OBJECTIVES AMONG MANAGERS AND FOCUSERS INSTEAD ON THE PROBLEM OF COORDINAT-ING THE DECISIONS OF SEVERAL IMPERSONALLY INFORMED AGENTS. 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 POSSIBILTY THAT BOUNDED RATIONALITY LIMITS THE MANAGERS' ABILITIES TO UNDERSTAND MULTIPLE MESSAGES.

MANAGEMENT IS FUNDAMENTALLY ABOUT DECISION MAKING, AND A USEFUL THE-ORY OF MANAGEMENT MUST CONSIDER THE BOUNDARIES OF RATIONALITY. IF INDIVIDUAL MANAGERS HAD UNLIMITED ACCESS TO INFORMATION, THEN THEY COULD PROCESS CONFIDENTLY 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 CAPACITIES OF INFOR-MA TION PROCESSING BUT CAN LEARN HOW TO ALLOCATE THEIR ATTENTION, THE ADVANTAGE OF SHARING RESPONSIBILITY FOR MANAGEMENT WHERE 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 DECISIONS 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 CONSISTING OF MANAGERS WITH LIMITED ATTENTION AND DERIVE SOME GENERAL PROPOSITIONS. FIRST, OPTIMAL ORGANIZATIONS ALWAYS ASSIGN MANAGERS TO ACQUIRE DIFFERENT INFORMATION. SECOND, THERE ARE SUPERINSENSITIVE RETURNS TO ABILITY, WHERE ABILITY IS DEFINED TO BE A VECTOR OF MANAGER'S SPEED IN PROCESSING VARIOUS KINDS OF INFORMATION. THIS MEANS THAT EVEN WHEN ABILITY IS ONE DIMENSIONAL, THERE WILL BE NO GENERAL SUPPOSITION THAT THE "ABSENT" MANAGER WILL BE HEADED AT THE TOP OF A HIERARCHY. HOWEVER, WE DO PROVIDE A SUFFICIENT CONDITION FOR THIS CONCLUSION WHEN THERE IS SERIAL PROCESSING OF INFORMATION IN SUCH A WAY AS TO GENERATE PROFESSIONAL HIERARCHIES. ORGANIZATIONS ARE PROPOSED TO BE MORE BROADLY THROUGHOUT THE ORGANIZATION. BOUNDS ON MANAGERIAL RATIONALITY ARE INTRODUCED IN SECTION III. THESE ALLOW US TO EXPLAIN THE "COMMANDS" AND HIERARCHICALLY-ORGANIZED OBJECTIVES WITHIN ORGANIZATIONS—IMPORTANT PHENOMENA WHICH ARE INCONSISTENT WITH TRADITIONAL TEAM MODELS. FOR EXAMPLE, IN THE MARSHAK AND RADNER (1972) MODEL, THE ROLE FUNCTION OF COMMUNICATION IS TO PASS INFORMATION AMONG MANAGERS. THIS IS NEVER THE CASE FOR A MANAGER TO RESTRICT THE WIDE SET OF ANOTHER MANAGER OR TO SPECIFY THE OBJECTIVE HE SHOULD ATTAIN, FOR A SUBORDINATE WITH UNLIMITED ATTENTION AND CALCULATIVE POWER WHO KNOWS THE (OPTIMAL) STRATEGY OF HIS SUPERIOR COULD AVOID THE SET OF POSSIBLE PARADOCS FOR ANY INFORMATION HE RECEIVES AND WOULD BE LED TO FOLLOW THOSE INSTRUCTIONS VOLUNTARILY. IN OUR MODEL, IN WHICH MANAGERS CANNOT CON- SISTENTLY AND INITIATIVELY EXTRACT INFORMATION FROM OTHER DATA SOURCES, AN ASSUMPTION THAT THE WHOLE CONTEXT OF A MESSAGE IS CONTINUOUSLY PROCESSED WOULD BE OUT OF PLACE. INSTEAD, WE ASSUME THAT ONLY THE SURFACE CONTENT OF A MESSAGE LIKE "PRODUCE 800 WIDGETS" CAN BE GRANTED CONFIDENTLY; THE SUBTLE CONTENT, WHICH DEPENDS ON DRAWING AN INFERENCE FROM THE MESSAGE USING KNOWLEDGE OF THE SENDERS' RULES, 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 ASSIGNING NONPRODUCTION TARGETS, CAPITAL, AND OTHER RESOURCES TO DIVISION MANAGERS, WHO IN TURN REALIZE THE BUDGETED AMOUNTS TO THEIR SUBORDINATES, ETC., UNTIL THE BUDGETS AND TARGETS REACH THE "SHOPS" WHERE PRODUCTION TAKES PLACE. THE MODEL WE USE EMPLOY QUADRATIC COST FUNCTIONAL AND INFORMATION-REVEALING TECHNOLOGY BASED ON SAMPLING FROM NORMAL DISTRIBUTIONS. AMONG OUR FINDINGS ARE (I) THAT MANAGERS AT EACH LEVEL QUANTITIVELY FOCUS ATTENTION ONLY ON THOSE VARI- ABLES THAT DETERMINE THE MULTIPLIED PRODUCTIVITY OF RESOURCES AND THE MARGINAL COSTS OF PRODUCTION IN THE UNITS UNDER THEIR COMMAND, (II) THAT THERE IS A BIAS IN THE DEEP THRESHOLD HIERARCHIES (EVEN THROUGH PARALLEL PRO-CESSING OF INFORMATION IN SOURCES), AND (III) THAT ORGANIZATIONS HAVE MORE PRIOR INFORMATION ABOUT THE STRUCTURES AND MORE HIGH-LEVEL MANAGERIAL INFORMATION SYSTEMS WILL EMPLOY LESS ABLE MANAGERS, OR GIVE THEIR MANAGERS WIDER SPANS OF CONTROL, ETC.

