WORKING PAPER NO. 4

Information Structures, Task Structures, and Coordination Systems

by

Hideshi Itoh

Faculty of Economics
Kyoto University

Faculty of Economics,
Kyoto University,
Kyoto, 606 JAPAN

Faculty of Economics
Kyoto University
Kyoto, Japan 606

Information Structures, Task Structures, and Coordination Systems

by

Hideshi Itoh

Faculty of Economics
Kyoto University

April 1988

ABSTRACT

This paper concerns the design of the firm organization to obtain and use information efficiently in organization decision making. The focus is on coordination of shop managers' operating decisions through the choice of the organization structure such as the coordination system (hierarchical or horizontal) and information processing capacities of subordinates (specialists or generalists). Assuming that information acquiring, processing, and communication are costly, we show that in "volatile" environments, the optimal organization structure is the one typically found in Japanese firms, where coordination tasks are delegated to subordinates who are nonspecialized in tasks and information acquiring so that they can share each other's on-the-spot knowledge.

Information Structures, Task Structures, and Coordination Systems*

Hideshi Itoh†

Current Version, March 1988

Comments are welcome.

1. Introduction

This paper concerns the problem of coordination in organizational decision making that arises because of members' specialization in different information sources and their limited rationality. Since the firm is surrounded by uncertain, changing, complex environments, its managers must collect a huge amount of data from various information sources and process those data in order to acquire information or knowledge valuable in their decision making. Those managers are limitedly rational in information acquisition and processing: their attention to data and their information processing capacities are usually scarce resources.[1] One advantage of the organization is that it can have its members specialize in different information sources, which enables the firm to obtain more information than when it is managed by a single manager. However, even if all the members share the same organization goal, their independent decision making may be far from desirable from the organizational point of view because of differential information among members. We need to coordinate decisions by specialized members.

In hierarchical organizations, this coordination problem can usually be solved by authority. That is, when each member but one distinguished "top" manager has one and

* This is a modified version of a chapter in my dissertation. I would like to thank Masahiko Aoki, Hugo Hopenhayn, and David Kreps for their comments and encouragement. Remaining errors are, of course, my own.

† Graduate School of Business, Stanford University, Stanford, CA 94305; and Faculty of Economics, Kyoto University, Kyoto 606, Japan.

1 The development in mechanical information-processing systems in the today's world, such as computers, does not resolve this problem: the scarce resource is not data but human capabilities. See Simon (1974).

only one direct boss, each boss is able to coordinate decisions of his or her subordinates by receiving from them reports concerning the information they possess, and then by telling them what to do. One problem of this authoritative system is that attention or abilities of the upper-level managers are also limited, so that it is difficult to suppose that they can share all the knowledge of their specialized subordinates through reports. Hence the hierarchical coordination system may lose the opportunity to utilize on-the-spot information only available to lower-level members for the organizational decision making.

Therefore, even in the hierarchical organization, it may be of value to delegate coordination itself to lower-level members of the organization. In this "horizontal" coordination system, too much specialization will be harmful for the reasons mentioned above. Some knowledge sharing among subordinate members through horizontal communication or adoption of generalist-type members may be preferable.

By parametric analysis of a simple model, this paper examines the performance of these two kinds of coordination systems. More specifically, the model is described as follows: The firm is a two-tier hierarchy, consisting of a top manager and many subordinate shop managers. Each member of the organization observes noisy signals concerning some components of a multidimensional random variable that represents the uncertain environment. Each component may represent profitability of an industry where the firm operates, or it may represent technological conditions specific to each shop. The payoff to the firm, which is assumed to coincide with each member's objective, depends upon the subordinates' actions and the realization of the random variable. Specifically, we assume a quadratic form composed of three terms as the firm's objective function. The first term increases if the actions are closer to the state variable, the second one increases if, in each given shop, the actions are closer to each other, and the last term is higher as the actions in one shop are closer to the actions in others. For example, imagine that two workers try to discern the source of breakdown of two machines as precisely as possible. The first term is higher as their joint efforts detect the problem of each machine more closely. The increase in the second term comes from coordination of their actions for each machine. The last term increases if their joint actions are coordinated across machines.

The novel feature of the model is that the top manager can choose *information processing capacities* of subordinate members in the organization. The information processing capacity is defined as an ability of each member to perceive the environmental variable. In the model presented in Section 2, each subordinate observes each component of the state variable plus a noise term. All random variables are assumed to have a Normal distribution, hence the precision of each noise term represents the subordinate's ability to perceive that component. The firm will attempt to acquire employees with desirable expertise through searching labor markets or designing promotion and job rotation systems. We assume that better abilities are more costly because of training or search costs. Then one of our concerns is whether the top manager should choose as subordinate members many *generalists* who have good abilities for all components of the state variable, or choose many *specialists* who are excellent for some particular components, but poor for others.

In Itoh (1987), a simple one-top and one-subordinate organization was considered, and the optimal information processing capacity of the subordinate was examined, where the subordinates' information processing capacity may depend on a variable representing some global aspects of the environment observed by the top manager. Therefore, "strategic decisions" made by the top manager usually affect "operating decisions" subsequently made by subordinates because their capacities depend on what the top manager observes and decides. In this paper, however, this interdependence is ignored. Instead, I consider a one-top and two-subordinate organization and ask how to coordinate actions of two subordinates who may pursue idiosyncratic subobjectives *because of differential information.* That is, in this paper, strategic decisions by the top manager are assumed to be fixed, and the focus is on coordination of operating decisions implemented by several subordinates through the choice of the coordination system and their information processing capacities.

In this setting, two coordination systems mentioned above are defined as follows. In the *hierarchical coordination system,* the top manager collects information from shop managers and makes all organizational decisions. Then the top manager orders subordinates to implement her decisions.[2] Here we assume that each shop manager cannot infer the observation of the other subordinate from the command by the top manager.[3] This assumption

[2] As in the last paper, the top manager is assumed to be a female and each shop manager a male for the purposes of identification.

[3] Of course, we assume that subordinate members are Bayesian players in their information processing with their direct information sources. The point is that the inference from the order is too subtle for them to understand its informational content.

introduces another important aspect of limited rationality, different from limited attention, into the model. If subordinates had such inference abilities, they would recalculate the optimal decision based on their inference, so that there would be no role of orders.[4] Since the top manager decides everything under this system, it has some advantage in solving the coordination problem due to differential information.[5] However, there are some disadvantages, too. The top manager also requires a sufficient amount of time to read and understand the reports, and the more time she spends reading the reports, the more accurately she understands them while the more costs are created from delay of her decision making. Then generally the top manager cannot share in all the information subordinates have.

In the second coordination system, called the *horizontal coordination system,* the shop managers make their own decisions, without any suggestion by the top manager, possibly after sharing partially what they know by interchange of information. Note that under this system, the *task structure* matters: The task structure specifies who determines which actions. We consider two task structures: the specialized task structure makes each subordinate specialize in the actions relevant to his job shop while under the nonspecialized task structure, each subordinate is involved in the decision making in both job shops. Since each shop manager makes decisions depending on his knowledge, the resolution of the coordination problem will be imperfect. Also there are the same problems of limited attention to reports as in the case of the hierarchical system. Under this system, however, each subordinate can use for his decision the on-the-spot information directly collected by himself from information sources, which is not available to the top manager under the hierarchical system.

The questions asked in the paper are: Under each coordination system, does the top

[4] For more detailed arguments concerning this, see Geanakoplos and Milgrom (1985), who first pointed out the relation of such bounded rationality to the role of command in the organization.

[5] It is more general and realistic to assume that the top manager limits the sets of actions that the subordinates may undertake rather than chooses particular actions. Then the problem becomes the choice of the optimal mix of two systems. In this chapter, I consider two extreme organization structures to focus on the relation between the coordination system and information processing capacities. Another justification of the assumption in the main text is that the two-stage action choice described in this footnote may cause prohibitively expensive time delay for implementing actions.

manager choose generalists or specialists as shop managers? How is that choice affected by the environment? How does the optimal choice of the coordination system depend on the environment?

The basis of the model and the analysis is the theory of teams. (Arrow, 1985; Marschak, 1974; Marschak and Radner, 1972.) Team theory sets the stage for the analysis of the *information structure* (which specifies the information available to each member) as well as the *decision structure.* Recent research on the economics of organizations has put greater emphasis on the information to control subordinates in the firm hierarchy who pursue their objectives differing from the firm's goal. (See Holmström and Tirole, 1987, for a survey.) Since my concern in this paper is in the choice of the information structure to improve organizational decision making, I ignore incentive aspects of the organization, though the importance of conflicts within the organization are not meant to be underemphasized.[6] Therefore, the model examines the performance of several organization structures from the perspective of limited rationality (imperfect communication and limited attention), provided that incentive problems are resolved so that each member of the organization shares the same organizational goal. How such alignment is achieved is another important problem and is beyond the scope of this paper. We return to this problem in Section 7.

The analysis of the two coordination systems in this paper is related to the recent studies comparing American (or more generally, Western) management with Japanese management. (See Lincoln and McBride, 1987, for a survey.) According to those studies, especially Aoki (1988), the stylized nature of the coordination in the American firm is characterized by the hierarchical and authoritative coordination of specialized tasks, while the Japanese firm typically adopts a different kind of coordination system, with the following two main features: horizontal information flows and ambiguous job separation. Aoki compares the operational coordination system in the manufacturing department of the stylized Western firm and the typical Japanese firm, given the "strategic" decisions set by the top management under both firms. He asserts that in the Japanese firm, as is represented by the well-known "kanban" system, once initial preliminary production plans are set by the central planning unit, then operational coordination, without any further interruption by the

[6] Other models examining the information structure of the organization, based on team theory, include Aoki (1986), Geanakoplos and Milgrom (1985), and Green and Laffont (1986). These are discussed in Section 8.

center, is conducted at the shop floor levels through horizontal information flow. And he points out that the other feature, ambiguous and fluid job demarcation, is observed in the regular rotation of workers in Japanese work organizations.

Aoki argues that when scale economies are less important, and flexible and quick adaptation to "volatile" environments becomes imperative, the hierarchical organization of work, carefully supported by Williamson (1985) on efficiency grounds, may not economize transaction costs as well as the Japanese system does. His argument is extensive, including dynamic learning effects and incentive aspects, so that our model is too simple to cope with all of his assertions.[7] However, we show in our model, which is based solely on limited rationality such as communication costs and limited attention, that hierarchical coordination tends to depend more on specialization than does horizontal coordination, and that in highly uncertain, "volatile" environments wherein scale efficiencies are small and quick, flexible responses to changing environments are important, horizontal coordination *with nonspecialized capacities and nonspecialized tasks* works best. Therefore, insofar as our simple model captures some of the essential characteristics of both typical Western firms and Japanese firms correctly, the results are compatible with Aoki's arguments.

Though Aoki (1988) concentrates on production systems and does not offer any empirical evidence, Kagono *et al.* (1985) seem to support his arguments in a context more general than operating decisions. Their survey research comparing strategy and organization in American management and Japanese management shows that coordination in Japanese firms depends more on sharing of values and information among employees in different positions (*e.g.* between manufacturing departments and marketing departments, among different brand managers, and so on). They observe that in Japanese firms, each functional manager sometimes invades the functions of other managers and departments.

In addition, there is some literature comparing the skill formation within the firm through OJT (on-the-job training) between the U.S. and Japan, especially Koike (1977). (This is the book written in Japanese. Koike (1978) is the English translation of the last paper of the book. Also see Koike (1984) where he summarizes his main findings.) He investigated the skill formation process in shop floors in both the U.S. and Japanese firms,

[7] His reasoning about the disadvantage of the hierarchy is based on: (i) high inventory costs, (ii) rigidity of specialization, (iii) communication costs in the hierarchy, and (iv) the lack of incentive to respond to local shocks.

through extensive field research, and he finds that OJT for blue-collar workers in Japanese firms includes a wider variety of jobs than for blue-collar workers in American firms. He argues that this property of typical Japanese firms, together with the regular rotation of workers, makes the workers both skilled in a relatively wide range of jobs and familiar with the whole work process.

This paper is organized as follows. In Section 2, the model is introduced. In Section 3, some preliminary analysis is conducted. In particular, the simple case where there is no need for coordination across shops is examined. It is shown that when coordination across shops is unnecessary the horizontal system with the specialized task structure is always optimal. This result supports the M-form (multi-divisional form) when the firm operates in two unrelated industries. Then we move to the case in which coordination across shops is extremely important. In this case, the horizontal system with the specialized task structure is always dominated by the hierarchical system, so that thereafter we concentrate on the hierarchical system and the horizontal system with nonspecialized task structure where each manager has to be involved in the management of *both* shops. Section 4 examines the hierarchical coordination system. We show that under the hierarchical system, the top manager always makes subordinates specialize in different information sources. Section 5 is devoted to the analysis of the horizontal coordination system with nonspecialized task structures. Under the horizontal system, the optimal information processing capacities depend on several parameters. Based on the results in Sections 4 and 5, the optimal coordination system and information structure of the firm under the assumption of extreme needs for inter-shop coordination are derived in Section 6. We will see how the optimal structure depends upon some exogenous parameters characterizing the outside environment. In Section 7, some extensions and the implication of the model are discussed. Finally, Section 8 provides concluding remarks.

2. The Model

Consider a two-tier hierarchical organization of the firm whose members are one top manager (hereafter called top) and two subordinate shop managers 1 and 2. There is an uncertain environment represented by independent random variables X and Y, to which the firm has to adapt. For simplicity, assume X and Y are independently and Normally distributed with $\mathrm{E}(X) = \mathrm{E}(Y) = 0$, $\mathrm{Var}(X) = \sigma_X^2$, and $\mathrm{Var}(Y) = \sigma_Y^2$. We also use h_X and h_Y as precisions of X and Y, respectively.[8]

There are four actions a_1, a_2, b_1, and b_2 that the subordinates must implement. The objective of the organization is to implement the actions to minimize the value of the loss function $l(a_1, a_2, b_1, b_2; x, y)$ which is given by

$$\begin{aligned} l(a_1, a_2, b_1, b_2; x, y) = {} & (a_1 + a_2 - x)^2 + (b_1 + b_2 - y)^2 \\ & + \mu_A(a_1 - a_2)^2 + \mu_B(b_1 - b_2)^2 \\ & + \lambda[(a_1 + a_2) - (b_1 + b_2)]^2 \end{aligned}$$

where μ_A, μ_B, and λ are nonnegative constants.[9]

To give some simple interpretation to this loss function, imagine that X represents the condition of the machines in shop A and Y represents the machine condition in shop B. (Or similarly, we can imagine that two random variables represent economic conditions of two industries A and B.) The first two terms of the function show that the firm tries to detect the true condition of the machines in each shop as closely as possible. Each of these terms represents a direct effect of the shop-specific actions on each shop. On the other hand, the term with μ_A or μ_B does not depend upon the realization of the state variable. Instead, it shows that two actions a_1 and a_2 (or b_1 and b_2) implemented in shop A (shop B) are more desirable if they are closer to one another because of some economies of scale or synergy effect. This term might be called an intra-shop coordination term. Finally, the last term shows that the closer are the joint actions $a_1 + a_2$ in shop A and $b_1 + b_2$ in shop B the better the performance of the firm because of interdependence between shops. This term represents inter-shop coordination.

