]

Clearly, for the case of unobservable allocation of managerial attention, when Division 2's outcome is uninformative, there is no point in conditioning compensation on the realization of Division 2's outcome.[18] Doing so just increases risk, which must be compensated. When Division 2's outcome is informative, the principal needs to weigh the value of incremental information against the increased risk premium necessary to compensate the additional risk. Suppose most managers are type G; screening would be relatively unimportant. Suppose, in addition, that Division 2's outcome is not "much more" informative with respect to type than Division 1's outcome is. Then it would seem intuitive that, when allocation of managerial attention is observable, the principal should prefer divisional compensation to company-wide compensation. This intuition, however, is false, as shown in the next observation.

OBSERVATION 4. Suppose that it is (second-best) optimal to allocate managerial attention to Division 2, and that Division 2's outcome is informative with respect to managerial type. Disaggregate company-wide compensation then strictly dominates division-based compensation even if the allocation of managerial attention is observable.

---

[16] In our two-division setting, the cost of implementing a division-based measure is unlikely to be higher than that of a disaggregate company-wide measure, since the combination of division-based and aggregate company-wide measures is equivalent to a disaggregate company-wide measure.

[17] More precisely, we should have defined informativeness with respect to a given combination of $\{e_B, e_G\}$. Under our assumptions, however, Division 2's outcome is always informative for $\{0,1\}$ and $\{1,0\}$, and is never informative for $\{0,0\}$. Thus, the definition is meaningful only for the combination $\{1,1\}$.

[18] In this benchmark case, managerial talent is specific to the manager's own division. For example, in a firm with production and marketing divisions, a type G manager could be more productive in managing the production division than a type B manager, but the productivity of type G's managerial attention in coordinating with marketing could be indistinguishable from that of type B.

This observation is essentially Holmström's [1979] informativeness condition, adapted for our discrete setting. Thus, when Division 2's outcome is informative, the optimal compensation cannot be division-based even if allocation of managerial attention is observable. Screening is more efficiently attained by utilizing all information that discriminates between the two types even if Division 2's outcome is only slightly informative. When allocation of managerial attention is observable, the principal simply assigns $e_G = 1$ and compensates the manager with a disaggregate company-wide contract.

When allocation of managerial attention is unobservable, the principle has to design a contract that induces optimal allocation of managerial attention for the type G manager. (Since type B is paid an outcome-independent salary, there is no conflict in his assignment of managerial attention.) The screening contract takes into account the largest expected compensation type B could get, were he to adopt type G's contingent contract. If the principal wants to induce $e_G = 1$, the contract for type G must be a company-wide contract. A division-based contract cannot induce $e_G = 1$. Suppose, on the other hand, the principal wants to induce $e_G = 0$. Recall that $e_B = 1$. It might appear that the optimal contract for type G is also company-wide, since Division 2's outcome is now necessarily informative about type. However, as Proposition 2 below shows, it is never optimal for the principal to use a company-wide contract if he wants to induce $e_G = 1$, even though $e_B = 1$. In sum, in designing the optimal contract, the principal trades off the smaller screening cost under division-based compensation against the higher output under company-wide compensation. The principal compares the expected profit under (i) division-based compensation with $e_G = 0$ and (ii) disaggregate company-wide compensation with $e_G = 1$.

PROPOSITION 2. In ($P2$) and ($P3$), when the allocation of managerial attention is unobservable, the principal finds it optimal in some cases to adopt division-based compensation and induce type G managers to focus exclusively on their own divisions. For type B managers, on the other hand, the principal will offer outcome-independent compensation and induce these managers to allocate their managerial attention to both divisions.

While we can provide numerous examples showing the result of Proposition 2, it is difficult to derive conditions, expressed in terms of exogenous variables, under which Proposition 2 holds. The problem is that the optimal allocation of managerial attention is determined by the tradeoff between screening costs and productivity gains, but screening costs are a function of the optimal compensation schedule, which is endogenously determined. However, for the special case of square-root utility functions, we can show the following.

COROLLARY TO PROPOSITION 2. Consider the case of square-root utility functions and assume that Division 2's outcome is relatively uninformative about managerial type. Then, when outcome realizations are not too disparate (i.e., when $H - L$ is sufficiently small), divisional focus for a type G manager is optimal in either ($P2$) or ($P3$).

Proposition 2 and its corollary address the main question raised in the introduction: Why do companies induce managers to focus narrowly on their own division? The optimal allocation of managerial attention is determined by the trade-off between screening cost and productivity gain. If productivity gain is less than the increased screening costs (the additional risk premium required by type G and informational rent required by type B), then the principal is better off forgoing the desired allocation of managerial attention for type G managers. This trade-off is a function of the problem parameters.

We now compare the two company-wide measures. These measures are optimal if and only if the principal wants to induce allocation of managerial attention of $e_G = 1$.[19] Since an aggregate company-wide measure is a special case of a disaggregate company-wide measure, the former is weakly dominated by the latter. It is clear that when Division 2's outcome is informative, this dominance is strict. What is less obvious is the dominance relation when Division 2's outcome is uninformative. To minimize screening costs, the principal would want to focus on the type difference in Division 1 by compensating the type G manager with a higher reward for the high outcome there. But this compromises the principal's ability to induce the manager to allocate his managerial attention to Division 2. While the net effect is not immediately clear, the following proposition shows that the relation is, in general, strict.

PROPOSITION 3. In $(P2)$, a disaggregate company-wide measure strictly dominates an aggregate company-wide measure, *regardless* of whether Division 2's outcome is informative about managerial type, with the exception of a knife-edge case.

The above proposition implies that aggregate company-wide compensation is sufficient for disaggregate company-wide compensation only in the knife-edge case when the relative likelihood of achieving y_2 and y_3 is independent of type. Otherwise, the disaggregate company-wide measure strictly dominates, even when Division 2's outcome is uninformative about type. Clearly, as long as the cost of disaggregation is zero, there is no reason to consider aggregate company-wide performance measures. But, since disaggregate company-wide measures are likely to involve higher processing and verification costs, the principal needs to trade off the incremental cost of generating disaggregate company-wide measures against the benefit from disaggregate information.

5. *Implications for the Problem of Short-Term Perspective*

Short-term managerial perspective has been studied in a number of articles; see, for example, Narayanan [1985], Darrough [1987], Stein [1988], and Nagarajan, Sivaramakrishnan, and Sridhar [1995]. Extending our single-period model to a multiperiod framework allows us to consider

[19] Recall that, by definition, disaggregate company-wide compensation is different from division-based compensation.

such issues as whether managers should be compensated for the results of their decisions even after they leave the company. Deferred compensation is possible, for example, in the form of *vested* stock options that are exercisable only after retirement. In practice, the use of deferred-vested stock options is not common, even though such options would seem to be an effective vehicle for inducing managers to focus on long-term performance. Why are such options relatively rare? The adaptation of our model to a multiperiod framework suggests that, in some cases, it is too costly (in terms of screening costs) to induce managerial behavior that benefits the firm after the manager's departure.

Consider a two-period model. A manager is hired to run a firm for one period and then retires. The firm operates for another period, with another manager, and liquidates itself at the end of the second period. This setting captures the idea that managers have a shorter time horizon than the firm. The two periods parallel Divisions 1 and 2 in the previous model. (Investigating the conditions under which it is optimal to induce managers to remain with the firm for a longer period is beyond the scope of this model.)

The manager can make "short-term" decisions that focus only on the period of his job tenure, or he can make "long-term" decisions that take into consideration the postretirement period. A type G manager is more likely to generate a high outcome in the period of his tenure than a type B manager, for both kinds of decisions. By making long-term decisions, both managers can increase the probability of a high outcome after their retirements. The type difference in the postretirement outcome, however, would be much smaller than during the period of tenure because of the influence of other factors. Although both types of decisions produce positive expected net present values, long-term decisions produce higher expected net present values. Four possible outcomes exist as before: L^1L^2, L^1H^2, H^1L^2, and H^1H^2, where the superscripts now refer to the time periods. The manager has a time-additive utility function with a discount factor identical to that of the principal.