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 TO DEVOTE TO A VARIETY OF INFORMATION SOURCES. IF THE ALLOCATION TIME TO SOURCE IS Ac, THE MANAGER ACQUISITION INFORMATION REPRESENTED BY THE AUDIENCE EAc WHERE AC IS A PARAMETER OF THE MANAGER'S ABILITY TO PROCESS INFORMATION. IF THERE ARE N INFORMATION SOURCES (WHERE N MAY BE AN INTEGERS), THEN THE MANAGER'S ABILITY CAN BE EXPRESSED AS A = A1, A2, , AN. THE MANAGER'S INFORMATION AT TIME t IS THEN THE JUMPS OF HIS INFORMATION FROM EACH INFORMATION SOURCE AT TIME t. HENCE, THE SET OF INFORMATION WITHIN A MANAGER OF ABILITY A AT TIME t IS

\[ L_t = \left\{ L_{t-1} + \alpha_i \sum_{i=1}^{n} \gamma_i \delta_{i} \right\} \]

(1)
Hierarchies

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

Proposition 1

In this section, we imagine an organization with a given effectuation of decisions. Each decision must be undertaken no later than its deadline, also given as a quasirational task. Assume that the organization has limited ability to acquire and use information, which imposes costs on each decision process. Therefore, the organization's decision problem is to decide which managers to employ, how to assign decisions to managers, and how to make decisions. A decision assigns a decision to a manager if the decision's cost is lower than the cost of assigning the decision to the manager. The assumption that the organization can assign decisions to managers implies that the cost of assigning a decision to a manager is lower than the cost of making the decision.

Proof: A manager of ability \( a \) and \( b \) could make decisions at each point of time and thus at all of the two managers' deadlines, the same information as any manager and \( c \) together. Since the managers of ability \( a \) and \( b \) would be hired at a position of the two unless his wage was higher than the sum of the organization's and the wages of the two managers. Since in general the information acquired by the boss manager is one of the old jobs will improve his decision in the other job and their organizational ability need to be considered in the same way. The second part of the proposition is proved in the same way. If there are two steps to Proposition 2, the first identifying a form of increasing returns to ability and the second extending the result of the organizational to hiring more managers of ability, the opposite proposition can be used.