Each shop manager implements two actions. The *task structure* of the organization assigns two actions to each manager. Essentially, there are two structures we should consider. The *specialized task structure* assigns actions a_1 and a_2 to subordinate 1, and b_1 and b_2 to subordinate 2. Under this structure, each subordinate specializes in one shop, and there is no way he can influence the direct effect of the other shop. The other structure,

[8] Unless otherwise noticed, the same notations are used throughout the paper: For a random variable V, the variance and the precision are written as $\sigma_V^2 = 1/h_V$. Also the realization of V is written by the small letter v.

[9] This loss function is the same as one used by Green and Laffont (1986) if $\lambda = 0$.

called the *nonspecialized task structure*, assigns actions a_i and b_i to subordinate i, so that the direct effect of each shop is determined by the joint actions of two subordinates and each shop manager can provide some help to the other shop manager. These two task structures may have clearer interpretation when applied to two industries instead of two shops. Suppose that X and Y represent some uncertain economic conditions of industries A and B, respectively, and that a_1 and b_1 are, say, marketing-related actions and a_2 and b_2 production-related actions. Then the term with μ_A or μ_B represents an inter-functional effect; for each industry, the firm needs to coordinate the marketing decision and the production decision. In this example, if the nonspecialized task structure is adopted, the firm has a U-form (unitary form) structure. Each subordinate is a representative of a functional division. On the other hand, if the specialized task structure is adopted, the firm has a M-form (multidivisional form) structure. In this case, each subordinate specializes in one of two industries.[10]

Now I describe the information structure of the organization. There are two stages in information processing in the firm: the *information acquiring stage* and the *communication stage*. In the information acquiring stage, each subordinate spends time acquiring information concerning X and Y, and at the end of the information stage, subordinate i receives signals $X_i = X + \epsilon_i$ and $Y_i = Y + \eta_i$, where ϵ_i and η_i are Normally distributed with mean zero and are independent of the other variables. The precisions h_{ϵ_i} and h_{η_i} depend on his *information processing capacities* p_i and q_i and time allocations as follows:

$$h_{\epsilon_i} = p_i t_i \qquad \text{and} \qquad h_{\eta_i} = q_i(1 - t_i)$$

where t_i is the time used by shop manager i for gathering and processing information about X. It is assumed that he can use one unit time for information acquiring, so that the time $1 - t_i$ is used for information about Y. Abler subordinates are assumed to be more costly, with the cost, called the *information capacity cost*, taking the linear form $K_C p_i + K_C q_i$ where $K_C > 0$ is a constant marginal cost of information processing capacities. This cost comes from either training subordinates to have desirable abilities or from searching for and paying subordinates with better information processing capacities.

[10] I am grateful to Mâcé Mesters for reminding me of the task assignment problem and suggesting this interpretation.

Then in the communication stage, communication occurs. Communication patterns depend on the coordination system. There are two feasible coordination systems, the hierarchical coordination system and the horizontal coordination system, one of which the top manager chooses. The two coordination systems differ in where *communication channels* are installed and in who makes the action choices. One communication channel makes one-way communication possible. For example, if subordinate 1 wants to send some reports to top, one communication channel is necessary between subordinate 1 and top. Two-way communication needs two channels. (But see below, especially footnote 11.) The cost of installing the communication channel can be ignored since we compare two systems both of which require two communication channels. In the model analyzed, there is no restriction in dimensionality of reports: Subordinate i can send the values of both X_i and Y_i. However, it may be required to write some *summary* reports before sending to top. This idea can be formalized by assuming limited dimensionality in communication channels. Some implications of limited dimensionality will be discussed in Section 7.

The hierarchical coordination system. Figure 1 shows the information flow and the action choice under the hierarchical system. In this coordination system, two communication channels are installed between top and two subordinates. Each channel is used by each subordinate to send his message to top.[11] Each subordinate i sends his information X_i and Y_i to top. We assume that top has exogenously given abilities to understand the information concerning both X and Y, but that more attention to reading reports from shop managers leads to more delay in her decision making and so creates more costs. We formalize this simply as follows. Top observes random variables of the form $X_i + \theta_i$ and $Y_i + \nu_i$ where θ_i and ν_i are independently and Normally distributed with mean zero and the precisions $r_i = \hat{p}t_{X_i}$ and $s_i = \hat{p}t_{Y_i}$, respectively. We assume that top has the information processing capacity $\hat{p}$, which is exogenously given and is the same regardless of information sources.

[11] It is assumed that the top's suggestion about the actions can be sent to subordinates without using communication channels. Following Arrow (1974), I assume that it is the difficulty in communicating complex information that makes the limited rationality of the decision makers matter. It is assumed that there is no difficulty in just understanding and following the orders. (However, as is discussed in Section 1, shop managers cannot understand the information content of the orders.) As Arrow states, authority economizes transmission and handling of information by using the orders instead of retransmitting the information collected.

She determines t_{X_i}, the time spent for the subordinate i's report about X, and t_{Y_i}, the time for i's report about Y, for each i. For simplicity, we regard r_i and s_i as choice variables, instead of t_{X_i} and t_{Y_i}. The higher is r_i or s_i, the more precisely top can understand the message from subordinate i concerning X or Y, respectively, since she spends more time in reading it. However, higher precisions lead to more costs of delay in decision making. We assume that these costs are linear in time spent for reading reports. Let $\hat{K}_D$ be the constant marginal cost of delay. Then the costs are of the form $\hat{K}_D t_{X_i} + \hat{K}_D t_{Y_i}$. Since we use r_i and s_i as decision variables, we rewrite the linear cost function as $K_D r_i + K_D s_i$ with $K_D = \hat{K}_D/\hat{p} > 0$ called a constant marginal *decoding cost*. Higher K_D reflects more need in quick decision making, given information processing capacities of top fixed.

Let $L(a_1, a_2, b_1, b_2 \mid Z_1, Z_2)$ be defined by

$$L(a_1, a_2, b_1, b_2 \mid Z_1, Z_2) = \mathrm{E}[l(a_1, a_2, b_1, b_2; X, Y) \mid Z_1, Z_2]$$

where $Z_i = (X_i + \theta_i, Y_i + \nu_i)$. Then in the hierarchical system, top chooses the *decision structure* α_1, α_2, β_1, and β_2, which are the functions of Z_1 and Z_2, such that for each Z_1 and Z_2,

$$(\alpha_1(Z_1, Z_2), \alpha_2(Z_1, Z_2), \beta_1(Z_1, Z_2), \beta_2(Z_1, Z_2)) \in \operatorname*{arg\,min}_{a_1, a_2, b_1, b_2} L(a_1, a_2, b_1, b_2 \mid Z_1, Z_2).$$

Note that the choice of the task structure does not matter under the hierarchical system, since all the actions are chosen by top and the subordinates have limited rationality in understanding subtle messages.

The horizontal coordination system. Figures 2a and 2b show the information flow and the action choice under the horizontal system with nonspecialized task structure and with specialized task structure, respectively. In the horizontal system, communication channels are installed between two subordinates. Each channel sends a message from one subordinate to the other. Subordinate 1 sends X_1 and Y_1 to subordinate 2, using one channel, and subordinate 2 sends X_2 and Y_2 to subordinate 1, using the other channel. Similar to top in the hierarchical system, each shop manager faces the tradeoff between benefits of better decision and costs of delay in decision making. Subordinate i observes the signal $Z_j = (X_j + \theta_j, Y_j + \nu_j)$ $(i \neq j)$. The cost of delay in his decision making is assumed to be of the form $\mathcal{K}_D(p_i)r_j + \mathcal{K}_D(q_i)s_j$ where the marginal cost $\mathcal{K}_D(\cdot) > 0$ is a decreasing function and r_j (s_j) is the precision with which the manager i reads the report from manager j concerning X (Y, respectively). Higher r_j means that manager i spends more time in reading the report concerning X. As his information processing capacity p_i concerning X is higher, he can understand the report concerning X from subordinate j more easily, so that the cost is less. After the interchange of information, each subordinate *independently* chooses his actions based on his original information and the signal received from the other subordinate. For example, when the nonspecialized task structure is adopted, subordinate i chooses the decision structure α_i and β_i, both of which are the functions of X_i, Y_i, and Z_j, such that for a given decision structure α_j and β_j of subordinate j,

$$(\alpha_i(X_i, Y_i, Z_j), \beta_i(X_i, Y_i, Z_j)) \in \arg\min_{a_i, b_i} L_i(a_i, b_i; \alpha_j, \beta_j \mid X_i, Y_i, Z_j) \tag{1}$$

for all X_i, Y_i, and Z_j, ($i, j = 1, 2$, and $i \neq j$), where

$$L_i(a_i, b_i; \alpha_j, \beta_j \mid X_i, Y_i, Z_j) = \mathrm{E}[l(a_i, \alpha_j(X_j, Y_j, Z_i), b_i, \beta_j(X_j, Y_j, Z_i); X, Y) \mid X_i, Y_i, Z_j].$$

Here $L_i(a_i, b_i; \alpha_j, \beta_j \mid X_i, Y_i, Z_j)$ is the expected value of the loss function when subordinate i chooses action a_i and b_i, given his signals (X_i, Y_i, Z_j) and the decision structure of subordinate j, α_j and β_j. Note that each subordinate i cannot exactly infer the actions implemented by the other subordinate j from his available information. Therefore, subordinate i has some conjectured decision structure of subordinate j, and responds optimally to it. We adopt the Bayesian-Nash equilibrium as an appropriate solution concept under the horizontal system, where the decision structure $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ is an equilibrium if the equation (1) is satisfied for $i, j = 1, 2$. In equilibrium, each subordinate has the correct conjectures concerning the other subordinate's decision structure.[12] Similar arguments apply to the case of the specialized task structure.

The decision problems in the model are summarized as follows. There are six decisions the organization has to make: (i) the coordination system; (ii) information processing capacities $I_C = (p_1, p_2, q_1, q_2)$; (iii) the precisions with which subordinates or the top manager

[12] This solution is equivalent to the person-by-person satisfactory rule in team theoretic terms. Thus, we can alternatively define the optimal decision structure (under the horizontal system with nonspecialized task structure) as the rules α_i and β_i for $i = 1, 2$, which top specifies *ex ante* to minimize the unconditional expected gross loss, given by $\mathrm{E}[l(\alpha_1(X_1, Y_1, Z_2), \beta_1(X_1, Y_1, Z_2), \alpha_2(X_2, Y_2, Z_1), \beta_2(X_2, Y_2, Z_1); X, Y)]$.

read reports, denoted by $I_D = (r_1, r_2, s_1, s_2)$; (iv) the time allocation by the subordinates $I_T = (t_1, t_2)$; (v) the task structure; and (vi) the decision structure. We call the decisions (ii)–(iv) the *information structure* of the firm[13] and denote $\mathcal{I} = (I_C, I_D, I_T)$; once this and the task structure are fixed for each coordination system, (vi) is solved as above. The optimal solutions of (i)–(vi) are called the *optimal organization structure.*

The analysis of the model follows the standard procedure of the theory of teams: First, for each coordination system, information structure, and task structure, we derive the optimal decision structure. Then we derive the optimal information structure, and finally the task structure and the optimal coordination system (if possible).

3. Preliminary Analysis

The abstract model presented in the last section is more complicated than it looks. In particular, it is very difficult to obtain even the optimal decision structure explicitly with λ as a parameter. Therefore we examine two extreme cases $\lambda = 0$ and $\lambda \to +\infty$. The first case corresponds to the case where inter-shop coordination is unnecessary. For example, this case will apply if the firm diversifies into two completely unrelated industries. The second case is that the inter-shop coordination is crucial as in the production system with assembly lines. The organization must determine the decision structure such that the λ-term almost always becomes zero. In these two cases, the tradeoff between two coordination structures becomes quite simple, as is shown below.

In the rest of this section, we provide some preliminary analysis of each coordination system in the two extreme cases of the inter-shop coordination term. The main results are that when $\lambda = 0$, the optimal organization structure is the horizontal coordination system with the specialized task structure, while when $\lambda \to +\infty$, this system is dominated by the hierarchical system for each information structure. Therefore, in the case where inter-shop coordination is extremely important, we can concentrate on the comparison between the hierarchical system and the horizontal system with the nonspecialized task structure, which will be conducted in the following sections.

[13] The pattern of communication channels is determined by the choice of the coordination system, so that it is not included in the information structure.

First consider the horizontal coordination system. Suppose that the specialized task structure is adopted so that subordinate 1 chooses a_1 and a_2, and subordinate 2 selects b_1 and b_2. Then the firm can ignore the μ_A-term and the μ_B-term by setting $a = a_1 + a_2$, $b = b_1 + b_2$, and $a_i = a/2$ and $b_i = b/2$, without affecting the value of the other terms. That is, the firm can resolve coordination within each shop completely. The serious disadvantage of this system becomes clear as λ takes a very high value because a and b are selected by different players independently. In particular, when $\lambda \to +\infty$, the expectation of the λ-term is zero if and only if, in equilibrium,

$$\alpha_1(X_1, Y_1, Z_2) + \alpha_2(X_1, Y_1, Z_2) = \beta_1(X_2, Y_2, Z_1) + \beta_2(X_2, Y_2, Z_1)$$

holds for *almost all* possible values of the signals observed by the shop managers. This is possible only when they set $\alpha_1 + \alpha_2$ and $\beta_1 + \beta_2$ to some same *constant* action almost everywhere. Clearly this is not desirable in view of the direct effect. Thus, when the coordination across shops is extremely important, the performance of the horizontal system with the specialized task structure will be very poor.

On the other hand, under the horizontal system with nonspecialized task structure, the intra-shop coordination terms become zero if and only if the decision structure of each subordinate is independent of all his information, by logic similar to above. Therefore, the performance will be quite bad as μ_A or μ_B takes very high values. However, when $\lambda \to +\infty$, the firm can set the expectation of the λ-term to zero. Subordinate i can choose his actions such as $\alpha_i \equiv \beta_i$ and still can utilize his on-the-spot information for his decision making. Hence the nonspecialized task structure under the horizontal system is likely to perform very well if $\lambda \to +\infty$ and μ_A and μ_B are small.

Finally, consider the hierarchical coordination system. Under this system, as with the horizontal system with the specialized structure, the firm can always zero out the intra-shop coordination terms, by ordering the shop managers to take the same actions for almost all messages received by top. In addition, since all the actions are chosen by the top manager, a very high value of λ will not be so harmful as in the horizontal system with the specialized task structure. In fact, when $\lambda \to +\infty$, the firm enables the expectation of the inter-shop coordination term to be zero by setting $\alpha_1(\cdot) + \alpha_2(\cdot) = \beta_1(\cdot) + \beta_2(\cdot)$ for each pair of messages received, without affecting the intra-coordination terms. Therefore at least in the two extreme cases considered, the disadvantage of the hierarchical coordination system comes solely from the fact that the on-the-spot information is not available to top.

