objectives among managers and focuses instead on the problem of coordinating the decisions of several imperfectly informed actors. However, unlike classical team theory, we concentrate on the choice by managers of what to know, as well as what to do, and we allow the possibility that bounded rationality limits the managers' abilities to understand subtle messages.

Management is fundamentally about decision making, and a useful theory of management must come to grips with the bounds on human rationality. If individual managers had unlimited access to information that they could process costlessly and instantaneously, there would be no role for organizations employing multiple managers. On the other hand, once one acknowledges that individual managers are limited in their rates of information processing but can choose how to allocate their attention, the advantage of sharing responsibility for management when there is a time constraint becomes clear: The organization can bring more attention and information to bear on its decisions.

To take advantage of the information processing potential of a group of managers, it is necessary to have the managers attend to different things. But these differences are themselves the major cause of failure of coordination among the several managers. Our analytic perspective attempts to explain certain characteristics of organizations as the results of the desire for information specialization, the need to coordinate the activities of diverse managers, and the tension between these two objectives.

In Section II, we formulate a model of an organization composed of managers with limited attention and derive some general propositions. First, optimal organizations always direct managers to acquire different information. Second, there are superadditive returns to ability, where ability is defined to be a vector of the manager's speed in processing various kinds of information. Third, we find that even when ability is one dimensional, there can be no general supposition that the "ablest" manager will be placed at the top of a hierarchy. However, we do provide a sufficient condition for this conclusion when there is serial processing of information so that delays imposed by higher-level managers are propagated more broadly throughout the organization.

Bounds on managerial rationality are introduced in Section III. These allow us to explain the use of "commands" and idiosyncratic objectives within organizations—important phenomena which are inconsistent with traditional team models. For example, in the Marschak and Radner (1972) model, the *sole* function of communication is to pass information among managers: There is never any reason for a manger to restrict the feasible set of another manager or to specify the objective he should pursue, for a subordinate with unlimited attention and calculating power who knows the (optimal) strategy of his superior could infer the set of possible justifications for any instruction he receives and would be led to follow those instructions voluntarily. In our model, in which managers cannot costlessly and instantaneously extract information from other data sources, an assumption that the subtle content of messages is costlessly processed would be out of place. Instead, we assume that only the surface content of a message like "produce 100 widgets" can be grasped costlessly; the subtler content, which depends on drawing an inference from the message using knowledge of the sender's decision rule, can be inferred only at a cost.

The behavior of managers in a hierarchy with commands is the subject of Section IV. We consider a chief executive allocating production targets, capital, and other resources to division managers, who in turn reallocate the budgeted items to their subordinates, etc., until the resources and targets reach the "shops" where production takes place. The model we use employs quadratic cost functions and an information gathering technology based on sampling from normal distributions. Among our findings are (i) that managers at each level optimally focus attention only on those variables that determine the marginal productivity of resources and the marginal costs of production in the units under their command, (ii) that there is a limit to the depth of optimal hierarchies (even though parallel processing of information is used), and (iii) that firms with more prior information about parameters and more highly refined information systems will employ less able managers, or give their managers wider spans of control, or both.

In Section V, we review the related literature and offer concluding remarks.

II. The Basic Model and Some Simple Propositions

II.1. *Limited Managerial Attention*

We consider a manager with time τ to allocate to a variety of information sources. If he allocates time τ_i to source i, he acquires information represented by the partition $I^i_{\alpha_i\tau_i}$, where α_i is a parameter of the manager's efficiency for processing information of this type and I^i_s becomes increasingly fine as s increases. If there are N information sources (where N may be any positive integer or $+\infty$), then the manager's *ability* is $\alpha = (\alpha_1, \ldots, \alpha_N)$. The manager's information at time τ is then the join of his information from each information source. Hence, the set of feasible information partitions for a manager of ability α and time τ is

$$I_{\alpha\tau} = \left\{ I^1_{\alpha_1\tau_1} \vee \cdots \vee I^N_{\alpha_N\tau_N} \,\middle|\, \sum_i \tau_i \le \tau \right\}. \tag{1}$$

This model of managerial attention is the analytical focus of the paper. In particular, in all that follows it is assumed that acquiring information is the *only* time-consuming activity.

II.2. *Some Simple Propositions*

In this section, we imagine an organization with a given collection of decisions. Each decision i must be undertaken no later than its deadline t_i. Also given is a managerial wage function $w(\alpha, \tau)$ specifying the cost of employing a manager of ability α who works τ hours per week. The organization problem is then to decide (1) which managers to employ, (2) how to assign decisions to managers, (3) how to divide information processing tasks, and (4) when to make the decisions. (A decision may optimally be made early if it affects how other managers should allocate their attention.) For the simple propositions that we develop here, it suffices to note that if in organization design A each decision is made by a better-informed manager than in organizational design B, and with the same total wage bill, then the payoff to design A is higher than the payoff to design B.

PROPOSITION 1 (Necessity of asymmetric information). *If $w(\alpha, \tau) > 0$ for $\alpha > 0$, $\tau > 0$, then each manager who is hired must make at least one decision at some time t at which no other manager has finer information.*

The proof is obvious: Otherwise the organization could do better by laying off the manager, saving the wage, and assigning his decisions at each time t to a manager who is better informed at that time. Note that we also use our hypothesis that information processing is the only time-consuming activity (otherwise it might pay to hire an extra manager who knew nothing more in order to speed up implementation).