Compensation can be outcome-independent or outcome-contingent. If outcome-independent, compensation consists of an outcome-independent salary and possibly a pension (also outcome-independent). If outcome-contingent, compensation might depend only on the outcome of the job-tenure period or, alternatively, on the outcomes of the two periods. (Given our assumptions, it can be shown that it is never optimal to have outcome-contingent second-period compensation when the first-period compensation is outcome-independent.) We refer to the first as "short-term compensation" and the second as "long-term compensation." Note that these correspond to the division-based and company-wide compensation of the earlier section. Disaggregate long-term compensation differentiates L^1H^2 from H^1L^2, whereas aggregate long-term compensation does not.

Similar to Proposition 2, we can establish here a parallel result, stated without a formal proof.

PROPOSITION 2′. When the time perspective of managerial decisions is unobservable, the principal finds it optimal, in some cases, to use short-term, outcome-contingent compensation and induce a type G manager to focus only on his tenure period. For a type B manager, on the other hand, the principal will use long-term outcome-independent compensation and induce this manager to consider postretirement consequences.

The counterintuitive prediction that the good-type manager focuses on the short term is again a consequence of the tension between screening and productivity.[20] Recall that the label "good" refers to relative productivity, and not to adopting the optimal time perspective. Alternatively, if the productivity gain from doing so is sufficiently large, both types will be induced to take long-term perspectives. For the type B manager, incentive compatibility is never binding, since he is paid outcome-independent compensation.[21] For the type G manager, it might be too costly to induce long-term decisions. The reason, however, can be quite different for the case of observable decisions compared with the case of unobservable decisions.

Suppose the manager's decision is observable, as, for example, when the decision is a project choice. Suppose further that the marginal productivity of managerial decision in the postretirement period is the same for both types of managers. It is then optimal to use short-term (tenure-period) compensation for efficient screening and to instruct managers to make the optimal project choice.

In contrast, when the manager's decision is unobservable, the principal must trade off efficient decisions against the cost of screening, since inducing managers to take a long-term perspective requires long-term compensation. Proposition 2′ implies that it is sometimes optimal to induce a short-term focus to save screening costs. Thus, an apparent short-term focus could be due to either long-term decisions that are observable and therefore enforceable, or to screening considerations, which dominate the productivity gain.

Another result (parallel to that of Proposition 3) is the following.

PROPOSITION 3′. A disaggregate long-term performance measure dominates an aggregate long-term measure, regardless of whether the post-retirement period is informative. This dominance is strict with the exception of a knife-edge case.

This finding of Proposition 3′ differs from the prescription in Holmström and Milgrom [1987], in which linear aggregation is shown to be sufficient in their moral hazard setting. In our model, other than in a

---

[20] The tenure period typically represents a number of years. One would therefore expect that screening considerations would be mitigated. Note, however, in some industries (e.g., high tech), the turnover of managers is so high that the horizon of managers could be quite short. In such industries, screening considerations are likely to be more significant.

[21] This does not necessarily imply that in the subperiods within the manager's tenure period, his salary is constant; for example, the salary might increase with seniority.

knife-edge case, linear aggregation is not optimal (in the absence of disaggregation costs).

This two-period model can be enriched to study issues such as how managerial reputation is developed and how compensation schemes evolve. We conjecture that the informational rent given to type in ($P2$) will erode over time. Cross-sectional and time-series differences in the level of compensation are expected to increase with time, while the sensitivity of pay to performance is expected to decrease. In contrast, the sensitivity is expected to persist, if contingent compensation is designed for incentive purposes. This issue, among others, was examined by Murphy [1986], who analyzed compensation data to test the incentive- and learning-model-based hypotheses. The data generally supported the learning hypothesis. Murphy's work is consistent with the conjecture underlying our paper that, in a managerial context, differences in talent may be more important than effort aversion.

6. *Discussion*

Our analysis identified circumstances under which it is optimal to induce managers to maximize divisional rather than corporate objectives or to take short-term perspectives rather than long-term perspectives. Excluding a number of important aspects of organizational setup, such as collusion and reputation, we showed that the trade-off between allocation of managerial attention and screening costs, *by itself*, could explain why inefficient managerial focus is induced. We also showed that the optimal pay-to-performance relation need not be large, a result consistent with observations made by Jensen and Murphy [1990].

One important restriction is that we take the organization as given and focus on the hiring of only one manager. If the principal were to hire two managers for both divisions simultaneously, then the principal would compensate each manager conditional on the contract chosen by the manager of the other division, assuming that each manager's contract choice is ex post verifiable. While there would be twice as many contracts (for each division), the results would be essentially the same as those in the current paper for each combination of managers. Alternatively, if the choice of contracts is not verifiable, then the managers would have to be compensated, taking into account expectation over managerial type hired in the other division. Again, our analysis should hold in this case, as long as we preclude collusion among division managers.[22]

Another simplifying assumption is nonaversion of effort. This assumption highlights the trade-off between screening and costless allocation of managerial attention. If managers were (somewhat) effort averse, the op-

---

[22] The possibility of collusion is beyond the scope of this paper. For discussion on collusion, see Tirole [1986], Baiman, Evans, and Nagarajan [1991], and Laffont and Martimort [1995].

timal menu would be modified. Since an outcome-independent contract cannot induce effort, the contract for type *B* would have to become outcome contingent. We expect that the main results will continue to hold. Any of (*P*1), (*P*2), and (*P*3) would be optimal for some parameters. In (*P*2), the optimal solution would involve a menu of two contracts with a type *G* contract that is more performance sensitive. Also, the basic trade-off between allocation of managerial attention and screening would remain the same. In (*P*2) and (*P*3), type *G*, and possibly type *B*, would, in some cases, be induced to focus their attention exclusively on Division 1.

We have assumed that a manager could be one of two types. However, neither the optimality of an outcome-independent contract for the worst type nor the result that the best type enjoys no informational rent is sensitive to this assumption. Furthermore, even in an *N*-type case, hiring the worst-type manager would be optimal when the productivity of worse types is not too low and/or when the market alternative of the better types is sufficiently high (i.e., either the (*P*1) or the (*P*2) setting is optimal). We also expect the result of Proposition 2 to hold for certain parameter values.

Our analysis also points out a potentially costly consequence of merging business units like divisions. While such mergers could result in productivity gains, higher compensation becomes necessary as screening of talent becomes more difficult. If a merger does not change the information available, screening costs could be minimized by using the original disaggregate performance measures. Treating the newly merged division as if it still consists of separate divisions, however, might be unacceptable on horizontal equity grounds.

While our discussion has been normative, our model also suggests two empirical hypotheses. First, the model implies that managers who focus on their own division are more likely to be higher-quality managers (paid by outcome-dependent compensation) than managers who assign managerial attention to other divisions. Thus, ceteris paribus, managers who are induced to take a narrow perspective should be paid, on average, more than those who are induced to adopt a company-wide perspective.

Second, when managerial talent is company specific, our model predicts that the firm will attempt to hire better managers only. The compensation would be outcome-dependent, characterized by a relatively small slope. The reason for the small slope is that when talent is not transferable, the market alternatives for managers of different types are similar. A principal can then screen out bad managers by imposing a small amount of risk. When talent is transferable, on the other hand, hiring only better types requires compensation that is more performance sensitive. In this case, it is more likely that the principal will find it optimal to hire either type; this hiring choice will be implemented by a menu of contracts. Thus, observing a menu of contracts is consistent with managerial talent being transferable.

APPENDIX A

Proofs

Proof of Proposition 1

We prove our result for the case of unobservable managerial attention. The case of observable managerial attention is a special case of the proof below (with step 4 omitted).

In ($P2$), we need to consider at most two contracts in a menu, $\{ \underset{\sim}{s}^B, \underset{\sim}{s}^G\}$ (possibly only one contract is chosen by both types). We proceed in six steps. First, we show that the participation constraint for G must be binding. Second, we show that a pooling outcome-contingent contract is dominated by a menu consisting of an outcome-independent contract, w, for B and a contingent contract for G. Third, we show that offering a single outcome-independent contract cannot be optimal in ($P2$). In step 4, we establish incentive compatibility of the contract considered in the previous step. Last, we show that B receives an informational rent (step 5) and establish incentive compatibility of the variation contract (step 6).