The main result of this section is summarized in the following theorem.

THEOREM 1. (i) *If $\lambda = 0$, the optimal organization structure is the horizontal coordination system with the specialized task structure and the information structure such that each shop manager specializes in the information source only relevant to his shop and that no horizontal communication occurs.* (ii) *If $\lambda \to +\infty$, then for all information structures, the hierarchical coordination system is at least as good as the horizontal coordination system with the specialized task structure.*

This theorem is proved by the following three lemmas.

LEMMA 1. *Suppose $\lambda = 0$. Then under the horizontal coordination system with the specialized task structure, the optimal information structure $\mathcal{I} = (I_C, I_D, I_T)$ is given by $I_C = (p_1, 0, 0, q_2)$, $I_D = (0, 0, 0, 0)$, and $I_T = (1, 0)$ where $p_1 = \max[(K_C)^{-1/2} - h_X, 0]$ and $q_2 = \max[(K_C)^{-1/2} - h_Y, 0]$.*

PROOF: See Appendix. ∎

The information structure in the assertion means that subordinate 1 specializes in attending to X and subordinate 2 specializes in Y, and there is no communication between them. In the Appendix, it is shown that these capacities are optimal even under the assumption that $\mathcal{K}_D \equiv 0$, that is, there is no cost of communication between subordinates. Then the expected gross loss under the optimal decision structure is $\mathrm{Var}(X \mid X_1, X_2) + \mathrm{Var}(Y \mid Y_1, Y_2)$. The assertion in Lemma 1 states that under the optimal information structure, the expected gross loss is of the form $\mathrm{Var}(X \mid X_1) + \mathrm{Var}(Y \mid Y_2)$. That is, at the optimum, X_2 and Y_1 do not convey any information because manager 1 (manager 2) has no information processing capacity concerning Y (X) and so $1 - t_1 = 0$ ($t_2 = 0$, respectively).

In the proof in the Appendix, we see that top chooses the information processing capacities to achieve a target level of the posterior variance for each shop-specific environment. Under the optimal capacities, this target level is given as $\sqrt{K_C}$ for both shops. If the prior variance of a shop exceeds this level, top invests in the capacities of the shop manager and reduces the variance. If the prior variance is no higher than the target level, top does not try to reduce the variance further, so that there is no investment in information processing capacities. Communication between subordinates does not occur since they make decisions only relevant to their own job shop and they specialize in different information sources.

If the subordinates specialize in the same information source (say, X), they may reduce the variance of X, but to the *same* target level $\sqrt{K_C}$ while they cannot reduce the variance of Y. Therefore, specialization in the same information source is no better than specialization in different information sources.

Finally, suppose that each subordinate attends to both information sources. Then communication is valuable, and by horizontal communication and by the assumption of no costs of delay, they can share their information completely. However, under this nonspecialization, the target levels of the posterior variances are given by $\sqrt{K_C/t}$ and $\sqrt{K_C/(1-t)}$ where $t = t_1 = t_2$. Hence, the nonspecialization is inferior to the specialization in different information sources. Here limited attention matters. In fact, if there were no time constraint and subordinates' information processing capacities could be used as the precisions of the noise terms (that is, $p_i = h_{\epsilon_i}$ and $q_i = h_{\eta_i}$ for all i), then nonspecialization would perform as well as the specialization in different information sources (assuming that communication is costless).

Given the horizontal coordination system, the next lemma compares the specialized task structure (with the information structure in Lemma 1) to the nonspecialized task structure.

LEMMA 2. *Consider the horizontal coordination system. Then when $\lambda = 0$, the specialized task structure with the optimal information structure as in Lemma 1 is at least as good as the nonspecialized task structure.*

PROOF: To give the greatest advantage to the nonspecialized task structure, suppose $\mu_A = \mu_B = 0$ and no costs of delay. Since the optimal information structure under the specialized task structure has no communication, this assumption of no decoding cost does not give any advantage to the specialized structure. Then under the nonspecialized task structure, managers choose the decision structure satisfying

$$\alpha_1(X_1, X_2) + \alpha_2(X_1, X_2) = \mathrm{E}[X \mid X_1, X_2]$$

$$\beta_1(Y_1, Y_2) + \beta_2(Y_1, Y_2) = \mathrm{E}[Y \mid Y_1, Y_2].$$

Then the expected gross loss is of the form $\mathrm{Var}(X \mid X_1, X_2) + \mathrm{Var}(Y \mid Y_1, Y_2)$ under this decision structure. However, this is the same as the expected gross loss under the specialized task structure under the assumption of no decoding cost. Hence the nonspecialized task structure is no better than the specialized one. ∎

It follows from the proof that if the decoding cost were zero, the subordinates could share the same information, so that the performance of the nonspecialized task structure would be the same as that of the specialized one, *even if* μ_A or μ_B is strictly positive. This can be done by making the subordinates specialize in different information sources, by choosing the decision structure satisfying the equations in the proof, and by setting $\alpha_1 \equiv \alpha_2$ and $\beta_1 \equiv \beta_2$ for almost all values of the random variables. However, when there are positive decoding costs, this arrangement is no longer available, so that when $\mu_A > 0$ or $\mu_B > 0$, the nonspecialized task structure is strictly inferior to the specialized one.[14]

Finally, the following lemma compares the hierarchical system to the horizontal system with the specialized task structure.

LEMMA 3. (i) *Suppose* $\lambda = 0$. *Then for each information structure, the horizontal coordination system with the specialized task structure is at least as good as the hierarchical system.* (ii) *Suppose* $\lambda \to +\infty$. *Then for each information structure, the hierarchical coordination system is at least as good as the horizontal system with the specialized task structure.*

PROOF: (i) Given an information structure, when $\lambda = 0$, the optimal decision structure under the horizontal system with the specialized task structure solves

$$\min_a \mathrm{E}[(a - X)^2 \mid X_1, Y_1, Z_2] + \min_b \mathrm{E}[(b - Y)^2 \mid X_2, Y_2, Z_1].$$

On the other hand, the decision structure under the hierarchical system solves

$$\min_{a,b} \{\mathrm{E}[(a - X)^2 \mid Z_1, Z_2] + \mathrm{E}[(b - Y)^2 \mid Z_1, Z_2]\}.$$

[14] When $\mu_A = \mu_B = 0$, even if communication is costly, the nonspecialized structure can achieve the same performance as the specialized one because the optimal capacities are specialization in different information sources. That is, the following decision structure achieves this: $\alpha_1(X_1) = \mathrm{E}(X \mid X_1)$, $\beta_2(Y_2) = \mathrm{E}(Y \mid Y_2)$, and $\alpha_2 \equiv \beta_1 \equiv 0$. This decision structure shows that even under the nonspecialized task structure, each subordinate specializes in one of two shops and does not influence the other shop.

Clearly the former system is more "informative," so that the horizontal system with the specialized task structure is at least as good as the hierarchical system. (ii) As is discussed above, when $\lambda \to +\infty$, the optimal decision structure under the horizontal system with the specialized task structure either has to ignore all the information and set to the constant $\alpha \equiv \mathrm{E}(X) = 0$ and $\beta \equiv \mathrm{E}(Y) = 0$, or has to choose the decision structure such as $\alpha_2 \equiv -\alpha_1$ and $\beta_2 \equiv -\beta_1$. It is easily seen that the optimal decision structure under this second condition leads to $\alpha_i \equiv 0$ and $\beta_i \equiv 0$ for $i = 1, 2$, so that again the subordinates have to ignore all the information. However, under the hierarchical system, top can zero out the λ-term without ignoring the information Z_1 and Z_2 by setting $\alpha(Z_1, Z_2) = \beta(Z_1, Z_2)$. Hence the hierarchical system is more informative and at least as good as the specialized task structure under the horizontal system. ∎

Theorem 1 (i) follows from Lemmas 1, 2, and 3 (i). Theorem 1 (i) is very intuitive. It states that if the firm diversifies into two completely unrelated industries, the optimal organization structure is the M-form where each manager specializes in his division and the information source specific to his division. As Williamson (1975, 1985) states, the M-from is better than the U-form because the former can improve coordination by "grouping the operating parts into separable entities within which interactions are strong and between which they are weak (1985, p.283)." And the M-form is better than the hierarchical system since operating decisions are resolved at the divisions, which reduces the information overload of top. Of course, this result holds because the coordination across industries is unnecessary ($\lambda = 0$). As λ is higher, interactions between divisional units become stronger, and the advantage of the M-form may be canceled out by its disadvantage in inter-divisional coordination.[15]

On the other hand, Theorem 1 (ii) (which follows directly from Lemma 3 (ii)) states that when $\lambda \to +\infty$, we can exclude the horizontal system with the specialized task structure from consideration in order to derive the optimal organization structure. In the rest of this paper, we focus on the comparison between the hierarchical system and the horizontal system with the nonspecialized task structure under the assumption $\lambda \to +\infty$. Hereafter, we assume that the horizontal system has always the nonspecialized task structure.

Before starting the analysis, we adopt the following assumptions for simplification.

[15] A typical example of this is that autonomous activities of two related product divisions sometimes cause the problem that they produce the same kinds of products and compete with each other in the market.

ASSUMPTION (A-1). *$\mu_A = \mu_B = \mu$ and $\sigma_X^2 = \sigma_Y^2 = \sigma^2 = 1/h$.*

ASSUMPTION (A-2). *For all $r \geq 0$, $\mathcal{K}_D(r) \to +\infty$. That is, under the horizontal coordination system, no horizontal communication occurs between two subordinates.*

Assumption (A-1) is purely for simplicity of the analysis. We assume that two shops are symmetric in the sense that they have the same prior variance and that they have the same degree of importance of coordination within shop. The implication of relaxing this assumption is straightforward.

Solving even the optimal decision structure under the horizontal system with interchange of information is very difficult. Therefore we adopt Assumption (A-2) and consider the "pure" decentralized coordination system. (A-2) states that interchange of information and understanding the reports under the horizontal system will overload the managers prohibitively. Notice that this assumption gives the hierarchical system the best possible chance of being superior to the horizontal system, since permitting communication between subordinates always improves the performance of the horizontal system. I will discuss intuitively in Section 7 how the results change if interchange of information is allowed.

4. The Hierarchical Coordination System

Now we start the analysis of the case of $\lambda \to +\infty$ with the hierarchical system. As is discussed in Section 2, the μ-terms are zero in the optimum under the hierarchical coordination system. However, since $\lambda \to +\infty$, it is necessary to satisfy $a_1 + a_2 = b_1 + b_2$ for each pair of messages Z_1 and Z_2. Therefore, top chooses $a = a_1 + a_2$, given an information structure, to minimize $L(a, a \mid Z_1, Z_2) = \mathrm{E}[(a - X)^2 + (a - Y)^2 \mid Z_1, Z_2]$ for each Z_1 and Z_2. The optimal decision structure, denoted by α^H, is given by the following:

$$\alpha^H(Z_1, Z_2) = \frac{1}{2}\mathrm{E}[X \mid Z_1, Z_2] + \frac{1}{2}\mathrm{E}[Y \mid Z_1, Z_2]. \tag{2}$$

If there were no need in coordinating among shops, the optimal decision structure would be clearly $\alpha(Z_1, Z_2) = \mathrm{E}[X \mid Z_1, Z_2]$ and $\beta(Z_1, Z_2) = \mathrm{E}[Y \mid Z_1, Z_2]$. Hence the optimal structure α^H is the convex combination of these two structures. The optimal gross loss

$L^H(\mathcal{I})$ is given by

$$\begin{aligned} L^H(\mathcal{I}) &= \mathrm{E}[L(\alpha^H(Z_1, Z_2), \alpha^H(Z_1, Z_2) \mid Z_1, Z_2)] \\ &= \frac{1}{2}\{\mathrm{Var}(X) + \mathrm{Var}(X \mid Z_1, Z_2) + \mathrm{Var}(Y) + \mathrm{Var}(Y \mid Z_1, Z_2)\}. \end{aligned} \tag{3}$$

Then top chooses the optimal information structure $\mathcal{I}$ which minimizes the expected net loss $L^H(\mathcal{I}) + K_C \sum_{i=1}^{2}(p_i + q_i) + K_D \sum_{i=1}^{2}(r_i + s_i)$. The main result in this section is the following.

THEOREM 2. *Suppose $\lambda \to +\infty$ and (A-1). Then, under the hierarchical system, the optimal information structure is the specialization in different information sources. That is, one subordinate, say subordinate 1, specializes in X ($p_1 > 0$, $q_1 = 0$, and $t_1 = 1$) and the other subordinate (subordinate 2) in Y ($p_2 = 0$, $q_2 > 0$, and $t_2 = 0$).*

The proof of Theorem 2 is similar to that of Lemma 1. The procedure is to compare the specialization in different information sources with the specialization in the same information source and with the nonspecialization. The only difference from the proof of Lemma 1 is that we utilize the following lemma to compute the optimal information processing capacities given nonspecialization.

LEMMA 4. *Suppose (A-1). If the optimal information processing capacities under the hierarchical system make subordinates generalists in the sense that $p_i > 0$ and $q_i > 0$ for $i = 1, 2$, then their capacities satisfy $p_1 = q_1$ and $p_2 = q_2$, and they allocate their time evenly to each information source, that is, $t_1 = t_2 = 1/2$.*

PROOF: See Appendix. ∎

This lemma simplifies the calculation of the optimal information processing capacities and time allocation if nonspecialized capacities are optimal.