Proposition 2

The proof of Proposition 2 is similar to the proof of Proposition 1. The proof of Proposition 3 is similar to the proof of Proposition 2. However, the proof of Proposition 3 is more complex because it involves the interaction between the decision-making process and the information-processing process.

Proof: A manager of ability \( a \) and \( b \) could make decisions at each point of time and thus at all of the two managers' deadlines, the same information as any manager and \( c \) together. Since the managers of ability \( a \) and \( b \) would be hired at a position of the two unless his wage was higher than the sum of the organization's and the wages of the two managers. Since in general the information acquired by the boss manager is one of the old jobs will improve his decision in the other job and their organizational ability need to be considered in the same way. The second part of the proposition is proved in the same way.
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 the intuition, suppose that decision nodes are arranged in a tree and that information processing must be done iteratively, 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 ψ, from our notation and focus on the number of weeks i managers are assigned to a project. We suppose that the weekly wage is constant, so that we write W(i; a, b) = (w(i; a, b) = w(i)).

Proposition 1. Suppose that utility u is in one dimension and that the cost of the time spent by the manager at any node i is ρ(i; a, b), where ρ(i) is the time spent, a is the manager's ability, and ψ is a strictly concave function. If all information processing must be done serially and decision i precedes decision j, then at the optimum, a, = a,

Proof. If i precedes j and a, < a, then the firm can do better by replacing each manager by a manager of ability a, = a, + ψ(j), keeping the manager’s speed same, i.e., in the top job and j, in the other job. Here, u(i) = u(i; a, + ψ(j), a, - ψ(j), and l, = u(j; a, - ψ(j)). Note that in fact l, < l, so that no decision is reached later, and at least one decision is reached faster. Moreover, the wage bill is smaller, since by convexity, l, ≥ l, + ψ(j) = l, + ψ(i; a, + ψ(j), a, - ψ(j)).

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

Consider a variation from the optimal strategy in which manager M takes the time he spends by a while each of M’s direct subordinates spends a more time. Such a change is always feasible, as the marginal product of manager M’s time (i.e, wage paid for that hour) cannot be less than the sum of the corresponding marginal products of time (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 not (net) of wages at bonus doubles with each step up the hierarchy. This may help explain the traditional role of cost-saving devices and producers such as sophisticated decision support systems, executive secretaries, mobile telephones, company cars, and chauffeured limousines for managers and executives at the top of a hierarchy.

References

III. COMMANDS IN TEAMS

Under traditional models of "rational" decision making, a key aspect 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 access to and the signals he receives when he makes his decision.

This if manager i must choose d, i (or, he saves

\[ \max \{ (e - d; \ i, d; i), 0 \} \]

where

i
the organization's payoff
w
is the wage bill
w
is the state of the world
w
the decision made by other managers
i
the information the manager has acquired directly
w
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 i.e., 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 equation (2). The inclusion of ψ, as part of i’s information about ψ, in (2) amounts to an assumption that i can conditionally and instantaneously infer the significance of the signals communicated to him by other managers. In a model with in costs, where information processing is costly and time-consuming, and where i may have unalternative of learning the same information by looking directly at some report in the database, this assumption is suspect.

What is most remarkable about equation (2) is that, in an optimal team strategy, there is no role for "isolation" from any manager to any other. That is, at an optimum, a superior may communicate information to his subordinate, but he avoids the sort of actions that the subordinate might undertake, not does he directly set the objective the subordinate pursue.

To put the same point slightly differently, when communication consists of orders and the requirement to follow orders do not degrade an optimal performance of the team, then the manager can infer from the orders themselves that it is optimal to obey. Optimal orders convey their own justification. However, when managers are not perfectly experts at interpreting communications, there can be a separate role for instructions that limit the manager’s choice set.
Consider an example in which we have three productive units or shops, which are each capable of producing a single output. The cost of producing a unit in shops 2 or 3 is 2x, while in shop 1 it is 1 x = y2, where x is a random variable. Assume that manager M allocates output responsibility between shops 1 and 2, while the chief manager allocates output responsibility between manager M and shop 1. Suppose that the total output which must be produced is 2x, 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 = (x1 + 2x2) and 1 = x1 - 2x2.