STEP 1. Suppose to the contrary that the participation constraint for G is not binding. That is, $\phi \equiv \sum_i P_G(y_i \mid e_G^*)U(s_i^G) - \bar{U}_G > 0$, where e_G^* is the allocation of managerial attention chosen by G. By the optimality of e_G^*, we have $\forall\ e_G'$:

$$\sum P_G(y_i \mid e_G^*)U(s_i^G) \ \geq \ \sum P_G(y_i \mid e_G')U(s_i^G). \tag{1}$$

Consider the variation $\hat{s}_i = s_i^G - \delta$, where $\delta = s_i^G - U^{-1}(U(s_i^G) - \phi)$, and $\hat{\underset{\sim}{s}}_i^B = \underset{\sim}{s}_i^B$ if B adopted $\underset{\sim}{s}^B$ before (the variation) or offer no second contract if B adopted $\underset{\sim}{s}^G$ before. Type G will accept the new contract, since $\sum P_G(y_i \mid e_G^*)U(\hat{s}_i) = \sum P_G(y_i \mid e_G^*)(U(s_i^G) - \phi) = \bar{U}_G$. Furthermore, type G will continue to choose e_G^* over e_G', since by equation (1):

$$\begin{aligned}&\sum P_G(y_i \mid e_G')U(\hat{s}_i) = \sum P_G(y_i \mid e_G')U(\hat{s}_i^G) - \phi\\ &\leq \sum P_G(y_i \mid e_G^*)U(s_i^G) - \phi = \sum P_G(y_i \mid e_G^*)U(\hat{s}_i).\end{aligned}$$

As for B, there are two cases to consider:

Case 1: Suppose B chose $\underset{\sim}{s}^G$ before the contract variation. Then, given the above assumption of no alternative contract, B's response to $\hat{\underset{\sim}{s}}$ is either to: (i) drop out, in which case the principal is better off since by the assumed optimality of ($P2$), $EY_G(e_G) - \sum P_G(y_i|e_G)s_i > EY_B(e_B) - U^{-1}(\bar{U}_B)$ or (ii) continue and adopt $\hat{\underset{\sim}{s}}$, in which case the principal is also better off because expected compensation is reduced with no effect on allocation of managerial attention or productivity.

Case 2: Suppose B was choosing $\underset{\sim}{s}^B$ before the variation. Then the variation $\hat{\underset{\sim}{s}}$ has no effect on B as he finds $\hat{\underset{\sim}{s}}$ even less attractive than $\underset{\sim}{s}^G$, which he rejected prior to the variation.

STEP 2. Suppose that the optimal contract is an outcome-contingent pooling contract denoted by $\underset{\sim}{s}' \equiv \{s_1', s_2', s_3', s_4'\}$. From step 1, $EU_G(e_G^*; \underset{\sim}{s}') \equiv \sum P_G(y_i \mid e_G^*)U(s_i') = \bar{U}_G$, where e_G^* is the second-best allocation of managerial attention. Further, $EU_B(e_B^*; s') \geq \bar{U}_B$. It is also clear that an optimal contract must result in $EU_G(e_G^*; \underline{s}') > EU_B(e_B^*; \underline{s}')$, because otherwise the principal would be strictly better off offering a single outcome-independent contract, $w = U^{-1}(EU_G(e_G^*; \underline{s}'))$. Thus, the principal can strictly reduce his expected cost by offering a menu $\{w', \underline{s}'\}$ such that $w' = U^{-1}(EU_B(e_B; \underline{s}'))$. This $\{\underline{s}'\}$ is dominated by $\{w', \underline{s}'\}$. Note that this variation does not affect incentive compatibility and G does not adopt w'.

STEP 3. Suppose, to the contrary, that the solution to ($P2$) involves the pooling outcome-independent contract $W_G = U^{-1}(\bar{U}_G)$. There are two cases to consider.

Case 1: Suppose the principal wants to induce $e_G = 1$. Consider the alternative menu $\{w', \underline{s}^G\}$, where $s_1^G = s_3^G = W_G - \mu(\varepsilon)$ and $s_2^G = s_4^G = W_G + \varepsilon$, ε is sufficiently small, $w = U^{-1}(\max\{EU_B(0; \underline{s}^G), EU_B(1; \underline{s}^G)\})$, $\mu(\varepsilon) = W_G - U^{-1}((1 + Q_1)U(W_G) - Q_1 U(W_G + \varepsilon))$, and $Q_1 = \dfrac{P_G^2(e_G)}{1 - P_G^2(e_G)}$. Incentive compatibility of this menu for both type G and B is established in step 4. By construction, G is indifferent. It is also clear that this variation makes B worse off. Since by assumption $\bar{U}_G > \bar{U}_B$ (otherwise ($P3$) is optimal), B would accept w when ε is sufficiently small. To show that the principal is better off with the above variation, we need to establish that the effect on the principal's expected cost is such that:

$$\begin{aligned}
&\rho\{[P_G(y_1|e_G) + P_G(y_3|e_G)](W_G - \mu(\varepsilon)) + [P_G(y_2|e_G) \\
&+ P_G(y_4|e_G)](W_G + \varepsilon) - W_G\} = \rho\,\{-(1 - P_G^2(e_G))\mu(\varepsilon) \\
&+ P_G^2(e_G)\,\varepsilon\} < (1-\rho)\,\{W_G - U^{-1}((1 - P_B^2(e_B))U(W_G - \mu(\varepsilon)) \\
&+ P_B^2(e_B)U(W_G + \varepsilon))\} \equiv (1-\rho)A(\varepsilon), \qquad (2)
\end{aligned}$$

where $A(\varepsilon) = W_G - U^{-1}((1 - P_B^2(e_B))U(W_G - \mu(\varepsilon)) + P_B^2(e_B)U(W_G + \varepsilon))$ and $\hat{A}(\varepsilon) = (1 - P_B^2(e_B))U(W_G - \mu(\varepsilon)) + P_B^2(e_B)U(W_G + \varepsilon)$. Note that $A'(0) = -(U^{-1})' \circ (\hat{A}(0))\hat{A}'(0)$, and $\hat{A}'(0) = \{P_B^2(e_B) - (1 - P_B^2(e_B))\mu'(0)\}\, U'(W_G(e_G))$.

Inequality (2) is equivalent to $P_G^2(e_G)\varepsilon - (1 - P_G^2(e_G))\mu(\varepsilon) < \dfrac{1-\rho}{\rho} A(\varepsilon)$. Note that $\mu(0) = 0$ and $A(0) = 0$. Thus at $\varepsilon = 0$, the right-hand side (RHS) and the left-hand side (LHS) of the last inequality are identically zero. We examine the derivatives of both sides of the inequality with respect to ε, at $\varepsilon = 0$, and show that it is zero for the LHS and is strictly positive for the RHS. Note $\mu'(\varepsilon)|_{\varepsilon=0} = -(U^{-1})' \circ (U(W_G)) \times (-Q_1)U'(W_G) = Q_1$, where the last equality follows from $\dfrac{\partial}{\partial x}[U^{-1}(U(x))] = 1$. Therefore, $\dfrac{\partial}{\partial \varepsilon} LHS|_{\varepsilon=0} = P_G^2(e_G) - (1 - P_G^2(e_G))\dfrac{P_G^2(e_G)}{1 - P_G^2(e_G)} = 0$.

On the other hand, $\dfrac{\partial}{\partial \varepsilon} RHS|_{\varepsilon=0} = \dfrac{1-\rho}{\rho} A'(0) = \dfrac{1-\rho}{\rho}\{-(U^{-1})' \circ \hat{A}(0)\}\, U'(W_G)\{P_B^2(e_B) - (1 - P_B^2(e_B))Q_1\} > 0$. This derivative is strictly positive if and only if $\frac{P_B^2(e_B)}{1-P_B^2(e_B)} < \frac{P_G^2(e_G)}{1-P_G^2(e_G)} \Leftrightarrow P_G^2(e_G) > P_B^2(e_B)$. Recall that $P_G^2(1) \geq P_B^2(1)$ by assumption. Thus, the strict inequality holds when $P_G^2(1) > P_B^2(1)$ under the variation considered so far.