Here we only show the derivation of the optimal information structure given the specialization in different information sources. Note that when subordinates are specialists, the message from each subordinate to top is one-dimensional. He sends the report concerning what he specializes in. Hence, we can denote $Z_1 = X_1 + \theta_1$ and $Z_2 = Y_2 + \nu_2$. We suppose without loss of generality that subordinate 1 specializes in X and subordinate 2 to

Y. When the subordinates specialize in different information sources as such, the expected gross loss is of the form

$$L^H(\mathcal{I}) = \frac{1}{2}(h_X^{-1} + h_Y^{-1}) + \frac{1}{2}\{\mathrm{Var}(X \mid Z_1) + \mathrm{Var}(Y \mid Z_2)\}$$
$$= \frac{1}{2}(h_X^{-1} + h_Y^{-1}) + \frac{1}{2}\left\{\left(h_X + \frac{p_1 r_1}{p_1 + r_1}\right)^{-1} + \left(h_Y + \frac{q_2 s_2}{q_2 + s_2}\right)^{-1}\right\}.$$

Notice $h_{\epsilon_1} = p_1$ and $h_{\eta_2} = q_2$ since $t_1 = 1 - t_2 = 1$. Then top chooses the information structure to minimize the expected net loss $L^H(\mathcal{I}) + K_C(p_1 + q_2) + K_D(r_1 + s_2)$. It can be shown that this objective function is strictly convex in (p_1, q_2, r_1, s_2), so that the first-order conditions are necessary and sufficient. The first-order conditions for $p_1 > 0$ and $r_1 > 0$ are given by

$$\begin{aligned} \left(h_X + \frac{p_1 r_1}{p_1 + r_1}\right)^{-2} \left(\frac{r_1}{p_1 + r_1}\right)^2 &= 2K_C \\ \left(h_X + \frac{p_1 r_1}{p_1 + r_1}\right)^{-2} \left(\frac{p_1}{p_1 + r_1}\right)^2 &= 2K_D. \end{aligned} \tag{4}$$

From two equations in (4), we obtain

$$r_1 = \frac{k_C}{k_D} p_1 \tag{5}$$

where $k_i = \sqrt{K_i}$ for $i = C, D$. Therefore

$$\frac{r_1}{p_1 + r_1} = \frac{k_C}{k} \tag{6}$$

where $k = k_C + k_D$. Substituting (5) and (6) into the first equation in (4) yields

$$h_X + \frac{p_1 r_1}{p_1 + r_1} = h_X + \frac{k_C}{k} p_1 = \frac{1}{\sqrt{2}k}. \tag{7}$$

By (7), (5), and the usual complementary slackness conditions for corner solutions, the optimal values of p_1 and r_1, denoted by $p_1^\star$ and $r_1^\star$, are as follows.

$$\begin{aligned} p_1^\star &= \max\left[k_C^{-1}\left((\sqrt{2})^{-1} - k h_X\right), 0\right] \\ r_1^\star &= \max\left[k_D^{-1}\left((\sqrt{2})^{-1} - k h_X\right), 0\right]. \end{aligned} \tag{8}$$

By the same procedure, we obtain

$$\begin{aligned} q_2^\star &= \max\left[k_C^{-1}\left((\sqrt{2})^{-1} - k h_Y\right), 0\right] \\ s_2^\star &= \max\left[k_D^{-1}\left((\sqrt{2})^{-1} - k h_Y\right), 0\right]. \end{aligned} \tag{9}$$

The optimal information structure is given by $\mathcal{I}^\star = (p_1^\star, 0, 0, q_2^\star; r_1^\star, 0, 0, s_2^\star; 1, 0)$. Let $C^H(\mathcal{I})$ be the total costs of information processing capacities and decoding. Then

$$\begin{aligned} L^H(\mathcal{I}^\star) &= \frac{1}{2}(h_X^{-1} + h_Y^{-1}) + \frac{1}{2}\{\min[\sqrt{2}k, h_X^{-1}] + \min[\sqrt{2}k, h_Y^{-1}]\} \\ C^H(\mathcal{I}^\star) &= \max[k((\sqrt{2})^{-1} - kh_X), 0] + \max[k((\sqrt{2})^{-1} - kh_Y), 0]. \end{aligned} \tag{10}$$

As in the optimal information structure in the last section, the firm seeks a target level, given by $\sqrt{2}k$, for each shop. When the prior variance is higher than this, top invests in information processing capacities of shop managers and spends time reading reports from them, while when the prior variance is no higher than the target level, she makes no investment.

When the subordinates specialize in the same information source, the target level is the same. Since they can reduce the variance of only one shop, this is inferior to the specialization in different information sources.[16] When nonspecialization is chosen, the target level of the posterior variance is larger than that under the specialization because of limited attention.

For later references, we summarize the results in the following corollary.

COROLLARY 1. *Under Assumption (A-1) and $\lambda \to +\infty$, the optimal information structure under the hierarchical system is given by $\mathcal{I}^H = (p^H, 0, 0, p^H; r^H, 0, 0, r^H; 1, 0)$ with*

$$\begin{aligned} p^H &= \max[k_C^{-1}((\sqrt{2})^{-1} - kh), 0] \\ r^H &= \max[k_D^{-1}((\sqrt{2})^{-1} - kh), 0]. \end{aligned} \tag{11}$$

The optimal values of the expected gross loss, the information costs, and the expected net loss are given by

$$\begin{aligned} L^H &= h^{-1} + \min[\sqrt{2}k, h^{-1}] \\ C^H &= \max[k(\sqrt{2} - 2kh), 0] \\ L^H + C^H &= h^{-1} + \min[2k(\sqrt{2} - kh), h^{-1}]. \end{aligned} \tag{12}$$

[16] This result partially depends upon the linearity of the decoding costs. If there were scale economies such as learning effects, so that top could read two summary reports concerning the same information source faster than reading two reports of different sources, then the specialization in the same information source might be better.

When $h < (\sqrt{2}k)^{-1}$, the firm invests in the information capacities of both subordinates, so that $L^H = h^{-1} + \sqrt{2}k < 2h^{-1}$ and $C^H = k(\sqrt{2} - 2kh) > 0$. Note that the target level $\sqrt{2}k$ does not depend on the prior precision h. In the next section, we will see that this is not necessarily true under the horizontal system. Also C^H is linearly decreasing in h. Hence the expected net cost is decreasing in the prior precision. The higher the prior precision is, the lower information costs are required to achieve the target. When $h \geq (\sqrt{2}k)^{-1}$, the firm does not accumulate any resource in subordinates since the prior variances of the environment are sufficiently low. Therefore the expected net loss is the same as the sum of the prior variances, $2h^{-1}$. The effect of the capacity cost and the decoding cost can be easily seen, too. An increase in these costs reduces the critical level of the prior precision, beyond which the firm makes no investments, and so increases the target level. Therefore the expected net loss also increases.

5. The Horizontal Coordination System

The optimal decision structure under the horizontal coordination system is more complicated since, as is seen in Section 3, in the optimum, the μ-term is not zero. Also a_1 and b_1, and a_2 and b_2 are independently selected by subordinates 1 and 2, respectively, who share the information imperfectly.

Since $\lambda \to +\infty$, the decision structure has to be chosen such that the λ-term is zero. Under the horizontal system, this can be done by setting, for each i, $\alpha_i = \beta_i$ for almost all values of the signals.

We can derive the optimal structure for every value of μ under Assumption (A-2) using the well-known result in team theory that the optimal decision function of each subordinate is linear in the signals he observes under the assumptions of Normal distributions and a quadratic loss function. (See Marschak and Radner, 1972.) The optimal decision structure is of the form

$$\begin{aligned}\alpha_i(X_i, Y_i) = {} & \frac{1}{2}\frac{(1+\mu)h_X^2 - (1-\mu)h_X h_{X_j}}{(1+\mu)^2 h_X^2 - (1-\mu)^2 h_{X_1} h_{X_2}} \mathrm{E}[X \mid X_i] \\ & + \frac{1}{2}\frac{(1+\mu)h_Y^2 - (1-\mu)h_Y h_{Y_j}}{(1+\mu)^2 h_Y^2 - (1-\mu)^2 h_{Y_1} h_{Y_2}} \mathrm{E}[Y \mid Y_i]\end{aligned} \tag{13}$$

where $i \neq j$. Because of serious need for inter-shop coordination, each decision structure depends on the information concerning both shops, X_i and Y_i. One simple observation is that when μ goes to infinity, all the actions go to zero. When the coordination within each shop is extremely important, the firm needs to achieve $\alpha_1 \equiv \alpha_2$. Under the horizontal system, however, this is possible only when both subordinates ignore almost all their information, so that they simply set $\alpha_1(\cdot) = \alpha_2(\cdot) = \frac{1}{2}\mathrm{E}(X) + \frac{1}{2}\mathrm{E}(Y) = 0$ almost everywhere. Clearly in this case, the hierarchical system works better since top can use the noisy messages from subordinates as well as achieve $\alpha_1 \equiv \alpha_2$.

The optimal expected gross loss is, by tedious calculations, obtained as

$$\begin{aligned} L(\mathcal{I}) = {} & \frac{1}{2}\left\{h_X^{-1} + h_X^{-1}\left(1 - \frac{(1+\mu)(h_{X_1}+h_{X_2})h_X - 2(1-\mu)h_{X_1}h_{X_2}}{(1+\mu)^2 h_X^2 - (1-\mu)^2 h_{X_1}h_{X_2}}\right)\right\} \\ & + \frac{1}{2}\left\{h_Y^{-1} + h_Y^{-1}\left(1 - \frac{(1+\mu)(h_{Y_1}+h_{Y_2})h_Y - 2(1-\mu)h_{Y_1}h_{Y_2}}{(1+\mu)^2 h_Y^2 - (1-\mu)^2 h_{Y_1}h_{Y_2}}\right)\right\}. \end{aligned} \tag{14}$$

By Assumptions (A-1) and (A-2), the problem is to obtain the optimal information processing capacities $I_C = (p_1, p_2, q_1, q_2)$ and the time allocation $I_T = (t_1, t_2)$ to minimize $L(\mathcal{I}) + K_C \sum_{i=1}^{2}(p_i + q_i)$. If we obtain the interior solution ($p_i > 0$, $q_i > 0$ and $t_i \in (0,1)$ for all i), then the optimal information processing capacities involve nonspecialization. Otherwise, the optimal capacities involve specialization either in the same information source or in different information sources. Since the second-order conditions are complicated, it is not easy to derive the optimal information capacities and the optimal time allocation directly from (14). Therefore, we consider separately the three cases mentioned above: the case of nonspecialization (called the N-horizontal system), the case of the specialization in different information sources (the SD-horizontal system) and the case of the specialization in the same information source (the SS-horizontal system). Then we compare the performance of the three cases to obtain the optimal information processing capacities and the optimal coordination system.

3.5.1. The N-Horizontal System

If the N-horizontal system is the optimal horizontal system, both subordinates are generalists in the sense that $p_i > 0$, $q_i > 0$, and $t_i \in (0,1)$ for $i = 1, 2$. Though I am almost sure that the result similar to Lemma 4 ($p_i = q_i$ and $t_i = 1/2$) holds under the horizontal system, I could not prove it formally.[17] Thus, we take this as an assumption:

[17] It is easily shown that $p_i = q_i$ and $t_i = 1/2$ satisfy the first-order necessary conditions.

ASSUMPTION (A-3). *Suppose (A-1) and (A-2). If the optimal capacities under the horizontal system are generalists, then assume $p_i = q_i$ and $t_i = 1/2$ hold for $i = 1, 2$.*

By Assumption (A-3) and symmetry, the optimal capacities satisfy $p_1 = q_1 = p_2 = q_2$ so that $h_{X_1} = h_{Y_1} = h_{X_2} = h_{Y_2}$ holds. Let $g(p,t) = hpt/(h + pt)$ be the precision of the signal observed by a subordinate when p is his information processing capacity and he spends time t for information acquiring. Then (14) is rewritten as

$$\begin{aligned} L^N(\mathcal{I}) &= \frac{1}{h}\left\{1 - \frac{2g(p,\frac{1}{2})}{(1+\mu)h + (1-\mu)g(p,\frac{1}{2})}\right\} + h^{-1} \\ &= \frac{1+\mu}{p + (1+\mu)h} + h^{-1}. \end{aligned} \tag{15}$$

The firm chooses p to minimize $L^N(\mathcal{I}) + 4K_C p$. Clearly this objective function is strictly convex in p. Thus, by solving the first-order condition for the interior solution, we obtain the optimal information processing capacity p^N under nonspecialization as follows.

$$p^N = \max\left[\frac{\sqrt{m}}{2k_C} - mh, 0\right] \tag{16}$$

where $m = 1 + \mu$. Hereafter we use m rather than μ as the parameter representing the importance of the intra-shop coordination. Note that $m \geq 1$, and when $m = 1$, there is no need in coordination within each shop. The optimal values of the expected gross loss, the information cost, and the expected net loss are given as follows.

$$\begin{aligned} L^N &= \min\left[2\sqrt{m}k_C, h^{-1}\right] + h^{-1} \\ C^N &= \max\left[2\sqrt{m}k_C - 4mk_C^2 h, 0\right] \\ L^N + C^N &= \min\left[4\sqrt{m}k_C - 4mk_C^2 h, h^{-1}\right] + h^{-1}. \end{aligned} \tag{17}$$

The firm pursues a target level $2\sqrt{m}k_C$ if the prior variance exceeds this level. Note that *this target level is independent of the prior variance* as is the target level under the hierarchical system. This is the most important consequence of knowledge sharing between the shop managers in our quadratic model: Since the shop managers can share information concerning both X and Y (partially), they can coordinate their actions such as to set and achieve the target level independent of the prior precision. Because of this property, when $h < \left(2\sqrt{m}k_C\right)^{-1}$, the expected net loss is linear and decreasing in h. What is different from the hierarchical system, however, is that both this critical value of h and the target level depend upon the importance of the intra-shop coordination effect m. If m becomes larger, then the target level increases, so that the performance of the system becomes worse. The reason is that as the intra-shop coordination term is larger, on-the-spot information is less valuable: as we saw in Section 3, deeper knowledge makes it more difficult for each subordinate to take the similar actions. Therefore the firm makes less investments in the information processing capacities, which results in the higher expected net loss.

3.5.2. The SD-Horizontal System.

Here it is assumed that subordinate 1 specializes in X and subordinate 2 in Y. Then since $q_1 = p_2 = 0$, $h_{Y_1} = 0$ and $h_{X_2} = 0$ hold in (13). Also $\mathrm{E}[X \mid X_2] = \mathrm{E}(X) = 0$ and $\mathrm{E}[Y \mid Y_1] = \mathrm{E}(Y) = 0$. Therefore, the optimal decision structure under the SD-horizontal system is given by

$$\alpha_1^{SD}(X_1) = \frac{1}{2m}\mathrm{E}[X \mid X_1] \qquad \text{and} \qquad \alpha_2^{SD}(Y_2) = \frac{1}{2m}\mathrm{E}[Y \mid Y_2].$$

Since subordinate 2 specializes in Y, he does not receive any signal about X. Therefore, though he is involved in the management of job shop A as well as job shop B, his actions depend only on the information specific to job shop B. Then subordinate 1 chooses his actions by considering the tradeoff between the direct effect and the intra-shop coordination effect. If his action depends heavily on his on-the-spot information X_1, then since subordinate 2 has no knowledge concerning X, coordination is very difficult, whence the performance will not be good. On the other hand, too much attention to the coordination within each shop loses the advantage of on-the-spot information. Therefore, the optimal action by subordinate 1 depends on m. As above, the higher m is, the lower is the weight on $\mathrm{E}[X \mid X_1]$. When the intra-shop coordination effect is more important, subordinate 1 utilizes his on-the-spot information less.