The significance of Proposition 1 lies in the extension of this work to situations in which incentives may not be in complete harmony. In order to take full advantage of an employed manager's information processing capacity, the firm must allow him to become asymmetrically informed at least some of the time. But this may permit him to pursue his own objectives at the expense of the firm's goals.

PROPOSITION 2 (Superadditive returns to ability). *Suppose that an organization hires managers of ability α and β who each work τ hours per week. Then the wage function satisfies $w(\alpha, \tau) + w(\beta, \tau) \le w(\alpha + \beta, \tau)$. Similarly, if the organization hires two managers of ability α who work τ and τ' hours per week, respectively, then $w(\alpha, \tau) + w(\alpha, \tau') \le w(\alpha, \tau + \tau')$.*

Proof. A manager of ability $\alpha + \beta$ could acquire at each point of time (and thus at all of the two managers' deadlines) the same information as managers α and β together. Thus the manager of ability $\alpha + \beta$ would be hired in place of the two unless his wage was higher than the sum of the wages of the slower managers. (Note that in general the information acquired by the fast manager in one of the old jobs will improve his decision in the other job and thus organizational utility leading to a strict inequality.) The second part of the proposition is proved in the same way. ∎

There are two parts to Proposition 2, the first identifying a form of increasing returns to ability and the second explaining the resistance of organizations to hiring part-time managers at only proportionally reduced wages, the predilection of many managers to work very long hours, and the tendency for managers to work full time for a single firm rather than part time for several.

The proof of Proposition 2 makes use only of the obvious and simple fact that, holding the assignment of managers to jobs fixed, it is better to have managers who work faster and reach decisions earlier. But there are additional advantages to replacing several managers with a single one. Combining tasks under a single manager improves the coordination between the tasks. Also, there may be an opportunity for further improvement as the single manager fine-tunes his choice of an information system, receives more information before beginning each task, and avoids the almost inevitable duplication in information processing that accompanies the use of multiple managers. The effect of improved coordination among tasks is illustrated in the extended example studied in Section IV, in which there are superadditive returns to time and ability despite diminishing returns to information processing *in each job*.

Our third proposition concerns how managers are assigned to jobs. Is there any sense in which it pays to assign abler managers to higher-level jobs? Since ability is multidimensional, one cannot speak generally about assigning the ablest manager to the top job. Even if there is a manager with an absolute advantage in every kind of information processing, he may have a comparative advantage in processing information that is most relevant to "lower-level" jobs. Also, if a higher-level manager is defined to be one who controls the resource allocation among lower-level managers, then it is easy to construct examples in which the returns to ability are lower in higher-level jobs, because the information processing required by these jobs is very easy (or so difficult that even the ablest manager cannot process enough information to make much difference). For example, the people who allocate research funds may be less able (in an optimal organization) than the researchers who receive the grants.

There is one case, though, in which one may expect the best managers to be assigned to the "top" jobs in a hierarchy. This is the case where the decisions to be made are such that effective information processing for lower-level decisions cannot begin until higher-level decisions are made; it is the kind of situation to which Simon alluded in the opening quotation of this paper.

To formalize this intuition, suppose that decision nodes are arranged in a tree and that information processing must be done *serially;* that is, information processing for decision i cannot begin until all decisions at preceding nodes have been made. In our next proposition we hold hours per week constant, suppressing τ_i from our notation, and focus on the number of weeks t_i managers are assigned to a project. We suppose that the weekly wage is constant, so that we write $W(t_i, \alpha_i, \tau_i) = t_i w(\alpha_i, \tau_i) = t_i w(\alpha_i)$.

PROPOSITION 3. *Suppose that ability α is one dimensional and that the cost of the time spent by the manager at any node i is $t_i w(\alpha_i)$, where t_i is the time spent, α_i is the manager's ability, and w is a strictly convex function. If all information processing must be done serially and decision i precedes decision j then, at the optimum, $\alpha_i \geq \alpha_j$.*

Proof. If i precedes j and $\alpha_i < \alpha_j$, then the firm can do better by replacing each manager by a manager of ability $\bar{\alpha} = (\alpha_i t_i + \alpha_j t_j)/(t_i + t_j)$, letting the manager spend time $\bar{t}_i$ in the top job and $\bar{t}_j$ in the other job, where $\bar{\alpha}\,\bar{t}_i = \alpha_i t_i$ and $\bar{\alpha}\,\bar{t}_j = \alpha_j t_j$ and $\bar{t}_i + \bar{t}_j = t_i + t_j$. Note that in fact $\bar{t}_i < t_i$, so that no decision is reached later, and at least one decision is reached earlier. Moreover, the wage bill is smaller, since by convexity, $(\bar{t}_i + \bar{t}_j)w(\bar{\alpha}) < t_i w(\alpha_i) + t_j w(\alpha_j)$. ■

Observe that in Proposition 2 we showed that the wage function must be superadditive in ability, holding τ constant. Convexity is a special case of superadditivity (provided that $w(0) = 0$). We can gain some additional insight into the problem by looking more closely at the marginal value of time at various levels of a hierarchy of managers doing serial processing.