To establish the strict inequality of equation (2) when Division 2's outcome is uninformative, we need to consider the following variation: $s_1^G = W_G - \mu_1(\varepsilon_1) - \mu_2(\varepsilon_2)$, $s_3^G = W_G - \mu_1(\varepsilon_1) + \varepsilon_2$, $s_2^G = s_4^G = W_G + \varepsilon_1$, $\mu_1(\varepsilon_1) = W_G - U^{-1}((1 + Q_2)U(W_G) - Q_2U(W_G + \varepsilon_1))$, and $\mu_2(\varepsilon_2) = W_G - \mu_1(\varepsilon_1) - U^{-1}\{(1 + Q_3)U(W_G - \mu_1(\varepsilon_1)) - Q_3U(W_G - \mu_1(\varepsilon_1) + \varepsilon_2)\}$, where $\varepsilon_2 << \varepsilon_1$, $Q_2 = P_G^2(1)/(1 - P_G^2(1))$, and $Q_3 = P_G^1(1)/(1 - P_G^1(1))$.

As with the first variation, this variation implies that the principal is better off, type G is indifferent, type B is worse off (but still accepts the contract), and incentive compatibility is maintained for both types.

Case 2: Suppose the optimal $e_G = 0$. Consider the alternative menu $\{w, \underset{\sim}{s}^G\}$ where $s_1^G = s_2^G = W_G - \mu(\varepsilon)$ and $s_3^G = s_4^G = W_G + \varepsilon$; w and $\mu(\varepsilon)$ are defined as in Case 1 with $Q_4 = \frac{P_G^1(e_G)}{1-P_G^1(e_G)}$. The proof is then identical to the first part of Case 1 and is therefore omitted.

STEP 4. Case 1: To establish incentive compatibility of the contract in step 3, we need to check for $t = B,G$:[23] $\sum P_t(y_i|1)U(s_i^G) \geq \sum P_t(y_i|0)U(s_i^G)$, or $(1 - P_t^2(1))U(W_G - \mu(\varepsilon)) + P_t^2(1)U(W_G + \varepsilon) \geq (1 - P_t^2(0))U(W_G - \mu(\varepsilon)) + P_t^2(0)U(W_G + \varepsilon)$, or $P_t^2(1)\{U(W_G + \varepsilon) - U[W_G - \mu(\varepsilon)]\} \geq P_t^2(0)\{U(W_G + \varepsilon) - U[W_G - \mu(\varepsilon)]\}$, or $P_t^2(1) \geq P_t^2(0)$, which is met by assumption. Thus, both types would choose $e_t = 1$ under the variation. Of course, when the optimal (second-best) menu is offered, B will adopt any recommended e_B, since he gets the outcome-independent contract (i.e., the certainty equivalent of the utility he gets under the variation).

Case 2, where we induce $e_G = 0$, is dealt with similarly, with the applicable variation introduced above.

STEP 5. Suppose to the contrary that the optimal menu in ($P2$) involves no informational rent for type B, $w = U^{-1}(\bar{U}_B)$. In this case, the principal is strictly better off switching to ($P3$) in contradiction to the optimality of ($P2$). As in step 3, there are two cases to consider: $e_G = 1$ and $e_G = 0$. Consider first the case of $e_G = 1$. Here, a single contract $\underset{\sim}{s}''$ is offered instead of $\{w, \underset{\sim}{s}\}$, where $s_1'' = s_1$, $s_2'' = s_2$, $s_3'' = s_3 - \mu(\varepsilon)$ and $s_4'' = s_4 + \varepsilon$, where $\mu(\varepsilon) = s_3 - U^{-1}[U(s_3) + Q_5U(s_4) - Q_5U(s_4 + \varepsilon)]$ and $Q_5 = \frac{P_G(y_4|1)}{P_G(y_3|1)}$. Incentive compatibility of this contract is established in step 6. Clearly, G is unaffected. For B, note that by hypothesis $\max\{\sum P_B(y_i|0) U(s_i), \sum P_B(y_i|1)U(s_i)\} = w$ prior to the variation. The variation (weakly) reduces B's expected utility because:

[23] This inequality also applies to type B, since we must have:

$$w = U^{-1}(\max\{EU_B(1; \underset{\sim}{s}^G), EU_B(0; \underset{\sim}{s}^G)\}).$$

$$\begin{aligned}&\Sigma P_B(y_i\,|e_B)\;[U(s_i'') - U(s_i)]\\ &= P_B(y_3\,|e_B)(U(s_3 - \mu(\varepsilon)) - U(s_3)) + P_B(y_4\,|e_B)(U(s_4 + \varepsilon) - U(s_4))\\ &= [P_B(y_4\,|e_B) - Q_5 P_B(y_3\,|e_B)](U(s_4 + \varepsilon) - U(s_4))\\ &= \left\{\frac{P_B^2(e_B)}{1-P_B^2(e_B)} - \frac{P_G^2(1)}{1-P_G^2(1)}\right\}[U(s_4+\varepsilon) - U(s_4)] \le 0,\end{aligned}$$

where the last inequality is strict if $e_B = 0$ or $e_B = 1$ and Division 2's outcome is informative.

The principal's expected cost clearly goes up; but it is also apparent that the increase in cost converges to zero as $\varepsilon \to 0$. The effect on gross output, however, is bounded away from zero for any given ρ, since by assumption $M \equiv \Sigma[P_G(y_i\,|e_G) - P_B(y_i\,|e_B)]\; y_i > 0$ and the gain is $(1-\rho)M > 0$. When $e_G = 0$, the proof is similar with the exception that the single contract considered is s''' where $s''' = s_1 - \mu(\varepsilon)$, $s_2''' = s_2$, $s_3''' = s_3 + \varepsilon$, $s_4''' = s_4$, and $\mu(\varepsilon)$ is defined similarly to the first case.

STEP 6. We show that the contract considered in step 5 is incentive compatible. Consider the case of $e_G = 1$. We need to establish $\Sigma P_G(y_i\,|1)\,U(s_i'') \ge \Sigma P_G(y_i\,|0)\,U(s_i'')$. We know from step 5 that by construction $P_G(y_3|1)(U(s_3) - U(s_3'')) = P_G(y_4|1)(U(s_4'') - U(s_4))$, and $s_2'' = s_2$, $s_3'' = s_3$. Also by hypothesis, $\Sigma P_G(y_i\,|1)\,U(s_i) \ge \Sigma P_G(y_i\,|0)\,U(s_i)$. Incorporating these relations, we only need to show $P_G^2(1) > P_G^2(0)$, which is assumed in section 2.

Proof of Observation 4

When allocation of managerial attention is unobservable, it is clear that company-wide compensation is necessary to induce $e_G = 1$. If allocation of managerial attention is observable, however, the principal can just instruct $e_G = 1$ even under division-based (*DB*) compensation. Thus, we need to show that the principal can lower his compensation cost by switching to *DB* compensation from disaggregate company-wide when allocation of managerial attention is observable, holding $e_B = e_G = 1$. Since this can be shown using a technique similar to that used in the proof of Proposition 1, we just outline, but skip the details of, the proof here. We let the optimal *DB* compensation be $s_1 = s_2 = s_l$ and $s_3 = s_4 = s_h$, and consider a variation such that $s_1^* = s_l - \mu(\varepsilon)$ and $s_2^* = s_l + \varepsilon$, where $\mu(\varepsilon) = s_l - U^{-1}((1 + Q_6)U(s_l) - Q_6 U(s_l + \varepsilon))$, and $Q_6 = \dfrac{P_G(y_2|1)}{P_G(y_1|1)}$. This variation leaves the type G manager with the same expected utility, $EU_G(1)$. We can show that for sufficiently small ε values, the principal's overall expected cost is reduced by this variation. (Details are available upon request.)

Proof of Proposition 2

The existence of situations in which divisional focus is optimal is established in the Corollary to Proposition 2. We prove here that in such situations, the optimal contract for type G is indeed division-based. That is, to induce $\{e_G = 0,\ e_B = 1\}$, the principal should use a menu $\{w,\ \underline{s}\}$, where $\underline{s}$ is division-based $(s_1 = s_2,\ s_3 = s_4)$.