Then by substituting $h_{Y_1} = h_{X_2} = 0$ into (14) and using (A-1) the optimal expected gross loss is given by

$$L^{SD}(\mathcal{I}) = \frac{1}{h}\left\{1 - \frac{g(p,1)}{mh}\right\} + h^{-1} \tag{18}$$

where $g(p,1) = h_{X_1} = h_{Y_2} = hp/(h+p)$ for $p = p_1 = q_2$. Then the firm chooses p to minimize $L^{SD}(\mathcal{I}) + 2K_C p$. The optimal information processing capacity p^{SD} is obtained as follows.

$$p^{SD} = \max[(\sqrt{2m}k_C)^{-1} - h, 0]. \tag{19}$$

Notice that $p^{SD} > 0$ when $h < (\sqrt{2m}k_C)^{-1}$. The optimal values of the expected gross loss, the information cost, and the expected net loss are given by

$$\begin{aligned} L^{SD} &= \min\left[\frac{1}{mh}(m - 1 + \sqrt{2m}k_C h), h^{-1}\right] + h^{-1} \\ C^{SD} &= \max\left[\frac{1}{mh}(\sqrt{2m}k_C h - 2mk_C^2 h^2), 0\right] \\ L^{SD} + C^{SD} &= \min\left[\frac{1}{mh}(m - 1 + 2\sqrt{2m}k_C h - 2mk_C^2 h^2), h^{-1}\right] + h^{-1}. \end{aligned} \tag{20}$$

Note that unless $m = 1$, the target level depends on the prior precision h, which is the important difference from the hierarchical system and the N-horizontal system. When $m = 1$, the coordination within each shop is unnecessary. Therefore as the prior precision decreases, the expected net loss increases only because of the linear increase in the information cost. However, under the specialization in different information sources, the information the subordinates obtain is stochastically independent, so that they share no knowledge and so they cannot resolve the intra-shop coordination at all. Thus, when $m > 1$, decreasing the prior precision results in increasing the expected net loss at an increasing rate.

3.5.3. The SS-Horizontal System.

Without loss of generality, it is assumed that both subordinates specialize in X. Then since $q_1 = q_2 = 0$, $h_{Y_1} = h_{Y_2} = 0$ holds in (13). In addition, $\mathrm{E}[Y \mid Y_1] = \mathrm{E}[Y \mid Y_2] = \mathrm{E}(Y) = 0$. Therefore the optimal decision structure, under Assumptions (A-1) and (A-2), is given by

$$\alpha_i^{SS}(X_i) = \frac{1}{2}\frac{(1+\mu)h^2 - (1-\mu)hh_{X_j}}{(1+\mu)^2h^2 - (1-\mu)^2h_{X_1}h_{X_2}}\mathrm{E}[X \mid X_i] \tag{21}$$

for $i, j = 1, 2$ and $i \neq j$. By symmetry ($h_{X_1} = h_{X_2}$), the optimal expected gross loss is obtained as

$$L^{SS}(\mathcal{I}) = \frac{1}{h}\left\{1 - \frac{g(p,1)}{(1+\mu)h + (1-\mu)g(p,1)}\right\} + h^{-1} \tag{22}$$

for $p = p_1 = p_2$. The firm chooses p to minimize $L^{SS}(\mathcal{I}) + 2K_C p$. The optimal capacity p^{SS} under the SS-horizontal system is given by

$$p^{SS} = \max\left[\frac{m}{2}\left((\sqrt{2m}k_C)^{-1} - h\right), 0\right]. \tag{23}$$

The optimal values of the expected gross loss, the information cost, and the expected net loss are obtained as

$$\begin{aligned} L^{SS} &= \min\left[\frac{1}{2h}\left(1+\sqrt{2m}k_C h\right), h^{-1}\right] + h^{-1} \\ C^{SS} &= \max\left[\frac{1}{2h}\left(\sqrt{2m}k_C - 2mk_C^2 h\right), 0\right] \\ L^{SS}+C^{SS} &= \min\left[\frac{1}{2h}\left(1+2\sqrt{2m}k_C h - 2mk_C^2 h^2\right), h^{-1}\right] + h^{-1}. \end{aligned} \tag{24}$$

Under the SS-horizontal system, for all m, the target level in L^{SS} depends on the prior variance. However, the reason for this is quite different from the reason under the SD-horizontal system. In (24), when $h < \left(\sqrt{2m}k_C\right)^{-1}$, L^{SS} is of the form

$$L^{SS} = \frac{1}{2}\left(\sqrt{2m}k_C + h^{-1}\right) + h^{-1}.$$

The prior variance $1/h$ always appears in the parentheses not because the shop managers cannot share knowledge about Y but simply because they completely ignore the information source Y. Therefore, as h decreases, the expected net loss increases at an increasing rate if $h < \left(\sqrt{2m}k_C\right)^{-1}$. However, since both subordinates focus on X and share knowledge about it, they can coordinate the part of their decision relevant to X very well: the "target level" related with this information source, given by $\sqrt{2m}k_C$, is independent of h and is lower than either target level of the N-horizontal system or the SD-horizontal system.

In contrast to the hierarchical system, the horizontal system can have all the three kinds of information structures at the optimum, and which of them is the best depends upon exogenous parameters. In the next section, we examine this and derive the optimal coordination system.

6. Comparisons and the Optimal Organization Structure

We compare the four systems derived in the last two sections by the *optimal cost saving*, which is defined as the expected net loss under no information ($2h^{-1}$) minus the optimal value of the expected net loss ($L+C$) for each system. Let $S^i = 2h^{-1} - (L^i + C^i)$ be the optimal cost saving for each system $i = H, N, SD, SS$. Then they are given as follows:

$$
\begin{aligned}
S^H(h,k) &= h^{-1}\left(1-\sqrt{2}kh\right)^2\delta\left\{h < \left(\sqrt{2}k\right)^{-1}\right\} \\
S^N(h,m,k_C) &= h^{-1}\left(1-2\sqrt{m}k_C h\right)^2\delta\left\{h < \left(2\sqrt{m}k_C\right)^{-1}\right\} \\
S^{SD}(h,m,k_C) &= \left(mh\right)^{-1}\left(1-\sqrt{2m}k_C h\right)^2\delta\left\{h < \left(\sqrt{2m}k_C\right)^{-1}\right\} \\
S^{SS}(h,m,k_C) &= \left(2h\right)^{-1}\left(1-\sqrt{2m}k_C h\right)^2\delta\left\{h < \left(\sqrt{2m}k_C\right)^{-1}\right\}
\end{aligned}
\tag{25}
$$

where $\delta\{E\} = 1$ if E holds, and $= 0$, otherwise.

The graph of the optimal cost saving under each system as a function of h is drawn in Figures 3–6. Each graph has a similar shape: It is smooth, decreasing, and convex in h. Particularly, it is equal to zero when the prior precision is equal to or higher than its critical value, while if the prior precision is smaller than the critical level, the optimal cost saving is strictly increasing at an increasing rate as h decreases. The effect of m on the optimal cost saving is also shown in the figures. The optimal cost saving under the three horizontal coordination systems decreases as the importance of intra-shop coordination increases. Though not shown in the figures, the effect of an increase in the marginal cost of capacities or decoding is also straightforward. An increase in the marginal cost of decoding K_D reduces the optimal cost saving under the hierarchical system. The increase in the marginal cost of capacities K_C reduces the optimal cost saving under each system.

Some ordering among four systems is easily established by comparing the optimal cost saving of each system in (25).

Proposition 1. *Suppose (A-1), (A-2), (A-3), and* $\lambda \to +\infty$.

(i) $S^N(h,m,k_C) \geq S^H(h,k)$ *for all* h *if and only if* $m \leq k^2/(2k_C^2)$.

(ii) $S^{SD}(h,m,k_C) \geq S^{SS}(h,m,k_C)$ *for all* h *and* k_C *if and only if* $m \leq 2$.

Assertion (i) shows that the ordering between the hierarchical system and the N-horizontal system does not depends upon the prior precision. If $m < k^2/(2k_C^2)$, then for *all* h, the N-horizontal system is at least as good as the hierarchical system (strictly better for small h). Since this condition means $2\sqrt{m}k_C < \sqrt{2}k$, the N-horizontal system can save the loss by making investments in the capacities for higher precisions than the hierarchical system to achieve lower target levels. In addition, when the firm makes investments under both systems, the slope of the information cost function under the hierarchical system is steeper, whence S^N increases faster than S^H as h decreases. If $m > k^2/(2k_C^2)$, then for *all* h the hierarchical system is at least as good as the N-horizontal system since the former can save the loss for higher h in order to achieve a lower target, and the optimal saving S^H increases faster than S^N. Finally when $m = k^2/(2k_C^2)$, the optimal cost saving of both systems coincides for all h, since the expected net losses coincide.

The disadvantage of the hierarchical system comes solely from the fact that it cannot utilize on-the-spot knowledge held by the shop managers and that the top manager has to spend time to read their reports, which causes delay of decision making. A parameter associated with this cost of delay is given by k_D. On the other hand, the disadvantage of the N-horizontal system is due to imperfect intra-shop coordination, which is more serious the higher is m. Proposition 1 shows that the ordering of these two systems is determined by the comparison between these disadvantages only.

Why does the prior precision have no role in determining the order between these two systems? Because both of them can set and achieve target levels independently of the prior variance $1/h$. In fact, the optimal values of the expected gross loss under these two systems have very similar forms. Under the hierarchical system, the firm incurs the costs of both information processing capacities and information decoding. By (6), the precision of the noise term in the signal observed by top is equal to $p^H r^H/(p^H + r^H) = (k_C/k)p^H$. Hence the expected gross loss under the hierarchical system is

$$L^H = \left(h + \frac{k_C}{k}p^H\right)^{-1} + h^{-1}.$$

The first term is the posterior variance obtained by top if she had an information processing capacity p^H and attended to X for time k_C/k. The time $1 - (k_C/k)$ is left idle because of the loss due to limited attention of top. On the other hand, by (15), the expected gross loss under the N-horizontal system is

$$L^N = \left(h + \frac{1}{m}p^N\right)^{-1} + h^{-1}.$$

The first term is the posterior variance obtained if a manager with capacity p^N attended to an information source for time $1/m$. Though there is the positive μ-term under the N-horizontal system in addition to the direct effect term, subordinates who have nonspecialized capacities and allocate their time evenly to each information source can achieve the same expected gross loss as if $\mu = 0$ and each subordinate specialized in different information sources and spent time $1/m$ only for his information source. The time $(m-1)/m$ is left idle, which represents the loss due to imperfect intra-shop coordination.

As an example, suppose $k_C = k_D$, so that both systems have the same information cost function $4K_C p$. (Note $p^H = r^H$ if $k_C = k_D$.) The critical value for m is given by $k^2/(2k_C^2) = 2$. Then when $m = 2$, both L^N and L^H are given by $(h+\frac{1}{2}p)^{-1}+h^{-1}$. Therefore $p^H = p^N$ holds, and they have the same optimal cost saving. When $m < 2$, under the N-horizontal system, the subordinates put more weight on each conditional expectation in their decision structure in order to exploit the advantage of the on-the-spot information: the malcoordination due to more aggressive use of the on-the-spot information is less harmful. The loss of the N-horizontal system represented by the idle time above also decreases. Similarly, when $m > 2$, the weight on each conditional expectation in the optimal decision structure of the N-horizontal system decreases and the on-the-spot information is utilized less, because of the greater importance of the intra-shop coordination effect.[18]

Assertion (ii), which compares the SD-horizontal system and the SS-horizontal system, is similar to assertion (i); the ordering does not depend on h. We can show that for each $h < \left(\sqrt{2m}k_C\right)^{-1}$, the slope of S^{SD} is steeper than that of S^{SS} if $m < 2$ and is flatter if $m > 2$. When $m = 2$, the optimal cost saving under both systems coincides.

To make clear the relation between these two systems, first consider the case that there is no intra-shop coordination term ($m = 1$). Then the SS-horizontal system is inferior to the SD-horizontal system because the SS-horizontal system cannot reduce the prior variance of the environment Y specific to shop B, so that the expected net loss increases drastically when h decreases. While the expected net loss of the SD-horizontal system also increases as h decreases, the firm can reduce the expected net loss *partially* to the target level independent of h, by attending to both information sources. Thus, L^{SD} does not increases so fast as L^{SS} does as h decreases. Without any need for intra-shop coordination, the SD-horizontal system uses on-the-spot information more efficiently than the SS-horizontal system.

[18] When k_D is different from k_C, these systems are different in the information cost function as well as the expected gross loss. Therefore even when $m = k^2/(2k_C^2)$, p^N is not equal to p^H.

However, when the intra-shop coordination term is present ($m > 1$), this advantage of the SD-horizontal system may be lost. Under the SD-horizontal system, each subordinate specializes in a *different* information source, while, under the SS-horizontal system, both subordinates specialize in the *same* information source. Therefore, the SS-horizontal system is better in coordinating the actions between the subordinates. Proposition 1 (ii) states that, under $m < 2$, the advantage of the SS-horizontal system in better coordination is not sufficient to offset the disadvantage in the inefficient use of the on-the-spot information, so that the SD-horizontal system is still better for all h. When $m > 2$, however, the coordination is important sufficiently to make the SS-horizontal system dominate the SD-horizontal system for all h.

Other simple cases are when there is no intra-coordination term, that is, when $m = 1$, and when the intra-coordination effect is so important that $m \geq k^2/k_C^2$ holds. The best organization structure does not depend on h in these cases as the following proposition shows.[19]

PROPOSITION 2. *Suppose (A-1), (A-2), (A-3), and $\lambda \to +\infty$.*

(i) *When $m = 1$, the optimal organization structure is the SD-horizontal coordination system.*

(ii) *When $m \geq k^2/k_C^2$, the optimal organization structure is the hierarchical coordination system (with the optimal information structure in Corollary 1.)*

PROOF: (i) By Proposition 1, it is sufficient to compare this system with the N-horizontal system. By (15), when $m = 1$, $L^N = (p^N + h)^{-1} + h^{-1}$, while by (18), $L^{SD} = (p^{SD} + h)^{-1} + h^{-1}$: They have the same form of the expected gross loss. However, the information cost side is different. While the N-horizontal system needs to invest in four information processing capacities p_1, p_2, q_1, q_2, the SD-horizontal system needs only two capacities p_1 and q_2. Thus, the SD-horizontal system is better than the N-horizontal system. (ii) When $m \geq k^2/k_C^2$, the hierarchical system is better than the N-horizontal system by Proposition 1. For the comparison with the SD-horizontal system or the SS-horizontal system, note that when $m = k^2/k_C^2$, all of the hierarchical, SD-horizontal, and SS-horizontal systems have the same critical value of h; the firm makes investments in the information structure only if the prior precision is smaller than that same critical value. However, the optimal

[19] Of course, if h is sufficiently high, no investment in the information structure is optimal, hence all systems are indifferent.

cost saving under the hierarchical system increases faster than under the other two systems. As we saw in the last section, the target level under the hierarchical system is independent of h, while the target levels under the other two are not. ∎

When $m \in (1, k^2/k_C^2)$, the optimal organization structure depends on the prior precision as well as other parameters. To examine this case, for each $i = SD, SS$, define for $m \in [1, k^2/k_C^2]$,

$$\begin{aligned} N^i(m, k_C) &= \sup\{h \geq 0 \mid S^N > S^i, S^N > 0\} \\ H^i(m, k_C, k_D) &= \sup\{h \geq 0 \mid S^H > S^i, S^H > 0\} \end{aligned} \tag{26}$$

and

$$\begin{aligned} N(m, k_C) &= \begin{cases} N^{SD}(m, k_C) & \text{if } m \leq 2 \\ N^{SS}(m, k_C) & \text{if } m > 2 \end{cases} \\ H(m, k_C, k_D) &= \begin{cases} H^{SD}(m, k_C, k_D) & \text{if } m \leq 2 \\ H^{SS}(m, k_C, k_D)\cdot & \text{if } m > 2 \end{cases} \end{aligned} \tag{27}$$

Figures 7 and 8 display how $N(\cdot)$ and $H(\cdot)$ are determined. By definition, the N-horizontal system is better than both the SD- and SS-horizontal systems if $h < N(m, k_C)$. Similarly, the hierarchical system is better than both the SD- and SS-horizontal system if $h < H(m, k_C, k_D)$. In Appendix, we provide the exact forms of $N(\cdot)$ and $H(\cdot)$ and show the following properties.