Consider a variation from the optimal strategy in which manager M reduces the time he spends by Δ while each of M's direct subordinates spends Δ more time. Such a change is always feasible, so the marginal product of manager M's time minus the wage paid for that time cannot be less than the sum of the corresponding marginal products of time (net of wages) for M's direct subordinates. For example, if the hierarchy is symmetrical and each manager has at least two direct subordinates, then the marginal value of time (net of wages) at least doubles with each step up the hierarchy. This may help to explain the extensive use of time-saving devices and perquisites such as sophisticated decision support systems, executive assistants, mobile telephones, company jets, and chauffeured limousines for managers and executives at the top of a hierarchy.

III. COMMANDS IN TEAMS

Under traditional models of "rational" decision making, a key part of the specification is that a rational decision maker can adopt *any* decision strategy that depends only on what he knows. In these models, an optimal team strategy will have each manager maximizing the expected payoff of the organization, given the information he has acquired and the signals he has received when he makes his decision.

Thus if manager i must choose $d_i \in D_i$, he solves

$$\max_{d_i \in D_i} E[\pi(\omega, d_{-i}(\omega), d_i) \mid I_i, \sigma_i] \tag{2}$$

where

π is the organizational payoff
ω is the state of the world
d_{-i} are the decisions made by other managers
I_i is the information the manager has acquired directly
σ_i is the set of signals the manager has received from other managers.

From the point of view of manager i, the decisions made by others in the organization are random variables, because they are functions of their information. Equally, from the manager's point of view, the signals he receives are observed random variables, because they are functions of the information of those sending the signals. All of the manager's information appears behind the conditioning sign in problem (2). The inclusion of σ_i as part of i's information about ω in (2) amounts to an assumption that i can costlessly and instantaneously infer the significance of the signals communicated to him by other managers. In a model such as ours, where information processing is explicitly time-consuming, and where i may have the alternative of learning the same information by looking directly at some report in the database, this assumption is suspect.

What is most remarkable about expression (2) is that, in an optimal team strategy, *there is no role for "instructions" from any manager to any other*. That is, at an optimum, a superior may communicate information to his subordinate, but he never limits the set of actions that the subordinate may undertake, nor does he directly set the objective the subordinate pursues.

To put the same point slightly differently, when communication consists of orders and when the requirement to follow orders does not degrade the optimal performance of the team, then the managers can infer from the orders themselves that it is optimal to obey: *Optimal orders convey their own justification*. However, when managers are not perfectly adept at interpreting communications, there can be a separate role for instructions that limit the manager's choice set.

Consider an example in which we imagine that there are three productive units or shops, which are each capable of producing a single output. The cost of producing x units in shops 2 or 3 is $\frac{1}{2}x^2$, while in shop 1 it is $\frac{1}{2}(x - \gamma)^2$, where γ is a random variable. Assume that manager M allocates output responsibility between shops 1 and 2, while the top manager T allocates output responsibility between manager M and shop 3. Suppose that the total output which must be produced is x_T, that is, a large penalty must be paid by the team if output falls short. Then the (first-best) cost-minimizing output assignments to the shops are $x_1 = (x_T + 2\gamma)/3$ and $x_2 = x_3 = (x_T - \gamma)/3$.

Now suppose that both managers know x_T, but only manager T knows γ. Plainly, with "full rationality," T can set $x_3 = (x_T - \gamma)/3$ and assign the output target $x_M = (2x_T + \gamma)/3$ to his subordinate M, and M can then infer the value of $\gamma = 3x_M - 2x_T$. In this way, the cost-minimizing allocation can be attained. We emphasize that even if manager M were not constrained to set $x_1 + x_2 = x_M$, he would choose to do so given the team objective and the inference he can make from the target x_M communicated by T. But what happens if M cannot make such subtle inferences costlessly, or if the time cost of deducing γ from the instructions is so high that it is optimal for the manager to ignore that information?

The organization can sometimes overcome this bounded rationality problem by giving idiosyncratic subgoals to the managers, that is, by instructing them to pursue objectives different from the organization's overall objective. In the present example, if the manager M is instructed to minimize the objective $(x_1 + 2x_2 - x_T)^2$ subject to meeting the output target $(x_1 + x_2 = x_M)$, he will always choose the first-best allocation despite his inability to infer γ from the output target x_M.

It is unclear how effective this device of resetting objectives can be. First, it requires that the organization designer be able to anticipate what the "subtle content" of messages will be, in order to correctly specify modified objectives. Second, if managers are mobile and objectives cannot be instantaneously learned and unlearned, then distorting managerial objectives may entail significant costs. Finally, artificial objectives that restore first-best behavior in general are quite complex. In view of these difficulties, we henceforth make the extreme assumption that managers seek to minimize the expected costs of their units, given the actual cost functions and the manager's knowledge about relevant parameters.

IV. Hierarchies—An Example with Limited Managerial Attention

Our purpose in this section is to study the contribution of a complex management organization to the efficient allocation of limited resources. To simplify our task, we restrict our attention to hierarchies in which each

manager receives instructions and resources from just one superior. We analyze a quadratic example in which we are able to answer specific questions:

(a) What are the key trade-offs in designing an optimal hierarchy?

(b) How can one measure the contribution of a manager?

(c) How is the optimal management organization (e.g., the span of control) affected by the degree of uncertainty in environment?

(d) What limits the height of the hierarchy?

Suppose that the decision to be made is one of setting production targets and allocating resources among shops $t \in T$ in an environment without externalities. The set of resources available and the organization's output requirements are given by the vector $x_T \in \mathbb{R}^k$. All production takes place in the shops; the job of the higher levels of the hierarchy is simply to direct resources to their most productive use, that is, to assign the prespecified resources and production targets to minimize total cost.