Now suppose that both managers know x, but only manager T knows y. Plaut, with “full rationality,” T can set x1 = (1 - y) x and assign the output target x2 = (1 - y) x to his subordinate M, and M can then infer the value of y = 2x - x2. In this way, the cost-minimizing allocation can be achieved. We emphasize that even if manager M were not constrained to set x1 = (1 - y) x, he would choose to do so given the team objective and the inference he can make from the target x2 communicated by T. But what happens if M cannot make such useful inferences correctly, or if the team costs of deducing y 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 hierarchies 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 maximize the objective (x1 + 2x2) - x2 subject to meeting the output target (x1 + 2x2) - x2, he will always choose the first-best allocation despite his inability to infer y from the output target x2.

It is unclear how effective this device or restructing 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 cost. Finally, artificial objectives that restore first-best behavior in general are quite complex. In view of these difficulties, we have found that the extreme assumption that managers seek to minimize the expected costs of their units, given the actual cost function and the manager's knowledge of relevant parameters, is a reasonable approximation to the behavior of both managers.

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 the 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 j ∈ T in an environment without externalities. The set of resources available and the organization's output requirements are given by the vector x_j ∈ R^T. 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 production target x_j for resources and production targets to achieve total cost.

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

C_j(x_j, y_j) = \begin{pmatrix} x_j \end{pmatrix} M_j \begin{pmatrix} x_j \ y_j \end{pmatrix} - \begin{pmatrix} x_j \end{pmatrix} R_j \begin{pmatrix} x_j \end{pmatrix}

(3)

In this model, the vector parameters y_j of the shops are initially unknown. The other technological parameter B_j of shop j is a positive definite matrix which is known 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 intercept. The constant term in (3) is often chosen so that C_j(0, y_j) ≡ 0, thus we interpret C_j 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 (PM) in the tree.

Each manager M except the chief receives a quantity signal from his boss x_m ∈ R^T specifying how much to produce and with what centrally allocated resources and in turn passes instructions and resources (x_j) to (B_j) to his direct subordinates to satisfy the constraint x_j = x_m.

The organization's objective is to minimize expected total production costs, subject to meeting the specified output requirement and using the preallocated resources.

Now we introduce the "database" time, to make it clear. We assume that by devoting time s_j to studying affairs of shop j, a manager of ability α = (α_0, α_1, ... ) observes a statistic distributed like...
\[ y = e^{\text{variance } \sigma^2} \] where \( e \) has a \( k \)-variate normal distribution with mean zero and variance \( \sigma^2 \). In effect, we have assumed that the size of the sample observed is proportional to the attention devoted to any shop.
Each manager has time \( t \) 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, the component of a corresponding to the instructional 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: that one useful rule 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 to increasing costs to high-level management.

We analyze the organization problem in three steps. First, holding the hierarchy and the manager's information fixed, we optimize 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 hierarchy 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 these 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 (ii), 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:

\[ y^v = \sum_{x \in X} y_x \] (4)

\[ r^v = \left[ \sum_{x \in X} r_x \right]^2 - \left( \sum_{x \in X} r_x \right) \] (5)

and

\[ y^o = E(x) \] (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 \( X \). Then, he will allocate resources and assign output responsibility to minimize the quantity

\[ E \left[ \sum_{x \in X} c(x, y^2) \right] \] (7)

subject to the constraint

\[ \sum_{x \in X} x = x_o \] (8)

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

\[ y^o = E(x) \] (9)

This means that the random variable \( y^o = E(x) \cdot x_o \).
Proposition 4. The solution to manager M's problem is $x_i = y_j = \bar{y}_j - \bar{y}_j' + \bar{y}_j''$, where $\bar{y}_j = B_j I_{LM}$, where $B_j = B_j I_{LM} - \bar{y}_j'$. The conditional variable cost function $F(M)$ is given by

$$F(M) = \frac{1}{2} \sum \left( d_i - y_j R_i - \bar{y}_j - y_j' - y_j'' + d_i \right)$$