LEMMA 5. *$N(\cdot)$ and $H(\cdot)$ have the following properties for $m \in [1, k^2/k_C^2]$.*

(i) *$N(\cdot)$ and $H(\cdot)$ are continuous in their arguments.*

(ii) *$N(1, k_C) = H(1, k_C, k_D) = 0$ and $N(m, k_C) = H(m, k_C, k_D)$ for $m = k^2/(2k_C^2)$.*

(iii) *$H(m, k_C, k_D) = \left(\sqrt{2}k\right)^{-1}$ for $m = k^2/k_C^2$.*

(iv) *$N(\cdot)$ is increasing and concave in $m < 2$, decreasing and convex in $m > 2$, and decreasing in k_C.*

(v) *$H(\cdot)$ is increasing and concave in $m < 2$, increasing and convex in $m > 2$, and decreasing in k_D. In addition, it is increasing in k_C if $m > 2$, and is independent of k_C if $m < 2$.*

(vi) *$N(m, k_C) \geq H(m, k_C, k_D)$ for $m \leq k^2/(2k_C^2)$ and $N(m, k_C) \leq H(m, k_C, k_D)$ for $m \geq k^2/(2k_C^2)$.*

These properties are later used for figures concerning the optimal organization structure and for comparative statics.

Now we can provide the optimal organization structure for all values of the parameters.

THEOREM 3. *Suppose (A-1), (A-2), (A-3), and $\lambda \to +\infty$. Then the optimal organization structure is given as follows:*

(I) *When $m = 1$, the SD-horizontal coordination system is optimal for $h < \left(\sqrt{2}k_C\right)^{-1}$.*

(II) *When $m \geq k^2/k_C^2$, the hierarchical coordination system is optimal for $h < \left(\sqrt{2}k\right)^{-1}$.*

(III) *Suppose $m \in (1, k^2/k_C^2)$. Then*

 (i) *if $m \leq k^2/(2k_C^2)$ and $h \leq N(m, k_C)$, the N-horizontal coordination system is optimal;*

 (ii) *if $m \geq k^2/(2k_C^2)$ and $h \leq H(m, k_C, k_D)$, the hierarchical coordination system is optimal;*

 (iii) *if $m \leq 2$ and $\max[N(m, k_C), H(m, k_C, k_D)] \leq h \leq \left(\sqrt{2m}k_C\right)^{-1}$, the SD-horizontal coordination system is optimal;*

 (iv) *if $m \geq 2$ and $\max[N(m, k_C), H(m, k_C, k_D)] \leq h \leq \left(\sqrt{2m}k_C\right)^{-1}$, the SS-horizontal system is optimal.*

(IV) *For all m, if $h \geq \max\left[\left(\sqrt{2m}k_C\right)^{-1}, \left(\sqrt{2}k\right)^{-1}\right]$, no investment in the information structure is optimal.*

PROOF: Assertions (I) and (II) are by Proposition 2. Assertions in (III) are by Proposition 1, the definition of $N(\cdot)$ and $H(\cdot)$, and Lemma 5. For the last assertion (IV), note that $\left(\sqrt{2m}k_C\right)^{-1} > \left(\sqrt{2}k\right)^{-1}$ if and only if $m < k^2/k_C^2$ and both are equal for $m = k^2/k_C^2$. ∎

Figures 9–11 summarize the optimal coordination system for each $m > 1$ and h. Figure 9 is the case $k^2/(2k_C^2) < 1$. In this case, the decoding cost of top is so low compared to the capacity cost ($k_D < (\sqrt{2} - 1)k_C$) that the N-horizontal system is never optimal. Also since $k^2/k_C^2 < 2$, there is no region where the SS-horizontal system is optimal. Either the hierarchical system or the SD-horizontal system is optimal, depending on the parameters. The figure shows that as the intra-shop coordination is more important, the region where the SD-horizontal system is the best shrinks, while the region for the hierarchical system expands.

Figure 10 is the case that $1 < k^2/(2k_C^2) < 2$. Though the decoding cost of top is still lower than the capacity cost ($k_D < k_C$), the N-horizontal system is optimal for sufficiently low m and h. Also for relatively high m and h, the SS-horizontal system is optimal. The region where the hierarchical system is optimal shrinks both from below and from right as the decoding cost increases. If the decoding cost is higher than the capacity cost, Figure 10 changes to Figure 11, which is the case $k^2/(2k_C^2) > 2$. This condition is equivalent to $k_D > k_C$. Both regions for the N-horizontal system and the SS-horizontal system expand and the region for the hierarchical system continues to shrink as k_D increases.

The figures show that given a prior precision h, the optimal regions for the hierarchical system and the SS-horizontal system are located above the regions for the N-horizontal system and the SD-horizontal system, respectively. We have provided explanation for this after Proposition 1: The hierarchical system and the SS-horizontal system have coordination advantage for high values of m while the N-horizontal system and the SD-horizontal system utilize on-the-spot knowledge more effectively. Another observation is that given m, the hierarchical system and the N-horizontal system are better than the SS-horizontal system and the SD-horizontal system for smaller prior precisions. This is because the hierarchical system and the N-horizontal system can achieve their target levels independently of the prior precision. The target level under the SS-horizontal system or the SD-horizontal system is quite sensitive to the prior precision and increases drastically as h decreases, as is discussed in Section 5. The reason is that the SS-horizontal system completely ignores one aspect of the environment and the SD-horizontal system cannot coordinate between two shop managers at all. On the other hand, the SS-horizontal system and the SD-horizontal system have information cost advantage when the prior precision is relatively high. Under these systems, the information costs are of the form $2K_C p$ for the common information processing capacity p. Each shop manager can specialize in one of two information sources. However, when top invests in capacities under the hierarchical system, she also spends time for reports, which creates costs of delay. Thus, the information costs are $2K_C p + 2K_D r$. Under the N-horizontal system, each shop manager spends his time for both information sources and so the information costs are of the form $4K_C p$. Because of this, the SS-horizontal system and the SD-horizontal system enable top to start investing in information processing capacities for a higher prior precision level than the hierarchical system and the N-horizontal system, and they retain this cost advantage for relatively high prior precisions.

Finally, consider comparative statics with regard to the capacity cost parameter k_C. By Lemma 5 (v), increases in the capacity cost yield greater advantage to the hierarchical system. Since $k^2/(2k_C^2)$ is decreasing in k_C, the region of the N-horizontal system shrinks from above. Note that by Lemma 5 (iv), this region also shrinks from right. Not only the hierarchical system, but also the SD- and SS-horizontal systems can have greater relative advantage over the N-horizontal system as k_C increases: The marginal cost saving of the information processing capacity under the N-horizontal system is larger than any other system. This is because better capacities under the N-horizontal system do not hurt the intra-shop coordination so much as those under the other two horizontal systems and because the on-the-spot information, not available under the hierarchical system, is utilized in decision making. Similarly, compared with the hierarchical system, the SS-horizontal system depends more on the information processing capacity, so that when k_C increases, the region for the hierarchical system makes the region for the SS-horizontal system shrink from left. Interestingly, the boundary between the hierarchical system and the SD-horizontal system does not change with k_C. However, since no-investment region expands as k_C increases, the region for the SD- or SS-horizontal system shrinks from the right.

7. Implications and Extensions

3.7.1. As was discussed in Section 1, the original motivation of this study came from the comparison between Japanese systems and American systems. Typical American coordination systems are characterized by the hierarchical, authoritative coordination of the specialized tasks, which corresponds to the hierarchical coordination system in the model. On the other hand, the stylized Japanese system is characterized by (i) less emphasis on authority and more dependence on autonomous, horizontal coordination through sharing of values and information, and (ii) ambiguous and fluid job separations. The N-horizontal system (which has the nonspecialized task structure) in the model seems closest to this Japanese system since (ii) results in generalist-type workers who are skilled in a wide variety of jobs. Under the N-horizontal system, the top manager delegates coordination to subordinates who can share knowledge about various information sources since they are generalists.[20] Aoki (1988) asserts that the Japanese system performs better than the American system if the environment is "volatile." For example, Aoki explicitly characterizes the environment favoring the Japanese production system as lower gains from increased scale, short product life cycles, volatile demand shift from one variety of product to another, small batch sizes in production, and relatively greater need to shorten lead time from order to delivery.

The simple model presented here characterizes the environment by parameters m, h, k_C, and k_D. I call the environment more volatile the lower is h, the lower is m, and the higher is k_D: As h is lower, the environment is more uncertain and changeable. As m is lower, it is less important to coordinate actions between two subordinates in each shop and it is more important for each shop manager to respond flexibly to the changing environment utilizing their on-the-spot knowledge. Scale economies are not important in such a situation. Finally, in the volatile environment, quick response is more important. Thus, time is more valuable and it is likely that the decoding cost of top k_D is high. Though these parameters will be too simple to capture all aspects of "volatile" environments, I believe that they capture some elements of such environments.

Then the results derived from the model have some interesting implications for the comparison between the Japanese system and the American system. Generally speaking, as in Figures 9–11, the horizontal system is better than the hierarchical system for small m, which is consistent with the observations cited above. However, being horizontal is not sufficient: First, by Theorem 1 (ii), if coordination across shops is very important as in an assembly line, *we have to alter the specialized task structure under the horizontal system to the nonspecialized task structure*, and allow workers to affect the performance in each shop jointly. Second, even when the nonspecialized task structure is adopted, horizontal systems with *specialized* information processing capacities are inferior to the hierarchical system if h is small, that is, if the environment is very changeable. *We need not only the horizontal system but also generalists* to dominate the hierarchical system in a very volatile environment. The horizontal system enables the firm to exploit the on-the-spot information available only to subordinates. Even if m is small, however, coordination within each shop

[20] Knowledge sharing also occurs under the SS-horizontal system about limited aspects of the environment.

is still important, if the environment is very changeable. Specialized subordinates ignore this fact and pursue the minimization of the direct effect too much, so that from the organizational point of view, the performance is far from optimal. To avoid this problem, generalists are needed to improve the coordination between subordinates.

The effect of higher decoding costs has already been discussed after Theorem 3 in the last section. The increase in k_D expands the region that the N-horizontal system is optimal.

3.7.2. In our model, there are two kinds of specialization; specialization in tasks and specialization in information processing. Though the model does not assume any externality between these two such as learning by doing, we showed that in volatile environments, nonspecialization in both information processing and tasks is necessary to dominate the hierarchical system. However, as Koike (1984) argues, nonspecialization in tasks is likely to result in nonspecialized capacities through learning by doing. In our model, this can be formalized by assuming that nonspecialized task structure reduces the capacity costs. The introduction of such an externality is clearly favorable to the N-horizontal system.[21]

3.7.3. Based on several efficiency criteria, Williamson (1985) supports the authority relation mode in the production line with successive manufacturing stages. I do not intend to attack his argument since his objective is different from mine: He compares the economic merits of a simple hierarchy, which is the essential characteristic of the capitalist firm, with those of other primitive modes of ownership relations and contracting forms. The objective of this paper is to compare several organization structures of the capitalist firm. Neither ownership nor contracting is considered. Instead, the analysis is based solely on limited rationality such as communication costs and limited attention. However, it seems of value to give some comments on his arguments.

Williamson's authority relation mode has properties similar to the hierarchical coordination system: It pursues specialization of labor and adaptation to local and system shocks is conducted by the authority of the upper-level member. In my model, such a system is in fact optimal if uncertainty is relatively high (h is low) and greater need for intra-shop coordination exists (m is high), which corresponds to greater need for local shock responsiveness in Williamson's words. The model shows, however, that if there exist com-

[21] I am grateful to Masahiko Aoki for this observation.

munication limitations, local shock responsiveness of the authority relation will be reduced and the advantage of the specialization of labor will become ambiguous. Then, the capitalist firm may adopt an organizational form different from the hierarchical system, depending on which of intra-shop coordination and the use of the on-the-spot information is more important. Williamson considers the situation that the value of on-the-spot information is relatively low.

Another comment on Williamson's work concerns his argument of "selective intervention." Williamson himself modeled communication distortion problems in Williamson (1967) and concluded that the control loss by the top manager in the large organization limits the size of the optimal organization. Later in Williamson (1985), he criticizes this early work by saying that it does not permit selective intervention. He says, "Intervention at the top...always occurs *selectively*, which is to say only upon a showing of expected net gains (p.133, emphasis in the original)." My model shows, however, that the optimal information structure depends upon the adopted coordination system. Especially when the environment is changeable, the optimal capacities are either specialization or nonspecialization, depending on which of the hierarchical system or the N-horizontal system is the best. If the on-the-spot information is initially more valuable and the top manager decides not to intervene in the operating decision of subordinates, it is best to accumulate nonspecialized capacities in subordinates. If, however, the the intra-shop coordination becomes more important later, it is best for the top manager to accumulate specialized capacities in subordinates and to intervene. However, modifying the capacities once accumulated is costly, so that intervention may not be beneficial. Therefore, selective intervention may not be as easy as expected.

3.7.4. The analysis was conducted under Assumption (A-2) which says that there is no communication between subordinates under the horizontal system. If this assumption is relaxed, the performance of the horizontal system will improve, so that the region where the hierarchical system is superior will shrink. The more important effect is that the interchange of information may favor the SD-horizontal system more than the N-horizontal system and the SS-horizontal system because the interchange of information improves the coordination between subordinates under the SD-horizontal system while keeping the efficient use of the on-the-spot information by specialization. Therefore it is expected that the region where the N-horizontal system is the best may expand upward, but may shrink from the right. On the other hand, the marginal decoding costs of shop managers depend on the information processing capacities they have. For example, if $\mathcal{K}(0)$ is extremely high, so that each shop manager cannot understand a report concerning some information source, say X, without the knowledge capacity for X, then some degree of nonspecialization will be necessary. Therefore the effect of allowing horizontal communication depends on the balance of these two opposite effects. If specialization in different information sources does not hurt the ability to communicate very much, the advantage of generalists will be reduced. For example, in a "kanban" system, clearly the horizontal communication by "kanban" is important while job rotation among various shop floors is also common. The information to each shop conveyed by "kanban" is generally simple (the amount and timing of delivery of each type of parts produced in that shop), and so understanding the information will not take as much time as the center understanding and processing the information from all shops. Therefore in every day manufacturing and delivery, being generalists may not be valuable. However, if some emergent events like machine breakdown happen, nonspecialized workers can jointly detect the source of breakdown while coordinating their actions. Hence generalists are of value in this respect.