We assume that the costs incurred by a shop as a function of its output requirement and the resources supplied by higher-level management are expressible as a quadratic form:

$$C_t(x, \gamma_t) = \tfrac{1}{2}[(x - \gamma_t)'B_t(x - \gamma_t) - \gamma_t'B_t\gamma_t]. \qquad (3)$$

In this model, the vector parameters $\{\gamma_t\}$ of the shops are initially unknown. The other technological parameter B_t of shop t is a positive definite matrix which is known *a priori* to all members of the team. For the case $k = 1$, this means that each shop has a linear marginal cost function with known slope and unknown x intercept. The constant term in (3) has been chosen so that $C_t(0, \gamma_t) = 0$; thus we interpret C_t as being a variable cost function.

A hierarchy is a tree $(H, <)$, that is, a collection of managers (nonterminal nodes) and shops (terminal nodes) and a precedence relation which specifies the lines of authority and communication. When one manager M precedes another manager M' in the hierarchy $(M < M')$ we say that M' is a *subordinate* of M. The set of *direct subordinates* $S(M)$ consists of the immediate successors of M in the tree. M's *boss* is his immediate predecessor $P(M)$ in the tree.

Each manager M except the chief receives a quantity signal from his boss $x_M \in \mathbb{R}^k$ specifying how much to produce and with what centrally allocated resources and in turn passes instructions and resources (x_s) $(s \in S(M))$ to his direct subordinates to satisfy the constraint $\Sigma_{s \in S(M)} x_s = x_M$. The organization's objective is to minimize expected total production costs, subject to meeting the prespecified output requirement and using the prespecified resources.

Now we introduce the "database," time, and attention. We assume that by devoting time τ_t to studying the affairs of shop t, a manager of ability $\alpha = (\alpha_1, \ldots, \alpha_t, \ldots)$ observes a statistic distributed like

$\gamma_t + \varepsilon_t/\sqrt{\alpha_t\tau_t}$, where ε_t has a k-variate normal distribution with mean zero and variance S_t. In effect, we have assumed that the size of the sample observed is proportional to the attention devoted to any shop.[1] Each manager has time $\bar{\tau}$ available to spend processing information.

We assume, as explained in Section III, that it is cheaper for a manager to examine his database directly than to infer information from his instructions. (Formally, let the component of α corresponding to the instructions information source be zero.) We also assume that each manager aims to minimize the expected total costs of the organization or—equivalently since there are no external effects among units—the expected costs incurred by his unit, given his information, resources, and production target. As noted earlier, this may not be optimal for an organization that is free to manipulate its manager's objectives.

Several things are easily seen in this example. First, the need for several managers arises from the limited ability of any single manager to estimate the marginal costs at many shops within the allowed time. Second, low-level managers will optimally choose to limit their attention to the variables that affect only the small portion of the organization which they manage; they will therefore fail to notice opportunities for cost-saving transfers of resources among themselves. This gives rise to our third observation, which is that one useful role of high-level management is to recognize these missed opportunities and take advantage of them. Fourth, there is always some gain to coordinating activities at a high level, because there are always opportunities that lower-level managers with their specialized information will fail to perceive. As we see, however, when we optimize over the form of the hierarchy, there may be diminishing returns and increasing costs to high-level management.

We analyze the organization problem in three stages. First, holding the hierarchy and the manager's information fixed, we compute the optimal decision rules and the resulting payoffs. Then, we analyze how managers optimally allocate their attention. Finally, we make the hierarchy itself endogenous. For hierarchies with fixed information, it is possible to calculate explicitly the optimal team strategy. The calculation exploits the facts (established below) that with quadratic costs (and ignoring nonnegativity constraints) (i) the information of a manager affects his unit's expected fixed costs, but not its expected marginal costs; (ii) consequently, (a) a manager optimally makes the same allocation to his subordinate managers no matter what their information systems, and (b) the savings achieved by any manager M with any given information system does not depend on the output target assigned; and (iii) one can express the savings attributable to management as the sum of the reductions in expected fixed costs achieved by the individual managers in the hierarchy, where each term in the sum depends only on the corresponding manager's information system. This last observation is important, because it makes it possible to study separately the choice of information systems (that is, the allocation of attention) by managers in each job. Also important is observation (iib), since it justifies our assumption that team managers process information in parallel.

To conduct much of the analysis of the quadratic model, we need to define three constructs, as follows:

$$\gamma_M = \sum_{s \in S(M)} \gamma_s = \sum_{t > M} \gamma_t, \tag{4}$$

$$B_M = \left[\sum_{s \in S(M)} B_s^{-1}\right]^{-1} = \left[\sum_{M < t \in T} B_t^{-1}\right]^{-1}, \tag{5}$$

and

$$\gamma_s^M = E[\gamma_s \mid I_M].^2 \tag{6}$$

Now define the term *service* to mean a unit in the hierarchy consisting of a collection of shops. Consider the problem of a service manager M. Suppose that the manager has information I_M and is asked to produce the vector X_M. Then, he will allocate resources and assign output responsibility to minimize the quantity

$$E\left[\sum_{s \in S(M)} C_s(x_s, \gamma_s) \mid I_M\right] \tag{7}$$

subject to the constraint

$$\sum_{s \in S(M)} x_s = x_M \tag{8}$$

A similar model was studied by Cremer (1980), who established a variant of the following proposition:

[1] This specification can be generated exactly by specifying that a manager of ability α devoting attention τ_t to shop t observes the path of a Brownian motion with unknown drift γ_t and instantaneous variance S_t over the internal $[0, \alpha_t\tau_t]$. A manager who devotes less effective attention $\alpha_t\tau_t$ thus has strictly less (that is, coarser) information than another manager who devotes more.