Proof: Substituting the form of the cost function (6) into the objective (7) leads to a quadratic maximization with a linear constraint (6), whose solution is realized. Let $y_j + \bar{y}_j + \bar{y}_j'$ denote the prior expectation of $y_j$. According to Proposition 5, the maximum cost function could be achieved if there were no information available for the allocation decision (apart from the information reflected in the prior belief) is given by

$$E[\hat{y}_j] = \frac{1}{2} \sum \left( d_i - y_j R_i - \bar{y}_j - y_j' - y_j'' + d_i \right)$$

The expected savings attributable to the management at $N$ is defined to be the excess of the zero-information maximum expected cost given by (10) over the expected cost $E(C_{m_{ij}})_{ij}$ incurred with information $I_{LM}$. Using the pair of identities

$$E[y_j + \bar{y}_j + \bar{y}_j'] = y_j + \bar{y}_j + \bar{y}_j'$$

and

$$E[I_{LM} y_j] = I_{LM} E[y_j]$$

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

Corollary 1. The expected savings attributable to management at $N$ is

$$-E[I_{LM} Var[y_j]] + \sum_{\bar{y}_j''} E[I_{LM} Var[y_j]]$$

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. A statistic is sufficient for a random variable $\xi$ if for all subsets $A$ of the range of $\xi$

$$P(\xi \in A | I_{LM}) = P(\xi \in A | I_{LM})$$

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 bas his opinions only on summary statistics or aggregate information about low-level units, while the lower-level manager pays attention to their details. It must also be justified when higher-level managers base their opinions on executive summaries of reports about low-level units, on data tersely gathered by the lower-level managers, and on reports that the lower-level managers prepare.

Proposition 5. Suppose we are given a hierarchy $N$ and information $I_{LM}$ for each manager $M$, with the property that $I_{LM}$ is sufficient in $(y_j, \bar{y}_j, \bar{y}_j')$ for $y_j, \bar{y}_j, \bar{y}_j'$ SEMS. Then, the minimal team strategy is to choose $y_j, \bar{y}_j, \bar{y}_j'$ according to $y_j = \bar{y}_j + \bar{y}_j' - \bar{y}_j''$, where $\bar{y}_j = \bar{y}_j I_{LM} - \bar{y}_j'$. The expected total cost incurred by the hierarchy when the output target is $y_j$ in the zero-information-maximum expected cost for the entire hierarchy, given by

$$E[y_j + \bar{y}_j + \bar{y}_j'] - \sum_{\bar{y}_j''} E[I_{LM} y_j]$$

and the sum of the expected savings attributable to management at each level of the hierarchy:

$$\sum_{\bar{y}_j''} E[I_{LM} Var[y_j]] + \sum_{\bar{y}_j''} E[I_{LM} Var[y_j]]$$

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 $I_{LM}$ is sufficient in $(y_j, \bar{y}_j, \bar{y}_j')$ for manager $M$'s problem implies that any inability to draw inferences from $I_{LM}$ about $y_j$ 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 \( c_{m} = \text{STAMP} \). 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 size of the hierarchy, the share of all levels of management were perfect, the first-best outcome would be achieved. In terms of Proposition 5, \( \gamma_{m}^2 = \gamma \), for all \( m \) and \( i \), and the savings achievable 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 level. For example, by assigning high-level units to a small group, one ensures that there is little potential to increase the highest levels and the key to good performance becomes efficient management at the lowest levels of the hierarchy. When information processing at the highest levels is especially costly, as in the small-information-processing cost studied earlier, or when the information available for making high-level decisions is too 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 (the market wages of managers in different ability classes).