Suppose, to the contrary, $\underline{s}$ is a strictly disaggregate company-wide performance measure. Then, we show below that the principal would be better off adopting *DB* compensation. Recall from Proposition 1 that $EU_G(0; \underline{s}) = \bar{U}_G$. If allocation of managerial attention chosen is $\{e_G = 0,\ e_B = 1\}$, we must have $EU_G(0; \underline{s}) \geq EU_G(1; \underline{s})$ and $U(w) = EU_B(1; \underline{s}) \geq EU_B(0; \underline{s})$, or $U(w) = EU_B(0; \underline{s}) > EU_B(1; \underline{s})$.

Consider an alternative *DB* contract, $\underline{s}'$, for type G such that $s'_1 = s'_2 = W'$ and $s'_3 = s'_4 = W''$, where W' and W'' are chosen such that $U(W') = (1 - P_G^2(0))\,U(s_1) + P_G^2(0)\,U(s_2)$ and $U(W'') = (1 - P_G^2(0))\,U(s_3) + P_G^2(0)\,U(s_4)$. With this variation, G's expected utility will be $EU_G(0; \underline{s}') = EU_G(0; \underline{s})$. Furthermore, $EU_B(0; \underline{s}') = EU_B(0; \underline{s})$, where the last equality is due to the fact that $P_B^2(0) = P_G^2(0)$. The expected cost of compensating G, however, will be lower under $\underline{s}'$ than under $\underline{s}$ by the concavity of $U(\cdot)$. That is, $W' < (1 - P_G^2(0))\,s_1 + P_G^2(0)\,s_2$, $W'' < (1 - P_G^2(0))\,s_3 + P_G^2(0)\,s_4$, and $(1 - P_G^1(0))\,W' + P_G^1(0)\,W'' < (1 - P_G^1(0))\{(1 - P_G^2(0))\,s_1 + P_G^2(0)\,s_2\} + P_G^1(0)\{(1 - P_G^2(0))\,s_3 + P_G^2(0)\,s_4\}$.

Case *A*: Suppose the original disaggregate company-wide compensation is such that $W'' > W'$. Type G will still choose $e_G = 0$ over $e_G = 1$, because $EU_G(0; \underline{s}') = (1 - P_G^1(0))\,U(W') + P_G^1(0)\,U(W'') = EU_G(0; \underline{s}) > (1 - P_G^1(1))\,U(W') + P_G^1(1)\,U(W'') = EU_G(1; \underline{s}')$, since $P_G^1(0) > P_G^1(1)$ and $U(W'') > U(W')$. As to type B, we need to consider the two alternative cases regarding the original compensation.

Case *A*.1: $EU_B(1; \underline{s}) \geq EU_B(0; \underline{s})$. In this case, w in the original menu would have been set such that $U(w) = EU_B(1; \underline{s})$. Under the new menu, we have $EU_B(1; \underline{s}') = (1 - P_B^1(1))\,U(W') + P_B^1(1)\,U(W'') < (1 - P_B^1(0))\ U(W') + P_B^1(0)\,U(W'') = EU_B(0; \underline{s}')$. Thus, under the variation, B will prefer to take $e_B = 0$ (if he selects $\underline{s}'$). The principal can, however, induce B to take $e_B = 1$ by setting w' such that $U(w') = EU_B(0; \underline{s}')$. Since $EU_B(0; \underline{s}') = EU_B(0; \underline{s}) \leq EU_B(1; \underline{s})$, the principal can reduce B's expected compensation.

Case *A*.2: $EU_B(1; \underline{s}) < EU_B(0; \underline{s})$. Under the new menu, $EU_B(1; \underline{s}') = (1 - P_B^1(1))\,U(W') + P_B^1(1)\,U(W'') < (1 - P_B^1(0))\,U(W') + P_B^1(0)\,U(W'') = EU_B(0; \underline{s}') = EU_B(0; \underline{s})$. The noncontingent compensation for B is unchanged because $U(w) = EU_B(0; \underline{s}') = EU_B(0; \underline{s}) > EU_B(1; \underline{s}')$.

Case *B*: Suppose the original disaggregate company-wide compensation is such that $W'' \leq W'$. We show that such a menu is suboptimal and is dominated by another menu for which $W'' > W'$. To see that, we show first that under the assumed menu, $EU_B(\cdot; \underline{s}) \geq \bar{U}_G$.

By the definition of W' and W'', $EU_G(0; \underline{s}) = (1 - P_G^1(0))\,U(W') + P_G^1(0)\,U(W'')$, and $EU_B(0; \underline{s}) = (1 - P_B^1(0))\,U(W') + P_B^1(0)\,U(W'')$. Since $P_G^1(0) > P_B^1(0)$ and $W' \geq W''$ by assumption, the last two equations imply $EU_G(0; \underline{s}) \leq EU_B(0; \underline{s})$. As before, we have either *A*.1: $EU_B(0; \underline{s}) \leq EU_B(1; \underline{s})$ or *A*.2: $EU_B(0; \underline{s}) > EU_B(1; \underline{s})$. If *A*.1 holds, we get $EU_G(1; \underline{s}) \leq EU_G(0; \underline{s}) \leq EU_B(0; \underline{s}) \leq EU_B(1; \underline{s}) = U(w)$. If *A*.2 applies, $EU_G(1; \underline{s}) \leq EU_G(0; \underline{s}) \leq$

$EU_B(0; \underset{\sim}{s}) = U(w)$. Either way, $EU_B(\cdot; \underset{\sim}{s}) \geq \bar{U}_G$, since $EU_G(0; \underset{\sim}{s}) = \bar{U}_G$. Thus, the original menu $\{w, \underset{\sim}{s}\}$ is suboptimal. Since $\bar{U}_G > \bar{U}_B$, the principal can make himself better off offering a single outcome-independent contract $w' = U^{-1}(EU_G(0; \underset{\sim}{s}))$. (Note that we have already established in step 3, case 2, of Proposition 1, that such a pooling contract is suboptimal.)

Proof of Corollary to Proposition 2

Divisional focus on the part of type G managers might be optimal either in ($P2$) or ($P3$) if productivity gain is less than the increased cost of screening. Specifically, divisional focus is optimal if, in ($P2$):

$$\rho[P_G^2(1) - P_G^2(0) - P_G^1(0) + P_G^1(1)]\Delta$$
$$< \rho\{\sum P_G(y_i|1)s_i^{DCW} - \sum P_G(y_i|0)s_i^{DB}\} + (1-\rho)(w^{DCW} - w^{DB}). \quad (3)$$

In ($P3$), the parallel condition is:

$$[P_G^2(1) - P_G^2(0) - P_G^1(0) + P_G^1(1)]\Delta$$
$$< \{\sum P_G(y_i|1)s_i^{DCW} - \sum P_G(y_i|0)s_i^{DB}\}. \quad (4)$$

In the above inequalities, $s_i^{DCW}(s_i^{DB})$ denotes the compensation of type G when outcome is y_i, $i = 1, \ldots, 4$, and w^{DCW} (w^{DB}) is the outcome-independent compensation for type B under a disaggregate company-wide (DB) performance measure. Note that both vectors, $\underset{\sim}{s}^{DCW}$ and $\underset{\sim}{s}^{DB}$, are in general different in ($P2$) and ($P3$). The *LHS* of equation (3) (or (4)) is the expected net productivity gain, whereas the *RHS* is the incremental screening cost required to induce type G to allocate his managerial attention to Division 2. The question of divisional focus is of interest only when allocation of managerial attention is unobservable. When allocation is observable, the principal always specifies $e_G = 1$ and uses disaggregate company-wide compensation if Division 2's outcome is informative or DB compensation otherwise.

To establish the corollary for the unobservable case, we show that the *RHS* of (3) and (4) are strictly positive, if the allocation of managerial attention has to be induced and if Division 2's outcome is sufficiently uninformative, i.e., $P_G^2(1) = P_B^2(1) + \varepsilon$, $\varepsilon > 0$ and is sufficiently small. Inducing type G to allocate his attention to Division 2 requires additional risk premium beyond what is required for the purpose of screening when allocation of managerial attention is observable. We consider first the optimal compensation in the absence of the incentive-compatibility (IC) constraints and identify circumstances in which the *RHS* is strictly positive. For those cases, there exists a Δ, small enough, such that the *LHS* < *RHS*. This constitutes a lower bound for the *RHS*, since incorporating IC constraints would only (weakly) increase the *RHS*.