[2] This means that the random variable $\gamma_s^M(\omega) = E[\gamma_s \mid I_M(\omega)]$.

PROPOSITION 4. *The solution to manager M's problem is* $x_s = \gamma_s^M + B_s^{-1}\lambda_M$, *where* $\lambda_M = B_M(x_M - \gamma_M^M)$. *The conditional variable cost function for M is given by*

$$C_M(x_M \mid I_M) = E\left[\sum_{s\in S(M)} (x_s - \gamma_s)'B_s(x_s - \gamma_s) - \gamma_s'B_s\gamma_s \mid I_M\right] \tag{9}$$
$$= (x_M - \gamma_M^M)B_M(x_M - \gamma_M^M) - \sum_{s\in S(M)} \gamma_s^{M\prime}B_s\gamma_s^M.$$

Proof. Substituting the form of the cost function (6) into the objective (7) leads to a quadratic minimization with a linear constraint (8), whose solution is routine. ■

Let $\bar{\gamma}_s = E[\gamma_s]$ denote the prior expectation of γ_s. According to Proposition 5, the minimum cost that could be achieved if there were no information available for the allocation decision (apart from the information reflected in the prior beliefs) is given by

$$(x_M - \bar{\gamma}_M)'B_M(x_M - \bar{\gamma}_M) - \sum_{s\in S(M)} \bar{\gamma}_s'B_s\bar{\gamma}_s. \tag{10}$$

The *expected savings attributable to the management at M* is defined to be the excess of the zero-information minimum expected cost given by (10) over the expected cost $E[C_M(x_M \mid I_M)]$ incurred with information I_M. Using the pair of identities

$$E[(\gamma_s - \gamma_s^M)'B_s(\gamma_s - \gamma_s^M) \mid I_M] = \mathrm{tr}[B_s\,\mathrm{Var}(\gamma_s \mid I_M)] \tag{11}$$

and

$$E[\mathrm{Var}(\gamma_s \mid I_M)] = \mathrm{Var}(\gamma_s) - \mathrm{Var}(\gamma_s^M), \tag{12}$$

one obtains the following representation of the product of management, which is a variant of another result of Cremer (1980).

COROLLARY 1. *The expected savings attributable to management at M is*

$$-\mathrm{tr}(B_M\,\mathrm{Var}(\gamma_M^M)) + \sum_{s\in S(M)} \mathrm{tr}(B_s\,\mathrm{Var}(\gamma_s^M)). \tag{13}$$

Using Proposition 4 and Corollary 1 as building blocks, one can construct a cost function for the entire organization showing the savings attributable to management at all levels.

For this we need to recall the definition of a sufficient statistic.

DEFINITION. I is sufficient in (I, J) for a random variable ξ if for all subsets A of the range of ξ

$$P\{\xi \in A \mid I, J\} = P\{\xi \in A \mid I\}.$$

In the following proposition (only), we assume that each manager knows at least as much about the costs of his unit as his superior does. This restriction is reasonably well justified when a high-level manager bases his opinions only on summary statistics or aggregate information about low-level units, while the lower-level manager pays attention to finer details. It may also be justified when higher-level managers base their opinions on executive summaries of reports about low-level units, on shop tours guided by the lower-level managers, and on reports that the lower-level managers prepare.

PROPOSITION 5. *Suppose we are given a hierarchy H and information* I_M *for each manager M, with the property that* I_M *is sufficient in* $(I_M, I_{P(M)})$ *for* $(\gamma_s : s \in S(M))$. *Then, the optimal team strategy is for each manager M to choose* $(x_s : s \in S(M))$ *according to* $x_s = \gamma_s^M + B_s^{-1}\lambda_M$, *where* $\lambda_M = B_M(x_M - \lambda_M^M)$. *The expected total cost incurred by the hierarchy when the output target is* x_T *is the zero-information minimum expected cost for the entire hierarchy, given by*

$$(x_T - \bar{\gamma}_T)'B_T(x_T - \bar{\gamma}_T) - \sum_{s\in S(M)} \bar{\gamma}_i'B_i\bar{\gamma}_i \tag{14}$$

minus the sum of the expected savings attributable to management at each node of the hierarchy:

$$\sum_{M\in H\setminus T} [-\mathrm{tr}(B_M\,\mathrm{Var}(\gamma_M^M)) + \sum_{s\in S(M)} \mathrm{tr}(B_s\,\mathrm{Var}(\gamma_s^M))]. \tag{15}$$

Remark. Note that Proposition 5 refers to the *optimal* team strategy, without regard to restrictions on the objective or inferences of the managers throughout the hierarchy. The assumption that I_M is sufficient in $(I_M, I_{P(M)})$ for manager M's problem implies that any inability to draw inferences from x_M about $I_{P(M)}$ does not affect M's decision problem.