Another argument supporting the N-horizontal system is the number of communication channels needed. For example, when there are four subordinates, each subordinate needs to exchange and process information from the other three subordinates. Thus, the horizontal system will need twelve information channels, while the hierarchical system will need only four channels.[22] Therefore sometimes the implicit understanding of other subordinates by nonspecialized and homogeneous capacities may be better than the horizontal system with explicit communication and may characterize some aspects of Japanese firms.

3.7.5. One might ask why top herself does not use her time to collect information concerning the state variables in order to utilize the on-the-spot information or to reduce the decoding cost. Although I have not been explicit, I assume in the model that top has to attend to some variables concerning strategic decisions while subordinates are collecting

[22] In the "kanban" system, by virtue of the tree structure of the automobile production line (each shop supplying its parts to only one subsequent shop), it will need six channels when there are four shops.

and processing information concerning X and Y. Allocating her scarce time to information sources relevant to operating decisions such as X and Y is assumed to be prohibitively costly because it reduces the quality of strategic decisions drastically, so by ignoring this possibility, I compared two coordination systems based on the performance of operating decisions. If information activities relevant to strategic decisions are introduced into the model, there may be the possibility that top allocates some time to the information sources concerning operating decisions, at the cost of lower quality in strategic decisions.

3.7.6. In the model introduced, the source of the cost of the hierarchical system is that top has limited attention and that spending more time reading reports creates more costs of delay in her decision making. In the model presented, each subordinate sends what he observes. However, he may require to write some *summary* reports for the top manager. This idea can be formalized by assuming that communication channels have limited dimensionality; each communication channel can send only a one-dimensional value of variables. Under this assumption, each subordinate i is assumed to send a convex combination of his information $d_i X_i + (1 - d_i) Y_i$ to either top or the other subordinate, where $d_i \in [0, 1]$ is the choice variable of subordinate i.

Introducing this dimension restriction assumption into the model analyzed does not alter our analysis at all because specialization is optimal under the hierarchical system. When subordinates are specialized, limited dimensionality restricts nothing because both X and Y are one-dimensional. Therefore, under limited dimensionality, the specialization in different information sources is still optimal under the hierarchical system.

This limited dimensionality plays an important role if the information sources X and Y are multidimensional. Suppose that the top manager has unlimited attention ($K_D = 0$), and that each multidimensional information source is inseparable in the sense that each element of the multidimensional random variable cannot be observed separately: By spending time t_i on $X = (X^1, \ldots, X^n)$, subordinate i is assumed to observe $X_i = (X_i^1, \ldots, X_i^n)$ where $X_i^j = X^j + \epsilon_i^j$ with ϵ_i^j Normally and independently distributed with mean zero and precision $p_i^j t_i$. However, each subordinate can send only a one-dimensional signal through a communication channel. Then, without decoding costs, this creates a cost in the hierarchical system, and the on-the-spot information is of value as in the model presented in Section 2. This model with multidimensional information sources and limited dimensionality is much more complicated, so that the analysis is not easy. However, I conjecture that qualitative results will be the same. In addition, we can predict that when the dimension of the information sources increases, the region where the hierarchical system is optimal will shrink. This higher dimension can be interpreted as another measure of volatile environments.

3.7.7. The analysis in this paper was based on team theory and we completely ignored incentive problems. One might assert that since incentive aspects are ignored, the analysis would have given too much advantage on the horizontal system over the hierarchical system. However, the nonspecialization in information sources under the horizontal system can reduce some cost due to incentive problems. If the contribution of each subordinate is public information, nonspecialization will make the contribution of one subordinate more correlated with that of another, so that relative performance evaluation may reduce costs due to incentive problems. (See Holmström (1982) for detailed arguments.) The nonspecialized capacities under the horizontal system may work as a device for reducing opportunistic behavior by subordinates.

More importantly, the hierarchical system itself may create some additional costs under incentive-based models. When the hierarchical system is adopted, subordinates may distort their information in their favor as analyzed in Tirole (1986), or subordinates may spend too much time attempting to influence top's decisions to their advantage and attend to their local management less intensively than optimal from top's point of view, as is discussed in Milgrom (1988) and Milgrom and Roberts (1987). These authors conclude from those sources of the costs of the hierarchical system, that some degree of decentralization is desirable.

8. Concluding Remarks

This paper analyzed the hierarchical coordination system and the horizontal coordination system by a simple model based on limited rationality. We derived the optimal information structure and the task structure under each coordination system and the optimal organization structure of the firm by comparing them. The firm accumulates different kinds of information processing capacities under different coordination systems, depending on characteristics of the outside environment such as the prior precision of the environmental variable and the importance of two kinds of coordination effects. The main results are summarized as follows:

1. When coordination across shops (or industries into which the firm diversifies) is unnecessary, the optimal organization structure is the horizontal coordination system with the specialized task structure and the information structure such that subordinates specialize in different information sources. This corresponds to the M-form organization.

2. However, when inter-shop coordination is extremely important (or the firm operates in two very related industries), this M-form structure is never optimal. Intervention by the upper-level manager is better in information utilization and coordination. To utilize on-the-spot information, we require the nonspecialized task structure to make subordinates work jointly in each shop.

3. When the environment is volatile (so that the prior variance is high, on-the-spot information is valuable, and the communication is costly), the optimal organization structure requires the horizontal coordination system with nonspecialization in information sources as well as the nonspecialized task structure.

One drawback of the model is that it is so abstract and stylized that the interpretation of the model (for example, the payoff function of the firm) is not clear. The next step would be to apply the approach developed in this paper to more specific situations such as coordination problems in assembly lines or relationship between organization structures and diversification strategies of the firm.

Finally, I mention some related existing literature analyzing the information structure of the organization. There is no literature analyzing information processing capacities directly. However, some literature provides models of the organization where capacities of lower-level members are indirectly determined by the higher-level managers' decisions concerning organization. Geanakoplos and Milgrom (1985) formalize a model of hierarchies with multiple managers. Their general model is similar to mine in that, since attention is a divisible scarce resource and there are several information sources to be attended to, each member in the organization decides how to allocate his or her attention to various information sources. The top manager can control members' allocation decision by specifying decision times. However, members' information processing capacities are not under the top's control. Also their concern is in analyzing hierarchy; they do not provide any alternative organizational structure.

Aoki (1986) analyzes two kinds of information structure, called the vertical and the horizontal structure. His model is more dynamic. The cost of the vertical structure (which is similar to the hierarchical system in my model) comes from both top manager's imperfect observation of emerging events and *time delay* in implementing actions. On the other hand, when the horizontal structure is chosen, shop managers respond to local shocks repeatedly, which encourages them to improve their information processing capacities by *learning by doing*. Then they implement actions without delay, but imperfectly in terms of horizontal coordination. My model is static, but emphasizes the relation between the coordination system and the optimal information processing capacities such as specialists vs. generalists, which is not analyzed in his model.

Green and Laffont (1986) is similar in spirit to our model. They compare two communication systems, centralization and interchange of information, both of which install two communication channels. There are two agents who attend to two information sources to obtain some noisy signals, and each of which chooses action a_i and b_i to aim at minimizing a loss function similar to ours. There are several differences between their model and mine. First, they do not consider the choice of information processing capacities. The precisions of noise terms are exogenously given. Second, they focus purely on alternative *communication* systems: There are only two agents, and even under centralization, each agent chooses its actions, different from the hierarchical system in my model, where the third agent, called top, chooses all the actions. The centralization in their model means that all information is centralized in one agent. Third, communication is limited not because receivers have limited attention, but because channels have limited dimensionality as discussed in the last section. Each agent sends a convex combination of his signals. However, they only consider the case that the prior variances of the state variables are infinite, so that if ever communication occurs, mixed signals are not used. This simplifies their analysis greatly, though they cannot examine the effect of the prior variances on the optimal structure, which is the main concern in this paper. Their concern is in the effect of the exogenously given precisions of the noise terms on the optimal structure. For example, in the limiting case that $\mu \to \infty$ in the loss function, they show that the most accurate observer of each state variable should transmit his observation to the other agent. Their analysis seems to be relevant to an investigation of the conditions under which hierarchical relations in the sense of information concentration appear as an optimal structure. The analysis in this paper concerns the optimal organization structure of the already hierarchical firm.

References

AOKI, M. (1986). Horizontal vs. Vertical Information Structure of the Firm, *American Economic Review* **76**, pp. 971–983.

———(1988). "A Microtheory of the Japanese Economy: Information, Incentives, and Games," Cambridge: Cambridge University Press.

ARROW, K. J. (1974). "The Limits of Organization," New York: W. W. Norton.

———(1985). Informational Structure of the Firm, *American Economic Review* **75**, pp. 303–307.

GEANAKOPLOS, J. AND P. MILGROM (1985). A Theory of Hierarchies Based on Limited Managerial Attention, Cowles Foundation Discussion Paper No. 775.

GREEN, J. R. AND J.-J. LAFFONT (1986). Alternative Limited Communication Systems: Centralization versus Interchange of Information, *in* "Uncertainty, Information, and Communication: Essays in Honor of Kenneth J. Arrow, Volume III," (Walter P. Heller, Ross M. Starr, and David A. Starrett, Ed.), Cambridge: Cambridge University Press.

HOLMSTRÖM, B. (1982). Moral Hazard in Teams, *Bell Journal of Economics* **13**, pp. 324–340.

HOLMSTRÖM, B. R. AND J. TIROLE (1987). The Theory of the Firm, *forthcoming in* "Handbook of Industrial Organization," (Richard Schmalensee and Robert Willig, Ed).

ITOH, H. (1987). Information Processing Capacities of the Firm, *Journal of the Japanese and International Economies* **1**, pp. 299–326.

KAGONO, T., I. NONAKA, K. SAKAKIBARA AND A. OKUMURA (1985). "Strategic vs. Evolutionary Management: A U.S.-Japan Comparison of Strategy and Organization," Amsterdam: North-Holland.

KOIKE, K. (1977). "*Shokuba no Rodokumiai to Sanka: Roshi Kankei no Nichibei Hikaku* (A Comparative Study of Industrial Relations on the Shop Floor between the U.S. and Japan)," Tokyo: Toyo Keizai Shinpo Sha.

———(1978). Japan's Industrial Relations: Characteristics and Problems, *Japanese Economic Studies* **7**, pp. 42–90.

———(1984). Skill Formation Systems in the U.S. and Japan: A Comparative Study, *in* "The Economic Analysis of the Japanese Firm," (Masahiko Aoki, Ed.), Amsterdam: North-Holland.

LINCOLN, J. R. AND K. MCBRIDE (1987). Japanese Industrial Organization in Comparative Perspective, *Annual Review of Sociology* **13**, pp. 289–312.

MARSCHAK, J. (1974). "Economic Information, Decision, and Prediction: Selected Essays, Volume II," Dordrecht-Holland: D. Reidel Publishing Company.

MARSCHAK, J. AND R. RADNER (1972). "Economic Theory of Teams," New Haven: Yale University Press.

MILGROM, P. R. (1988). Employment Contracts, Influence Activities and Efficient Organization Design, *Journal of Political Economy* **96**, pp. 42–60.

MILGROM, P. AND J. ROBERTS (1987). Bargaining and Influence Costs and the Organization of Economic Activity, Research Paper 934, Graduate School of Business, Stanford University.

SIMON, H. A. (1974). Applying Information Technology to Organization Design, *Public Administration Review* **33**, pp. 268–278.

TIROLE, J. (1986). Hierarchies and Bureaucracies: On the Role of Collusion in Organizations, *Journal of Law, Economics and Organization* **2**, pp. 181–214.

WILLIAMSON, O. E. (1967). Hierarchical Control and Optimum Firm Size, *Journal of Political Economy* **75**, pp. 123–138.

————(1975). "Markets and Hierarchies: Analysis and Antitrust Implication," New York: Free Press.

————(1985). "The Economic Institutions of Capitalism," New York: Free Press.

Appendix

Proof of Lemma 1.

We first prove this lemma under the assumption that decoding costs are zero. Suppose that subordinate 1 specializes in X and subordinate 2 in Y. Communication is of no value since the information subordinate 2 has does not improve the decision making by subordinate 1 (in shop A) at all, and vice versa. Thus, given such an informational structure, the expected gross loss is of the form

$$L(\mathcal{I}) = \mathrm{Var}(X \mid X_1) + \mathrm{Var}(Y \mid Y_2) = (h_X + p_1)^{-1} + (h_Y + q_2)^{-1}.$$

Top chooses p_1 and q_2 to minimize $L(\mathcal{I}) + K_C p_1 + K_C q_2$. The optimal solution is given by $p_1 = \max\big[(K_C)^{-1/2} - h_X, 0\big]$ and $q_2 = \max\big[(K_C)^{-1/2} - h_Y, 0\big]$ and the optimal value of the expected gross loss is

$$L = \min\big[\sqrt{K_C}, h_X^{-1}\big] + \min\big[\sqrt{K_C}, h_Y^{-1}\big].$$

Let $\mathcal{L}$ be the optimal value of the expected net loss (the sum of L and the information costs). Then

$$\mathcal{L} = \min\big[\sqrt{K_C}(2 - \sqrt{K_C}h_X), h_X^{-1}\big] + \min\big[\sqrt{K_C}(2 - \sqrt{K_C}h_Y), h_Y^{-1}\big]. \qquad \text{(A1)}$$

Next consider the specialization in the same information source. Suppose, without loss of generality, that both managers attend to X. Clearly, communication is of no value. Then

$$L(\mathcal{I}) = \mathrm{Var}(X \mid X_1, X_2) + \mathrm{Var}(Y) = (h_X + p_1 + p_2)^{-1} + h_Y^{-1}.$$

Top chooses p_1 and p_2 to minimize $L(I) + K_C p_1 + K_C p_2$. The optimal solution satisfies $p_1 + p_2 = \max\big[(K_C)^{-1/2} - h_X\big]$ and

$$\begin{aligned} L &= \min\big[\sqrt{K_C}, h_X^{-1}\big] + h_Y^{-1} \\ \mathcal{L} &= \min\big[\sqrt{K_C}(2 - \sqrt{K_C}h_X), h_X^{-1}\big] + h_Y^{-1}. \end{aligned} \qquad \text{(A2)}$$

It is clear by comparing (A1) and (A2) that the former is better.