To compute the expected cost of compensation, consider the case where the principal specifies $e_G = 1$ with a disaggregate company-wide compensation scheme and $e_G = 0$ with a *DB* compensation scheme. We show below that the *RHS* of (3) is strictly positive when Division 2 is sufficiently uninformative.

Consider (*P*2). The expected cost, $EC(e)$, of compensation for the principal when managers have a square-root utility function is as follows. (The derivation of the explicit solution is omitted but is available upon request.)

$$EC(e) = \left\{\frac{\bar{U}_G}{2\rho - 1 + (1-\rho)R(e)}\right\}^2 \rho\,\{(1-\rho)^2 R(e)^2$$

$$+ (1-\rho)(3\rho - 1)\,R(e) + \rho(2\rho - 1)\}, \tag{5}$$

where $e \equiv (e_B, e_G)$, $R_i(e) = \dfrac{P_B(y_i \mid e_B)}{P_G(y_i \mid e_G)}$ and $R(e) = \sum_{j=1}^{4} \dfrac{P_B(y_j \mid e_B)^2}{P_G(y_j \mid e_G)}$, for $i = 1, \ldots, 4$.

The optimal menu $\{w, s\}$ is:

$$s_i(e) = \bar{U}_G^2 \left\{\frac{\rho + (1-\rho)\,(R(e) - R_i(e))}{2\rho - 1 + (1-\rho)R(e)}\right\},$$

$$w(e) = \left\{\frac{\rho \bar{U}_G}{2\rho - 1 + (1-\rho)R(e)}\right\}^2.$$

Note that under the disaggregate company-wide compensation, $R_i(e_G) = \dfrac{P_B(y_i \mid 1)}{P_G(y_i \mid e_G)}$ and $R(e_G) = \sum \dfrac{P_B(y_i \mid 1)^2}{P_G(y_i \mid e_G)}$ for $i = 1, \ldots, 4$. Since $e_B = 1$, for any optimal compensation scheme, only type *G*'s allocation of managerial attention is relevant. Under *DB*, we do not differentiate between y_1 and y_2 nor y_3 and y_4. Let $S(e_G)$ and $S_k(e_G)$, $k = l, h$ be the counterpart of $R(e_G)$ and $R_i(e_G)$ under *DB*, i.e., $S(e_G) = \dfrac{(1 - P_B^1(1))^2}{1 - P_G^1(e_G)} + \dfrac{(P_B^1(1))^2}{P_G^1(e_G)}$ and $S_l(e_G) = \dfrac{1 - P_B^1(1)}{1 - P_G^1(e_G)}$ and $S_h(e_G) = \dfrac{P_B^1(1)}{P_G^1(e_G)}$. Substituting the equation for expected cost of compensation for both disaggregate company-wide and *DB* compensation into equation (3), we can rewrite the *RHS* of (3) as:

$$\frac{\bar{U}_G^2 \rho (1-\rho)^2 [S(0) - R(1)]\{(2\rho - 1)^2 + (1-\rho)\{(1-\rho)S(0)R(1) + (2\rho - 1)[S(0) + R(1)]\}\}}{[2\rho - 1 + (1-\rho)R(1)]\,[2\rho - 1 + (1-\rho)S(0)]}. \tag{6}$$

It can be shown that the denominator and the inside of the curly brackets in the numerator are strictly positive. Thus, the sign of the *RHS* depends on the sign of $S(0) - R(1)$. Recall $P_G^2(1) = P_B^2(1) + \varepsilon$. We show below that $S(0) - R(1) > 0$ for sufficiently small ε.

$S(0) > \quad R(1)$

$$\Leftrightarrow \quad \frac{[P_G^1(0) - P_B^1(1)]^2}{[1 - P_G^1(0)]P_G^1(0)} > \frac{[P_G^1(1) - P_B^1(1)]^2}{[1 - P_G^1(1)]P_G^1(1)}$$

$$+ \frac{\varepsilon^2}{[1 - P_G^2(1)]P_G^2(1)} + \frac{\varepsilon^2 [P_G^1(1) - P_B^1(1)]^2}{[1 - P_G^1(1)]P_G^1(1)[1 - P_G^2(1)]P_G^2(1)}$$

$$\Leftrightarrow \quad P_G^1(0)P_G^1(1)[1 - 2P_B^1(1)] + [P_B^1(1)]^2 [P_G^1(0) + P_G^1(1) - 1]$$

$$> 0 \text{ as } \varepsilon \to 0. \qquad (7)$$

Note that, as $\varepsilon \to 0$, $S(0) - R(1)$ depends on Division 1 only. This is so because when Division 2's outcome is uninformative, the optimal compensation is division-based (when the *IC* constraints are not binding). Thus, the comparison of $S(0)$ and $R(1)$ is equivalent to the comparison of $S(0)$ and $S(1)$.

We need to consider three alternative cases to establish inequality in (7):

(1) $P_B^1(1) = \frac{1}{2}$. Inequality (7) holds, since $P_G^1(0) > P_G^1(1) > P_B^1(1) = \frac{1}{2}$.

(2) $P_B^1(1) < \frac{1}{2}$.

$$\begin{aligned} \text{The } LHS \text{ of } (7) &> P_G^1(0)P_G^1(1)[1 - 2P_B^1(1)] + [P_B^1(1)]^2[2P_B^1(1) - 1] \\ &= [1 - 2P_B^1(1)]\{P_G^1(0)P_G^1(1) - [P_B^1(1)]^2\} > 0. \end{aligned}$$

(3) $P_B^1(1) > \frac{1}{2}$.

$$\begin{aligned} \text{The } LHS \text{ of } (7) &= P_B^1(1)[1 - P_B^1(1)][P_G^1(0) + P_G^1(1) - 2P_B^1(1)] \\ &\quad -[2P_B^1(1) - 1][P_G^1(0) - P_B^1(1)[P_G^1(1) - P_B^1(1)] \\ &> [1 - P_B^1(1)]\{[P_G^1(0) + P_G^1(1) - 2P_B^1(1)]P_B^1(1) \\ &\quad - [2P_B^1(1) - 1][P_G^1(0) - P_B^1(1)]\} \\ &= [1 - P_B^1(1)]\{P_G^1(0)[1 - P_B^1(1)] - P_B^1(1)[1 - P_G^1(1)]\} > 0, \end{aligned}$$

where the first inequality results from substituting $[1 - P_B^1(1)]$ for $[P_G^1(1) - P_B^1(1)]$. This establishes that the *RHS* of (3) is strictly positive if Division 2's outcome is sufficiently uninformative. When we incorporate the *IC* constraints, the required expected compensation would (weakly) increase, impacting the *RHS* (but not the *LHS*) of the inequality.

Consider now the case of (*P*3). For brevity, we only provide highlights of the proof. Assume again that the *IC* constraints are not binding. When Division 2's outcome is informative, we know, from Observation 4, that the optimal compensation cannot be division-based. However, it can be shown that as Division 2's outcome becomes less informative ($P_G^2(1) \to P_B^2(1)$), the optimal company-wide compensation converges to the optimal *DB* compensation.

Let $EC(e_G)$ be the expected *DB* compensation cost, given e_G. Notice that the variable of interest—expected compensation cost under disaggregate company-wide with $e_G = 1$—converges to $EC(e_G = 1)$ as Division 2's outcome becomes less informative. We therefore compare the expected cost of *DB* compensation given $e_G = 1$ with that of *DB* compensation given $e_G = 0$, and show below that $EC(e_G = 1) > EC(e_G = 0)$. The expected *DB* compensation cost is:

$$EC(e_G) = \frac{1}{[P_G^1(e_G) - P_B^1(1)]^2} \times \{[(P_B^1(1))^2 + P_B^1(1) - 2P_B^1(1)P_G^1(e_G)]\bar{U}_G^2 + P_G^1(e_G)[1 - P_G^1(e_G)]\bar{U}_B^2 - 2P_G^1(e_G)[1 - P_G^1(e_G)]\bar{U}_B\bar{U}_G\},$$

where $P_G^1(e_G)$ depends on allocation of managerial attention of type *G*. We need to show that $EC(1) > EC(0)$. Recall that $P_G^1(0) > P_G^1(1)$. Let P_G be a continuous variable. Then:

$$\frac{\partial EC(\cdot)}{\partial P_G} = \frac{-[P_G(1 - P_B^1(1)) + P_B^1(1)(1 - P_G)](\bar{U}_G - \bar{U}_B)^2}{[P_G - P_B^1(1)]^2} < 0.$$

Thus, the *RHS* of (4) is strictly positive if Division 2's outcome is sufficiently uninformative. When we incorporate the *IC* constraints, the compensation could only increase, impacting the *RHS* (but not the *LHS*) of the inequality.