Proof. The first part of the Proposition follows by inductive application of Proposition 4 to demonstrate that the cost function at every position M in the hierarchy is a quadratic form with unknown parameters $\{\gamma_s; s \in S(M)\}$. The second part of the Proposition then follows from Corollary 1. ■

Proposition 5 gives some of the flavor of the organization design problem. Regardless of the form of the hierarchy, if information at all levels of management were perfect, the first-best outcome would be achieved. In terms of Proposition 5, $\gamma_s^M = \gamma_s$ for all M and s, and the savings attributable to management at all levels would add up to the excess of the zero-information expected cost over the full-information expected cost. The organization design determines how much of total potential savings can be achieved at each management node. For example, by arranging higher-level units s so that $\mathrm{Var}(\gamma_s)$ is small, one ensures that there is little potential gain to management at the highest levels and the key to good performance becomes effective management at the lower levels of the hierarchy. When information processing at the highest levels is especially costly, as in the serial information processing case studied earlier, or when the information available for making high-level decisions tends to be poor, it may pay to organize the hierarchy so that high-level decisions based on poor information are not too damaging to the organization's performance. On the other hand, when the kind of information available allows fast and effective high-level decision making and when talented decision makers are at a premium, the organization can be structured so that low-level decisions based on poor information are not too damaging. Of course, the precise determination of the optimal hierarchy depends on the information technology and the ability levels of the managers who are available (or the market wages of managers in different ability classes).

Note that the savings attributable to managers does not depend on the organization's output target x_T. Also, the summand corresponding to M is a term that depends only on the information I_M. Therefore, for a fixed hierarchy, the problem of choosing information systems optimally for all the managers is solved by maximizing the individual summands, provided that the solutions to these problems satisfy the "sufficiency condition" of Proposition 5. In that case, the original problem decomposes into the problems of choosing specialized information systems optimally for each individual manager separately. To develop this idea, we make the following additional assumption:

Assumption. The γ_t's are one dimensional ($k = 1$) and independently and normally distributed with prior variance $\sigma_t^2 = 1/r_t$.

It is convenient analytically to work with *precision* (the inverse of the variance) r_t, rather than with the variance itself. A manager may devote attention to observing any of the γ_t's. The information technology described earlier can now be restated as follows: The precision of the manager's observation of any γ_t is proportional to the time spent observing γ_t, where the constant of proportionality α_t may reflect both the manager's ability and the quality of the information system at the manager's disposal. A service manager's time allocation problem is then to choose times τ_t^M to devote to observing each γ_t in order to maximize the expected savings

$$\sum_{t \in S(M)} (B_t - B_M)\left[\frac{1}{r_t} - \frac{1}{r_t + \alpha_t \tau_t^M}\right] \tag{16}$$

subject to the constraints $\tau_t^M \geq 0$ and $\Sigma_{t \in S(M)}\, \tau_t^M \leq \bar{\tau}$.

Proposition 6. *The solution to the manager's allocation of the attention problem is characterized by a number λ such that each τ_t^M is the maximum of zero or the solution to*

$$r_t + \alpha_t \tau_t^M = [\alpha_t (B_t - B_M)/\lambda]^{1/2}. \tag{17}$$

Proof. This is a linearly constrained concave maximization problem and λ is the Lagrange multiplier of the total time constraint. ■

In studying formula (17) one sees that, in deciding how informed to become about the situation at a shop, the manager will weigh the sensitivity of the shop's costs to the allocation (measured by $B_t - B_M$) against the difficulty of gathering information about the situation (measured by $-\alpha_t$). The manager does not try to deepen his knowledge about the shops where his prior information is greatest, but instead seeks a target level of knowledge about each shop, based on the shop's characteristics. Evidently, a manager will optimally gather more information about a particular shop t if he or she has invested in increasing α_t, that is, in making is cheaper to acquire information, but will gather less additional information if the prior information is better.

Corollary 2. *In the fully symmetric case with $B_t = B$, and $r_t = r$ for all t, the optimal solution is $\tau_t^M = \bar{\tau}/n$, where n is the number of shops in the service. Let $\beta = \alpha\bar{\tau}/r$. The savings attributable to the manager M is $(n - 1)B\beta/[r(n + \beta)]$.*

These savings, regarded as a function of $(n, \bar{\tau})$, display increasing returns to scale. Thus, when two "half-time" managers each managing units of size n are replaced by a "full-time" manager managing a unit of size $2n$ (formally, this was accomplished by doubling $\bar{\tau}$), the expected cost savings are more than doubled. The economies achieved in this example reflect, not the possibility of reallocating the manager's limited attention in a better way, which led to the general superadditivity result of Proposition 2, but rather the better coordination that can be obtained when the hierarchy is modified to extend the authority of the manager, allowing him or her to reallocate resources across more shops.

Now consider the situation with all managers equally able. Suppose that the depth of the organization is limited to one level, so that the services are coordinated on the basis of prior information only. How large should each service be? That is, what is the optimal span of control? Mathematically, the problem can be expressed as maximizing the average savings per shop, net of wage costs. Using Corollary 2, the problem is

$$\text{Maximize}_{n \geq 1} \; \psi(n|r,\beta,B,w) = \frac{1}{n}\left[\frac{(n-1)B\beta}{r(n+\beta)} - w\right], \tag{18}$$

where w is the managerial wage. We impose the natural assumptions that w, B, β, and r are nonnegative.

Proposition 7. *If $w \geq B\beta/r$, the optimal service service size is $+\infty$ (no managers are hired). Otherwise, the optimal service size is an integer within one of*

$$N(\kappa,\beta) = \beta\frac{1+\kappa}{\beta-\kappa}\left(1 + \sqrt{1 + \frac{\beta-\kappa}{1+\kappa}}\right), \tag{19}$$

where $\kappa = wr/B$. Moreover, N is increasing in κ and first decreasing then increasing ("U-shaped") in β.