Note that the savings attainable to managers does not depend on the organization’s output target \( y_{n} \). Also, the optimized correspondence to \( l_{n} \) is a term that depends only on the information \( \gamma \). Therefore, for a fixed hierarchy, the problem of choosing information systems optimally for all the managers is solved by maximizing the individual value, provided that the solution to these problems satisfies the “sufficiency condition” of Proposition 5. In that case, the original problem decomposes into the choice of a system with the minimum information cost for each individual manager separately. To develop this idea, we make the following additional assumption.

**Assumption.** The \( y_{n} \) are one-dimensional (\( d = 1 \)) and independently and normally distributed with prior variance \( \gamma_{n}^{2} = \gamma_{0}^{2} \).

It is convenient analytically to work with the variance of the variance \( \gamma_{2} \), rather than with the variance itself. A manager may deviate attention to observing any of the \( y_{n} \). The information technology described earlier can now be recast as follows: The precision of the manager’s observation of any \( y_{n} \) is proportional to the time spent observing \( y_{n} \), where the constant of proportionality \( \alpha \), may reflect both the manager’s ability and the quality of the information system in the manager’s disposal. A service manager’s time allocation problem is then to choose times \( t_{n}^{2} \) to devote to observing each \( y_{n} \), in order to maximize the expected savings

\[
\sum (R_{n} - E_{n}^{2}) \text{subject to the conditions} \sum t_{n}^{2} = t \text{and} \sum t_{n}^{2} = t
\]

Proposition 6. The solution to the manager’s allocation of the attention problem is characterized by a number \( k \) such that each \( t_{n}^{2} \) is the maximum of zero or the solution to

\[
t_{n}^{2} = \frac{\alpha}{2} (R_{n} - E_{n}^{2})
\]

**Proof.** This is a linearly constrained concave maximization problem and \( t \) is the Lagrange multiplier of the total time constraint.

In studying formula (17) one sees that, as deciding how informed to become about the situation at a shop, the manager will weigh the positivity of the shop’s costs to the allocation (measured by \( R_{n} - E_{n}^{2} \)) against the difficulty of gathering information about the situation (measured by \( \gamma_{0}^{2} \)).

The manager does not try to keep his knowledge about the shops where his prior information is perfect, 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 if he or she has invested in increasing \( \gamma_{0} \), that is, in making it cheaper to acquire information, but will gather less additional information if the prior information is better.

**Corollary 2.** In the fully symmetric case with \( R_{n} = R \), \( s_{n} = s \), for all \( n \), the optimal solution is \( t_{n}^{2} = 0 \), where \( s \) is the number of shops in the service, \( t = 0 \), and \( \alpha \). The savings attributable to the manager \( M \) is \( s \gamma_{0}^{2} \). 

These savings, regarded as a function of \( s, \alpha \), display increasing returns to scale. That is, 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 \( t_{n}^{2} \), the expected observed 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 modeled to extend the authority of the manager, allowing him or her to reallocate resources across more shops.

How 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 1, the problem is

\[
\text{Maximize } \frac{\alpha}{\beta} \left( \frac{w}{\beta} + \frac{v'}{\beta} \right) \quad \text{subject to } N_{s} \beta \leq \frac{w}{\beta} + \frac{v'}{\beta},
\]

where \( w \) is the managerial wage. We impose the natural assumptions that \( w, \beta, \alpha \), and \( r \) are nonnegative.

**Proposition 6.** If \( w = \beta \alpha \), the optimal service size is \( v' = 0 \) (two managers are hired). Otherwise, the optimal service size is an integer within one of

\[
(19) \quad x_{0} = \left\lfloor \frac{w}{\beta - \alpha} \right\rfloor + 1
\]

where \( x_{0} = \text{optimal number of managers}. \) Moreover, \( N_{s} \) is increasing in \( w \) and first decreasing then increasing ("U-shaped") in \( \beta \).

**Proof.** One can verify that, when \( a \) is treated as a continuous variable, the problem is strictly quasi-convex in \( a \). Hence, the optimal integer solution 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 \( \beta \), proceed in outline as follows. Reciprocate (19) as a function of \( \beta \) by \( \beta = \frac{w}{\alpha} + \frac{x}{\alpha} = 1 + \frac{w}{\beta - x} \).