To summarize, we can show that the *RHS* of (3) and (4) are strictly positive when Division 2's outcome is uninformative. Therefore, there exists a sufficiently small value of Δ such that the corollary holds.

Proof of Proposition 3

Let e_G be the optimal choice of managerial attention. Recall that the optimal $e_B = 1$. We first consider the hiring policy of ($P2$) with the following three cases: *A.i*: $\frac{P_G(y_3|e_G)}{P_G(y_2|e_G)} > \frac{P_B(y_3|1)}{P_B(y_2|1)}$, *A.ii*: $\frac{P_G(y_3|e_G)}{P_G(y_2|e_G)} < \frac{P_B(y_3|1)}{P_B(y_2|1)}$, and *A.iii*: $\frac{P_G(y_3|e_G)}{P_G(y_2|e_G)} = \frac{P_B(y_3|1)}{P_B(y_2|1)}$.

Case *A.i*: Let the optimal aggregate company-wide compensation be $\{s_1, s_2 = s_3, s_4\}$. Consider a variation such that $s_2^* = s_2 - \mu(\varepsilon)$, $s_3^* = s_2 + \varepsilon$, $s_1^* = s_1$, and $s_4^* = s_4$ where $\mu(\varepsilon) = s_2 - U^{-1}((1 + Q_7)U(s_2) - Q_7U(s_2 + \varepsilon))$, ε is sufficiently small, and $Q_7 = \frac{P_G(y_3|e_G)}{P_G(y_2|e_G)}$. Assume that choice of managerial attention is unaffected (to be verified); then this variation leaves a type *G* manager with the same expected utility. We now show that the principal's overall expected cost is reduced by this variation. We have to establish:

$$\begin{aligned}
&\rho\{P_G(y_1|e_G)s_1 + P_G(y_2|e_G)(s_2 - \mu(\varepsilon)) + P_G(y_3|e_G)(s_2 + \varepsilon) + P_G(y_4|e_G)s_4 \\
&\quad - (P_G(y_1|e_G)s_1 + P_G(y_2|e_G)s_2 + P_G(y_3|e_G)s_2 + P_G(y_4|e_G)s_4)\} \\
&= \rho\{P_G(y_3|e_G)\varepsilon - P_G(y_2|e_G)\mu(\varepsilon)\} \\
&< (1-\rho)\{U^{-1}(P_B(y_1|1)\,U(s_1) + (P_B(y_2|1) + P_B(y_3|1))\,U(s_2) + P_B(y_4|1) \\
&\quad U(s_4)) - U^{-1}(P_B(y_1|1)\,U(s_1) + P_B(y_2|1)\,U(s_2 - \mu(\varepsilon)) + P_B(y_3|1) \\
&\quad U(s_2 + \varepsilon) + P_B(y_4|1)\,U(s_4))\} \\
&\equiv (1-\rho)\{U^{-1}(A_1) - U^{-1}(\hat{A}(\varepsilon))\} \equiv (1-\rho)A(\varepsilon), \qquad (8)
\end{aligned}$$

where $A(\varepsilon)=U^{-1}(A_1) - U^{-1}(\hat{A}(\varepsilon))$, $A_1 = P_B(y_1|1)\,U(s_1) + P_B(y_2|1)\,U(s_2) + P_B(y_3|1)\,U(s_2) + P_B(y_4|1)\,U(s_4)$, $A(\varepsilon) = P_B(y_1|1)\,U(s_1) + P_B(y_2|1)\,U(s_2 - \mu(\varepsilon)) + P_B(y_3|1)\,U(s_2 + \varepsilon) + P_B(y_4|1)\,U(s_4)$, and $\hat{A}(\varepsilon) = P_B(y_1|1)\,U(s_1) + P_B(y_2|1)\,U(s_2 - \mu(\varepsilon)) + P_B(y_3|1)\,U(s_2 + \varepsilon) + P_B(y_4|1)\,U(s_4)$. Note $A'(\varepsilon)|_{\varepsilon=0} = -(U^{-1})' \circ (\hat{A}(0))\hat{A}'(0)$, and $\hat{A}'(\varepsilon)|_{\varepsilon=0} = \{-P_B(y_2|1)\mu'(0) + P_B(y_3|1)\}U'(s_2)$. Inequality (8) simplifies to $P_G(y_3|e_G)\varepsilon - P_G(y_2|e_G)\mu(\varepsilon) < \frac{1-\rho}{\rho}A(\varepsilon)$. Note that $\mu(0) = 0$ and $A(0) = 0$. Thus at $\varepsilon = 0$, the *RHS* and the *LHS* of the above equation are identically zero.

We examine the derivatives of both sides of the last equation with respect to ε, evaluated at $\varepsilon = 0$, and show that it is zero for the *LHS* and strictly positive for the *RHS*:

$$\mu'(\varepsilon)|_{\varepsilon=0} = -(U^{-1})' \circ (U(s_2)) \times (-Q_7)\,U'(s_2) = Q_7.$$

Therefore, $\frac{\partial}{\partial\varepsilon} LHS\,|_{\varepsilon=0} = P_G(y_3|e_G) - P_G(y_2|e_G)\mu'(\varepsilon) = 0$. On the other hand:

$$\begin{aligned}
\frac{\partial}{\partial\varepsilon} RHS\,|_{\varepsilon=0} &= \frac{1-\rho}{\rho} \times A'(0) \\
&= -\frac{1-\rho}{\rho}\,(U^{-1})' \circ (\hat{A}(0)) \times U'(s_2)\left(P_B(y_3|1) - P_B(y_2|1)\frac{P_G(y_3|e_G)}{P_G(y_2|e_G)}\right) \\
&> 0 \\
&\Leftrightarrow \frac{P_G(y_3|e_G)}{P_G(y_2|e_G)} > \frac{P_B(y_3|1)}{P_B(y_2|1)}.
\end{aligned}$$

We can show that an optimal contract for type G must be monotone in y_i, with strict inequality for some y_i. Thus, if $e_G^* = 1$, $EU_G(1; \underline{s}) > EU_G(0; \underline{s})$. We can therefore choose ε such that $EU_G(1; \underline{s}^*) > EU_G(0; \underline{s}^*)$.

Case *A.ii*: The proof proceeds exactly the same as in Case *A.i*, with the exception that the variation now is $s_2^* = s_2 + \varepsilon$ and $s_3^* = s_2 - \mu(\varepsilon)$.

Case *A.iii*: In this the knife-edge case (in the space of all probability values), aggregate company-wide compensation is sufficient for disaggregate company-wide compensation.

Next, we consider the case of ($P3$). The variation we consider here maintains the expected utility of type B. We show that type G's expected utility will be higher with the variation than his reservation utility. The principal then can make himself better off by reducing compensation such that the type G manager receives exactly his reservation utility. Let the optimal aggregate company-wide compensation be $\{s_1, s_2 = s_3, s_4\}$ as before. Consider a variation such that $s_2' = s_2 - \mu'(\varepsilon')$, $s_3' = s_2 + \varepsilon'$, $s_1' = s_1$, and $s_4' = s_4$ where $\mu'(\varepsilon') = s_2 - U^{-1}((1 + Q_8)U(s_2) - Q_8 U(s_2 + \varepsilon'))$, ε' is sufficiently small, and $Q_8 = \dfrac{P_B(y_3|e_B)}{P_B(y_2|e_B)}$. This variation leaves a type B manager with the same expected utility.