Proof. One can verify that, when n is treated as a continuous variable, the problem is strictly quasi concave in n. Hence, the optimal integer solution is an integer adjacent to the optimal continuous solution (19), which can be found as the unique positive solution to the first-order condition. To show that (19) is U-shaped in β is tedious, but proceeds in outline as follows. Reexpress (19) as a function of $\delta = [1 + (\beta - \kappa)/(1 + \kappa)]^{1/2}$. Since δ is increasing in β on $\beta > \kappa$, it suffices to show that (19) is U-shaped in δ for $\delta > 1$. Differentiate (19) with respect to δ. The derivative is negative for δ near one, positive for δ large, and has just one zero on the range $\delta > 1$. ∎

Intuitively, parameter β represents a manager's ability to gather information as a fraction of the amount of information (precision) that is already embedded in the prior distribution. Parameter κ is a cost parameter; it is

proportional to the wage w and inversely proportional to B/r, which measures the marginal value of information in managing a service. The approximate solution N is increasing in κ—when wages rise relative to the marginal value of information, it is optimal to employ fewer managers. Also, the approximate solution is U-shaped in β, first decreasing and then increasing (provided that $\kappa > 0$). When managers are quite unproductive (β small), the firm economizes on wages by reducing the number of managers employed, so n is large. Indeed, when $\beta \leq \kappa$, it is optimal to set $n = \infty$, that is, to hire no managers at all, in order to avoid any wage costs. When managers are highly productive, a few managers can effectively capture almost all the possible savings, so that once again it is optimal to employ few managers, and the optimal n is large. Only when managers are of "moderate" productivity, able to make significant contributions but unable to manage very large units effectively without assistance, will the size of the optimal service be moderate and the number of managers hired be large.

We now turn our attention to the following questions: In what environments is talented management most valuable? and How do characteristics of the environment affect the optimal span of control?

Proposition 8. *For any fixed service size n, the marginal value of a service manager's ability is a decreasing function of the prior precision r: $\partial^2\psi/\partial\beta\partial r \leq 0$ (where ψ is defined by (18)).*

Now suppose as before that the rate of information processing α is proportional to the manager's ability, but let the constant of proportionality depend on the quality of the information system. Then, we may ask how improvements in the information system that supports managers affect the value of ability to the organization. The answer is given by the following proposition.

Proposition 9. *For any fixed service size n, the marginal value of a service manager's ability is a decreasing function of the quality of the information system: $\partial^2\psi/\partial\beta^2 \leq 0$. The approximate optimal service size given by (19) is increasing in the prior precision r.*

Suppose that firms operating in older, more stable industries have better prior information about their environments and more highly refined information systems for monitoring the environment than firms in newer or more rapidly evolving industries. According to Propositions 8 and 9, with the organization form held fixed, firms in newer and more rapidly evolving industries should employ more able managers than firms in older, more stable industries. Intuitively, there is more scope for good management to improve matters when information processing is more difficult and when there is less prior knowledge about the environment. Moreover, one can show that, with the quality of the managers held fixed, if the information systems are sufficiently good, then firms with better information systems should have larger services, that is, should employ fewer low-level managers per unit of output.

Our final exercise in the quadratic case is a particularly interesting one. We examine the value of high-level management using Proposition 5. The following calculation is illustrative. Suppose the organization consists of n shops in all, each with slope coefficient B and prior precision r about its intercept parameter. Further suppose that manager M has N_M direct subordinates, each of whom heads a subunit containing n_M/N_M shops. Finally, suppose that any manager's information comes from observing the details of the shops—there are no explicit aggregates available other than those that the manager constructs from observations.

In this model, it can be shown that each manager's time allocation problem is a symmetric, concave problem whose optimal solution specifies devoting equal time to studying the parameters of each shop $i > M$. Moreover, one can verify that $B_M = B/n_M$ for every manager M. Then, by using the formula for posterior variances associated with the normal sampling technology and applying Proposition 5, one obtains the following result.

Corollary 3. *The savings attributable to management at M when the manager's ability is α_M and the time available is τ is*

$$\frac{B\beta(N_m - 1)}{r(n_m + \beta)}. \tag{20}$$

Expression (20) tends to zero as the size of the units managed (n_M/N_M) tends to infinity. Hence, for any positive wage w, there is a limit to the size of the units that can be profitably incorporated into a larger organization. The particular form of (20) depends on our specification of the data available to high-level managers, particularly the absence of reliable aggregates. One can similarly show that even if there are aggregates available, if the aggregates are not "too good," meaning that they can do no more than enhance the top manager's rate of information processing by a fixed positive factor, then there will be a limit on the size of the units managed. This result emerges even though there are significant gains possible from coordination at all levels, and the only factor limiting the contribution of high-level management is its limited ability to assemble and process the information that would enable it to make a good decision.

V. Conclusion

There have been many previous studies of the economics of hierarchies, especially those seeking to analyze whether problems of coordination impose a limit on the size of an efficient firm. Williamson (1967) initiated the formal theory with a model in which the instructions given by the chief executive are distorted by a fixed amount as they pass through each succeeding level in the hierarchy. He concluded that the loss of control by the chief executive in large organizations limits the depth and size of an optimal organization. However, the analysis is biased by its implicit assumption that the manager of a subunit necessarily makes poorer decisions as a middle manager in a large organization than as chief of a smaller one.