Suppose that firms operating in older, more stable industries have better prior information about their environment and more highly refined information systems for monitoring the environment than firms in newer, 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 employ fewer high-level managers per unit of output.

Our final exercise on the quadratic case is particularly interesting. We examine the value of high-level management using Proposition 5. The following calculation is illustrative. Suppose the organization consists of a shops in each, with slope coefficients $\beta$ and prior precision $\kappa$ about its intercept parameter. Further suppose that manager $M$ has $N_M$ direct subordinates, each of whom heads a subunit containing $n$ managers. 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, $n$ can be shown that each manager's time allocation problem is a symmetric, concave problem whose optimal solution specifies devoting equal time to observing the parameters of each shop $s > M$. Moreover, one can verify that $n_M = \frac{n}{M}$ for every manager $M$. Then, by using the formula for positive measures associated with the normal sampling technology and applying Proposition 5, one obtains the following result:

**Corollary 3.** The salaries attributable to managers at $M$ when the manager's ability is $n$ and the time available is $t$.

$$\frac{\partial S_M}{\partial n_M} = \frac{1}{n_M} \left( \frac{n_M}{n} - \frac{1}{n} \right)$$

Expression (20) tends to zero as the size of the units managed ($n_M/n$) tends to infinity. Hence, for any positive wage $w$, there is a limit to the size of the units that can profitably be incorporated into a larger organization. The particular form of (20) depends on the specification of the $n$ 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 $\omega$ convergence even though there is significant gain 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.

**Hierarchies.**

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 (1975) indicated the form of a model in which the interactions given by the chief executive are divided by a fixed amount as they pass through each successive 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 power decisions at a middle manager in a large organization than at a chief of a smaller one.

Calvo and Wallich (1978) studied a relatively "small" model based on the idea that managers who are not monitored carefully will cheat their subordinates. Unlike Williamson, they found that the optimal size of an organization. Similarly, Beckman (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, Beckman concluded that the average cost of managing declines with the size of the hierarchy.

Kerr and Levhari (1981) reached the opposite conclusion in a model where the productivity of the firm depends on the time it takes to reach a decision, which is turn depends on the size 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 costly to have many tiers in the hierarchy, since (1) there is a fixed time cost to each tier, and (2) 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. Only as the span of control becomes 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 proportionate effect on the productivity of workers but is subject to diminishing returns with the size of the organization. He concluded that it is possible to place the chief manager at the top of the organization.

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 factor-cost or production function that is meant to reflect some unmodeled aspect of management's limited discretion making capacity. In contrast, our theory takes the discretion problem itself and its limits on managerial attention as principle and other competing constraints to these common bowed-out curves 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 render decisions so as to bring more information to bear than any single manager could bring alone. Interactions component for the predominant form of coordination among individual action makers by gaining higher-level managers the task of coordinating the lower-level decisions.

Who are organizational actors to have half-of-managers as only proportionately reduced actors? The answer is because a half-of-manager is more twice as productive as two half-of-managers, the full-time manager brings about whatever is required for each decision from 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 support more than by their subordinates' productivity advantage over their less diligent co-workers.

When does the subordinates belong to the top of the organization? Equality, when the value of ability in the top job likely be great. The answer is inevitably. But the value is greater when the top-level decision involves social relationships between the chief and his subordinates, because then a "low" chief causes costly delays to echo through the whole-organization.

How does the form of the hierarchy and the kind of managers is really depend on its environment? In the economic, resource allocation emphasis, it takes in managers and outcomes 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 order, more specialized activities employ fewer silo managers or give them smaller spans of control than firms in order, more stable industries.

Are these unilateral economies of scale to management? With social processing, the costs of management may grow substantially the hierarchy because especially many decision are delayed by higher-level activities. In a subordination, top-level management can profitably participate in social decisions. Even with parallel processing, top-level management is far removed from the operating decision and if pot-end activities are.

References