The principal's expected cost is reduced if $P_G(y_3|e_G)\varepsilon' - P_G(y_2|e_G)\mu'(\varepsilon') < 0$. Examine the derivative of the LHS of this inequality. Note that $\dfrac{\partial \mu'(\varepsilon')}{\partial \varepsilon'}\Big|_{\varepsilon' = 0} = -(U^{-1})' \circ (U(s_2))(-Q_8)U'(s_2) = Q_8$. Therefore, $\dfrac{\partial LHS}{\partial \varepsilon'} = P_G(y_3|e_G) - P_G(y_2|e_G)Q_8 = P_G(y_3|e_G) - P_G(y_2|e_G)\dfrac{P_B(y_3|e_B)}{P_B(y_2|e_B)}$. Then, $\dfrac{\partial LHS}{\partial \varepsilon'} = 0$, if $\dfrac{P_G(y_3|e_G)}{P_G(y_2|e_G)} = \dfrac{P_B(y_3|e_B)}{P_B(y_2|e_B)}$, and $\dfrac{\partial LHS}{\partial \varepsilon'} < 0$, if $\dfrac{P_G(y_3|e_G)}{P_G(y_2|e_G)} < \dfrac{P_B(y_3|e_B)}{P_B(y_2|e_B)}$. If $\dfrac{P_G(y_3|e_G)}{P_G(y_2|e_G)} > \dfrac{P_B(y_3|e_B)}{P_B(y_2|e_B)}$, the principal can reduce his expected compensation cost by another variation such that $s_3' = s_2 - \mu'(\varepsilon')$, $s_2' = s_2 + \varepsilon'$, $s_1' = s_1$, and $s_4' = s_4$.

REFERENCES

AMERSHI, A., AND S. DATAR. "Incomplete Contracts, Production Expertise, and Incentive Effects of Modern Manufacturing Practices." Working paper, Stanford University, 1993.

BAIMAN, S.; J. EVANS III; AND N. NAGARAJAN. "Collusion in Auditing." *Journal of Accounting Research* 29, no. 1 (Spring 1991): 1–18.

BANKER, R., AND S. DATAR. "Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation." *Journal of Accounting Research* 27, no. 1 (Spring 1989): 21–39.

BARTLETT, C. "Procter & Gamble Europe: Vizir Launch." Harvard Business Case no. 9–384–139, 1983.

BITETTI, C. M., AND K. MERCHANT. "Chemical Bank: Allocation of Profits." Harvard Business Case no. 9–184–047, 1983.

BLAXIL, M. F., AND T. M. HOUT. "The Fallacy of the Overhead Quick Fix." *Harvard Business Review* (July/August 1991): 93–101.

BUSHMAN, R.; R. INDJEJIKIAN; AND A. SMITH. "Compensation Contract in Hierarchical Organizations: Determinants of Incentive Pay for Business Unit Managers." Working paper, University of Chicago, 1995.

CALVO, G., AND S. WELLISZ. "Hierarchy, Ability, and Income Distribution." *Journal of Political Economy* 87, no. 5, pt. 1 (October 1979): 165–72.

DARROUGH, M. "Managerial Incentives for Short-Term Results: A Comment." *Journal of Finance* 62, no. 4 (September 1987): 1097–1102.

DAUCH, R. F. "Improving Resource Allocation and Control for New Product Development Activities." M.S. thesis, Massachusetts Institute of Technology, 1992.

"Executive Compensation." *Employee Benefit Plan Review* (February 1992).

FELTHAM, G. A., AND J. S. HUGHES. "Communication of Private Information in Capital Markets: Contingent Contracts and Verified Reports." In *Economic Analysis of Information and Contracts*, edited by G. A. Feltham, A. H. Amershi, and W. T. Ziemba, pp. 271–317. Boston, Mass.: Kluwer, 1988.

FOX, H. *Top Executive Bonus Plans.* New York: Conference Board, 1979.

GEANAKOPLOS, J., AND P. MILGROM. "A Theory of Hierarchies Based on Limited Managerial Attention." *Journal of Japanese and International Economies* 5, no. 3 (September 1991): 205–25.

GIBBS, M. "Assignment, Information, and Incentives: An Economic Approach to Process in Pay and Performance Appraisals." Working Paper, Harvard Business School, 1990.

HACKMAN, R., J., AND G. OLDHAM. *Work Redesign.* Reading, Mass.: Addison-Wesley, 1980.

HAMEL, G., AND C. K. PREHALAD. "Do You Really Have a Global Strategy?" *Harvard Business Review* (July/August 1985): 139–48.

HELLRIEGEL, D.; J. W. SOCUM; AND R. W. WOODMAN. *Organizational Behavior.* St. Paul: West Publishing Co., 1989.

HOLMSTRÖM, B. "Moral Hazard and Observability." *Bell Journal of Economics* 10, no. 1 (Spring 1979): 74–91.

HOLMSTRÖM, B., AND P. MILGROM. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55 (1987): 303–28.

______. "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization* 7 (Special Issue, 1991): 24–52.

ITOH, H. "Incentives to Help in a Multi-Agent Situation." *Econometrica* 59, no. 3 (1991): 611–36.

JACOBS, M. *Short-Term America: The Causes and Cures of Our Business Myopia.* Boston: Harvard Business School Press, 1991.

JENNERGREN, P. "On the Design of Incentives in Business Firms—A Survey of Some Research." *Management Science* (February 1980): 180–201.

JENSEN, M., AND K. J. MURPHY. "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98, no. 2 (1990): 225–64.

KANTER, R. M. "The Middle Manager as Innovator." *Harvard Business Review* (July/August 1982): 95–105.

LAFFONT, J.-J., AND D. MARTIMORT. "Collusion and Delegation." Mimeographed ms., 1995.

LAZEAR, E. "Salaries and Piece Rates." *Journal of Business* 59, no. 3 (1986): 405–31.

LEWIS, T., AND D. SAPPINGTON. "Choosing Workers' Qualifications: No Experience Necessary?" *International Economic Review* 34, no. 3 (August 1993): 479–502.

MAGEE, R. "Aggregate Measures of Performance." Working paper, Northwestern University, 1988.

MERCHANT, K. A. *Rewarding Results: Motivating Profit Center Managers.* Boston: Harvard Business School Press, 1989.

MURPHY, K. J. "Incentives, Learning, and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts." *Rand Journal of Economics* 17, no. 1 (Spring 1986): 59–76.

NAGARAJAN, N.; K. SIVARAMAKRISHNAN; AND S. SRIDHAR. "Managerial Entrenchment, Reputation and Corporate Investment Myopia." *Journal of Accounting, Auditing and Finance* 10, no. 3 (1995): 565–85.

NARAYANAN, M. P. "Managerial Incentives for Short-Term Results." *Journal of Finance* 60, no. 5 (1985): 1469–84.

OHMAE, K. "Companyism and Do More Better." *Harvard Business Review* (January/February 1989): 125–32.

RICART I COSTA, J. "Managerial Task Assignment and Promotions." *Econometrica* 56 (1988): 449–66.

SAHLMAN, W., AND M. E. BARRETT. "Del Norte Paper Company." Harvard Business Case no. 9–177–034, 1976.

SAPPINGTON, D. "Limited Liability Contracts between Principal and Agent." *Journal of Economic Theory* 29 (1983): 1–21.

SCITOVSKY, T. *Human Desire and Economic Satisfaction.* New York: New York University Press, 1986.

SIBSON & COMPANY. *1990 Annual Executive Compensation Report.* New York: Sibson, 1990.

______. *Trends and Issues in Rewarding Executive Performance.* New York: Sibson, 1991.

SPAULDING, L. B. "General Motors' Component Set Strategy: A Critical Assessment of Knuckle Commonization Capability." M.S. thesis, Massachusetts Institute of Technology, 1991.

STEIN, J. "Takeover Threats and Managerial Myopia." *Journal of Political Economy* 96, no. 1 (1988): 61–90.

STIGLITZ, J. "Monopoly, Nonlinear Pricing, and Imperfect Information: The Insurance Market." *Review of Economic Studies* (October 1977): 407–30.

THAKOR, A. V. "Game Theory in Finance." *Financial Management* (Spring 1991): 71–94.

TIROLE, J. "Hierarchies and Bureaucracies: On the Role of Collusion in Organization." *Journal of Law, Economics, and Organization* (Fall 1986): 186–214.

"Titanic Battle: GM Plans Offer to Pay $700 Million to Buy Out Its Critic H. Ross Perot." *Wall Street Journal* (December 1, 1986).

WEISS, A. "Job Queues and Layoffs in Labor Markets with Flexible Wages." *Journal of Political Economy* 88, no. 3 (1980): 526–38.