Calvo and Wellisz (1978) studied a related control loss model based on the idea that managers who are not monitored sufficiently will shirk their responsibilities. Unlike Williamson, they found that this imposes no limit on the size of an optimal organization. Similarly, Beckmann (1977) analyzed a model in which the productivity of a worker depends (in part) on the amount of supervision he receives. In turn, the productivity of supervisors depends on the amount of supervision they receive. Continuing in this way, Beckmann concluded that the average cost of managing declines with the size of the hierarchy.

Keren and Levhari (1983) reached the opposite conclusion in a model where the productivity of the firm depends on the time it takes to reach a decision, which in turn depends on the structure of the hierarchy. The total time taken by each manager consists of a fixed setup time plus time proportional to the number of his or her immediate subordinates. Thus, on the one hand, it is costly to have wide spans of control, since this causes managers to reach decisions slowly; on the other hand, it is clostly to have many tiers in the hierarchy, since (i) there is a fixed time cost to each tier, and (ii) there are then more managers who need to be managed and paid. They found that the unit costs of coordination eventually increase as the firm grows. Unless scale economies are correspondingly large, this limits the size of the firm.

Rosen (1982) analyzed the wages and job assignments of managers in a hierarchy. His analysis assumes a production function in which the ability of the top-level manager has a more than multiplicative effect on the productivity of workers but is subject to diminishing returns with the size of the organization. He concluded that it pays to place the ablest manager in the organization's top job and that larger organizations will employ abler managers at the top.

None of the formal theories described above gives an explicit role to bounded rationality. Instead, each begins with either an incentive problem or a reduced form cost or production function that is meant to reflect some unmodeled aspect of management's limited decision-making capacity. In contrast, our theory takes the decision problem itself and the limits on managerial attention as primitive and offers competing answers to these commonly asked questions about organizations and hierarchies:

What are the functions of managers, organizations, and hierarchies? In our theory, managers make decisions to advance the organization's objectives. Organizations with multiple managers delegate decisions so as to bring more information to bear than any single manager could bring alone. Hierarchies compensate for the resultant loss of coordination among individual decision makers by assigning higher-level managers the task of coordinating the lower-level decisions.

Why are organizations averse to hiring half-time managers at only proportionately reduced salaries? The answer is because a full-time manager is more than twice as productive as two half-time managers. The full-time manager brings more information to bear on each decision than the half-timers do, avoids some duplication in information processing, and coordinates a wider range of activities without the need for higher-level intervention or communication with other managers. Managers who work long hours may enjoy similar more-than-proportionate productivity advantages over their less diligent co-workers.

When does the ablest manager belong at the top of the organization? Equivalently, when is the value of ability in the top job likely to be great? The answer is not always. But the value is great when the top-level decision involves serial processing between the chief and his subordinates, because then a "slow" chief causes costly delays to echo through the whole organization.

How does the form of the hierarchy and the kind of managers it employs depend on its environment? In the quadratic resource allocation example, talent in managers and smallness of services are both more valuable the greater the prior uncertainty and the more difficult the information processing job. One possible application of this principle is that firms in newer, more unsettled industries may employ abler managers or give them smaller spans of control than firms in older, more stable industries.

Are there unlimited economies of scale to management? With serial processing, the costs of management rise exponentially up the hierarchy (because exponentially more decisions are delayed by higher-level activities), so it is unlikely that top-level management can profitably participate in serial decisions. Even with parallel processing, if top-level management is too far removed from the operating decisions (and if good aggregates[3] are not available for use in high-level decisions), then the top-level manager's contribution to the firm may be quite small (as demonstrated by our quadratic resource allocation example).

[3] Milgrom and Weber (1983) study the properties of prices as statistical aggregates used by high-level decision makers.

Our theory could be naturally extended in several directions. First, although we have compared the performance of alternative hierarchical forms, we have not compared hierarchies with other organizations, like markets, in which managers may take "orders" from many sources. Second, the theory could be combined with a theory of incentives, in which the desire to have managers acquire private information is tempered by the incentive problems that such information may sometimes create.

References

Beckmann, M. (1977). Management production functions and the theory of the firm, *J. Econ. Theory* 14, 1–18.

Calvo, G., and Wellisz, S. (1978). Supervision, loss of control and the optimal size of the firm, *J. Polit. Econ.* 87, 943–952.

Cremer, J. (1980). A partial theory of the optimal organization of bureaucracy, *Bell J. Econ.* 11, 683–693.

Keren, M., and Levhari, D. (1983). The internal organization of the firm and the shape of average costs, *Bell J. Econ.* 14, 474–486.

Marschak, J., and Radner, R. (1972). "Economic Theory of Teams," Yale Univ Press, New Haven/London.

Milgrom, P., and Weber, R. (1983). "Organizing Production in a Large Economy with Costly Communication," Cowles Foundation Discussion Paper

Rosen, S. (1982). Authority, control and the distribution of earnings, *Bell J. Econ.* 13, 311–323.

Simon, H. (1976). "Administrative Behavior," 3rd ed., MacMillan, New York.

Williamson, O. (1967). Hierarchical control and optimum firm size, *J. Polit. Econ.* 75, 123–138.

Williamson, O. (1985). "The Economic Institutions of Capitalism," Free Press, New York.