Finally consider the case that both managers are generalists, that is, $p_i > 0$, $q_i > 0$, and $t_i \in (0,1)$. Then since decoding costs are zero, the expected gross loss is of the form

$$\begin{aligned} L(\mathcal{I}) &= \mathrm{Var}(X \mid X_1, X_2) + \mathrm{Var}(Y \mid Y_1, Y_2) \\ &= (h_X + p_1 t_1 + p_2 t_2)^{-1} + (h_Y + q_1(1 - t_1) + q_2(1 - t_2))^{-1}. \end{aligned}$$

Top tries to find p_1, p_2, q_1, q_2, t_1, and t_2 that minimize $L(\mathcal{I}) + \sum_{i=1}^{2} K_C(p_i + q_i)$. The first-order conditions are

$$
\begin{aligned}
(h_X + p_1 t_1 + p_2 t_2)^{-2} t_i &= K_C \\
(h_Y + q_1(1-t_1) + q_2(1-t_2))^{-2}(1-t_i) &= K_C
\end{aligned}
\tag{A3}
$$

for $i = 1, 2$. Hence $t_1 = t_2 = t$. Given $t \in (0,1)$ as a parameter, we obtain from (A3),

$$
\begin{aligned}
p_1 + p_2 &= \max\big[(tK_C)^{-1/2} - h_X/t, 0\big] \\
q_1 + q_2 &= \max\big[((1-t)K_C)^{-1/2} - h_Y/(1-t), 0\big]
\end{aligned}
$$

and

$$
\begin{aligned}
L &= \min\left[\sqrt{\frac{K_C}{t}}, h_X^{-1}\right] + \min\left[\sqrt{\frac{K_C}{1-t}}, h_Y^{-1}\right] \\
\mathcal{L} &= \min\left[\sqrt{\frac{K_C}{t}}\left(2 - \sqrt{\frac{K_C}{t}} h_X\right), h_X^{-1}\right] + \min\left[\sqrt{\frac{K_C}{1-t}}\left(2 - \sqrt{\frac{K_C}{1-t}} h_Y\right), h_Y^{-1}\right].
\end{aligned}
\tag{A4}
$$

To compare (A4) with (A1), we show that for all $t \in (0,1)$, the first term in $\mathcal{L}$ in (A4) is at least as large as the first term in (A1), which is sufficient. First note that $(K_C/t)^{-1/2} < K_C^{-1/2}$ for $t \in (0,1)$. Thus if $h \geq K_C^{-1/2}$, both are equal to h_X^{-1}. If $(K_C/t)^{-1/2} \leq h < K_C^{-1/2}$, clearly the first term in (A1) is smaller. Finally, for $h < (K_C/t)^{-1/2}$,

$$
\frac{\partial}{\partial t}\left[\sqrt{\frac{K_C}{t}}\left(2 - \sqrt{\frac{K_C}{t}} h_X\right)\right] = 2\sqrt{K_C}\left(1 - \sqrt{\frac{K_C}{t}} h_X\right)\frac{d}{dt}\left(\frac{1}{\sqrt{t}}\right) < 0.
$$

Thus, the first term in (A1) is smaller. Therefore, for all $t \in (0,1)$, specialization in different information sources are better.

Now suppose that decoding costs are positive. Since we have shown that the optimal information structure under the assumption of zero decoding costs has no communication, the result does not change: specialization in different information sources is optimal.

Proof of Lemma 4.

The first-order necessary conditions for $p_i > 0$, $q_i > 0$, $r_i > 0$, $s_i > 0$, and $t_i \in (0,1)$ are given as follows:

$$
A^{-2}\frac{t_i r_i^2}{(p_i t_i + r_i)^2} = 2K_C; \tag{A5p}
$$

$$
B^{-2}\frac{(1-t_i)s_i^2}{(q_i(1-t_i) + s_i)^2} = 2K_C; \tag{A5q}
$$

$$A^{-2}\left(\frac{p_it_i}{p_it_i+r_i}\right)^2 = 2K_D; \tag{A5r}$$

$$B^{-2}\left(\frac{q_i(1-t_i)}{q_i(1-t_i)+s_i}\right)^2 = 2K_D; \tag{A5s}$$

$$A^{-2}\frac{p_ir_i^2}{(p_it_i+r_i)^2} = B^{-2}\frac{q_is_i^2}{(q_i(1-t_i)+s_i)^2} \tag{A5t}$$

where

$$A = h + \frac{p_1t_1r_1}{p_1t_1+r_1} + \frac{p_2t_2r_2}{p_2t_2+r_2}$$

$$B = h + \frac{q_1(1-t_1)s_1}{q_1(1-t_1)+s_1} + \frac{q_2(1-t_2)s_2}{q_2(1-t_2)+s_2}.$$

From (A5p), (A5q), and (A5t), we obtain

$$t_i = \frac{p_i}{p_i+q_i}. \tag{A6}$$

From (A5r) and (A5s),

$$A^{-1} = \sqrt{2}k_D\left(1+\frac{r_i}{p_it_i}\right)$$

$$B^{-1} = \sqrt{2}k_D\left(1+\frac{s_i}{q_i(1-t_i)}\right). \tag{A7}$$

Since t_i minimizes $A^{-1}+B^{-1}$, the first-order necessary condition is

$$-\frac{p_ir_i}{(p_it_i)^2} + \frac{q_is_i}{(q_i(1-t_i))^2} = 0. \tag{A8}$$

Substituting (A6) into (A8) yields

$$\frac{r_i}{p_i^3} = \frac{s_i}{q_i^3} \tag{A9}$$

Also substituting (A6) and (A7) into (A5t) yields

$$\frac{r_i^2}{p_i^3} = \frac{s_i^2}{q_i^3}. \tag{A10}$$

From (A9) and (A10), we obtain $r_i = s_i$, so that $p_i = q_i$ and $t_i = 1/2$.

The Derivation of $N(\cdot)$ and $H(\cdot)$

The derivation is straightforward. First consider $N^{SD}(\cdot)$. Since

$$\left(2\sqrt{m}k_C\right)^{-1} < \left(\sqrt{2m}k_C\right)^{-1}, \tag{A11}$$

by solving $S^N = S^{SD}$ for $h < \left(2\sqrt{m}k_C\right)^{-1}$, we obtain

$$N^{SD}(m,k_C) = \frac{\sqrt{m}-1}{\sqrt{2m}\left(\sqrt{2m}-1\right)k_C}. \tag{A12}$$

Since $N^{SD} < \left(2\sqrt{m}k_C\right)^{-1}$ by (A11), this is well defined. Similarly, solving $S^N = S^{SS}$ for $h < \left(2\sqrt{m}k_C\right)^{-1}$ yields

$$N^{SS}(m, k_C) = \frac{\sqrt{2}-1}{\sqrt{2m}k_C}. \tag{A13}$$

Again, since $N^{SS} < \left(2\sqrt{m}k_C\right)^{-1}$ by (A11), this is well defined.

Next, consider $H^{SD}(\cdot)$. First note that

$$\left(\sqrt{2}k\right)^{-1} < \left(\sqrt{2m}k_C\right)^{-1} \qquad \text{if and only if} \qquad m < k^2/k_C^2. \tag{A14}$$

Suppose h is smaller than both critical levels in (A14). Then by solving $S^H = S^{SD}$, we obtain

$$H^{SD}(m, k_C, k_D) = \frac{\sqrt{m}-1}{\sqrt{2m}k_D}. \tag{A15}$$

However, we can show

$$H^{SD} < \left(\sqrt{2}k\right)^{-1} \qquad \text{if and only if} \qquad m < k^2/k_C^2. \tag{A16}$$

Therefore for $m \leq k^2/k_C^2$, (A15) is valid. For $m \geq k^2/k_C^2$, we have

$$H^{SD}(m, k_C, k_D) = \frac{1}{\sqrt{2}k}. \tag{A17}$$

The similar procedure yields

$$H^{SS}(m, k_C, k_D) = \frac{\sqrt{2}-1}{2k - \sqrt{2m}k_C} \tag{A18}$$

for $m \leq k^2/k_C^2$. When $m \geq k^2/k_C^2$, $H^{SS}(\cdot)$ is equal to $H^{SD}(\cdot)$ in (A17). The functions $N(\cdot)$ and $H(\cdot)$ are derived from these directly.

The proof of Lemma 5 follows from $N(\cdot)$ and $H(\cdot)$ derived above. Concerning (i) and (ii), note that

$$N^{SD}(2, k_C) = N^{SS}(2, k_C) = \left(\sqrt{2}-1\right)/\left(2k_C\right),$$
$$H^{SD}(2, k_C, k_D) = H^{SS}(2, k_C, k_D) = \left(\sqrt{2}-1\right)/\left(2k_D\right),$$

and

$$N(m, k_C) = H(m, k_C, k_D) = \left(k - \sqrt{2}k_C\right)/\left(\sqrt{2}kk_D\right) \qquad \text{for} \qquad m = k^2/(2k_C^2).$$

Assertions (iii) and (iv) can be obtained by differentiating them. Finally, (v) holds because of Proposition 1 (i).

FIGURE 1. The hierarchical coordination system.

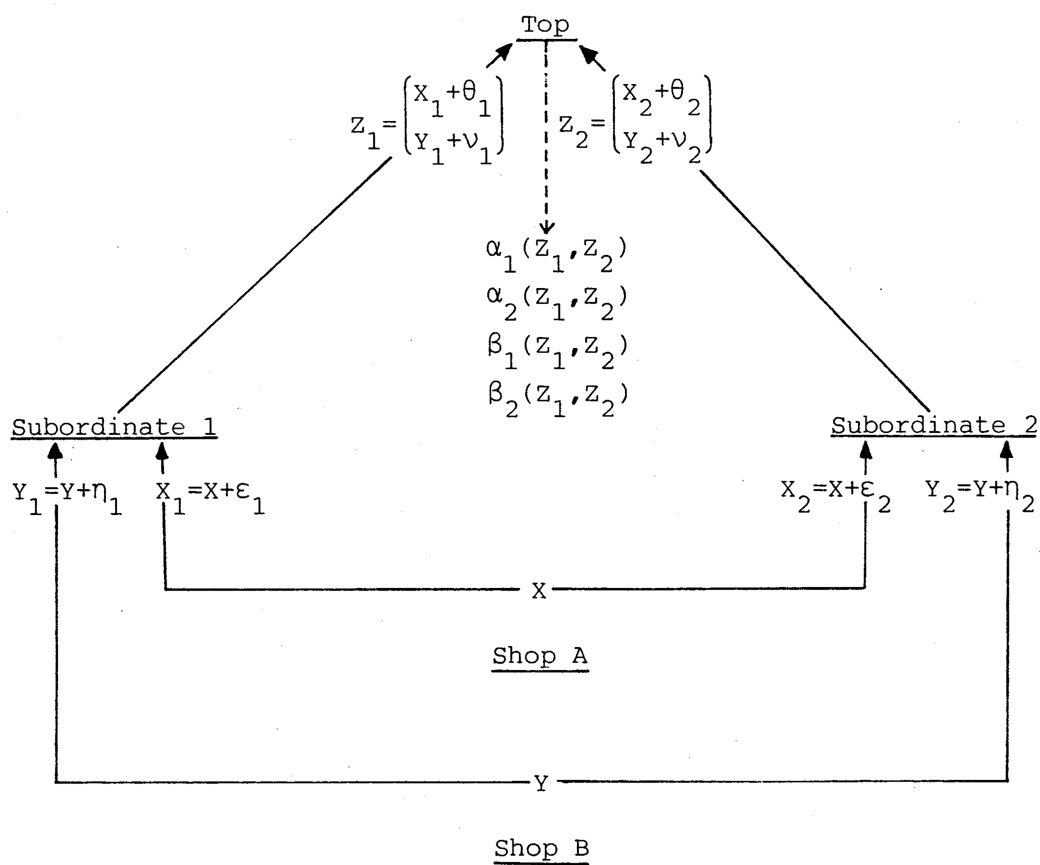

FIGURE 2a. The horizontal coordination system with nonspecialized task structure.

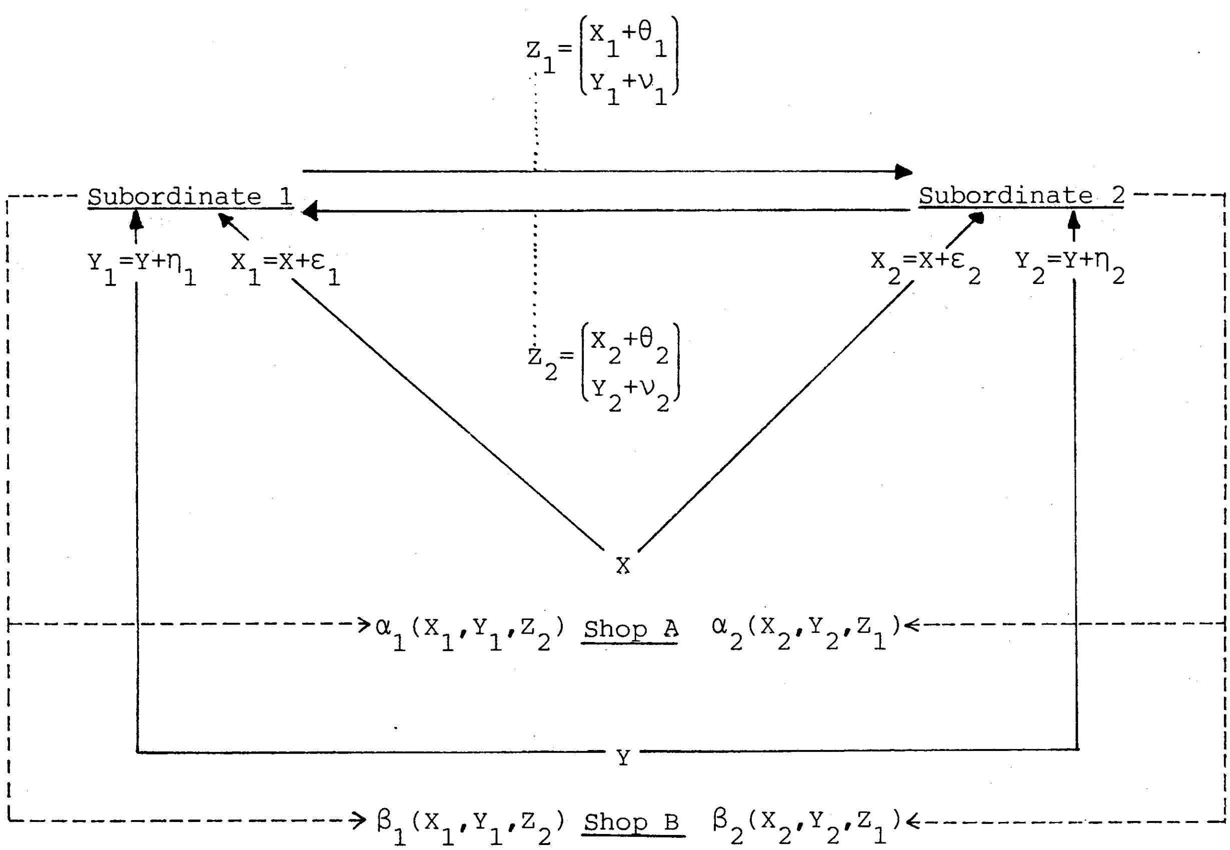

FIGURE 2b. The horizontal coordination system with specialized task structure.

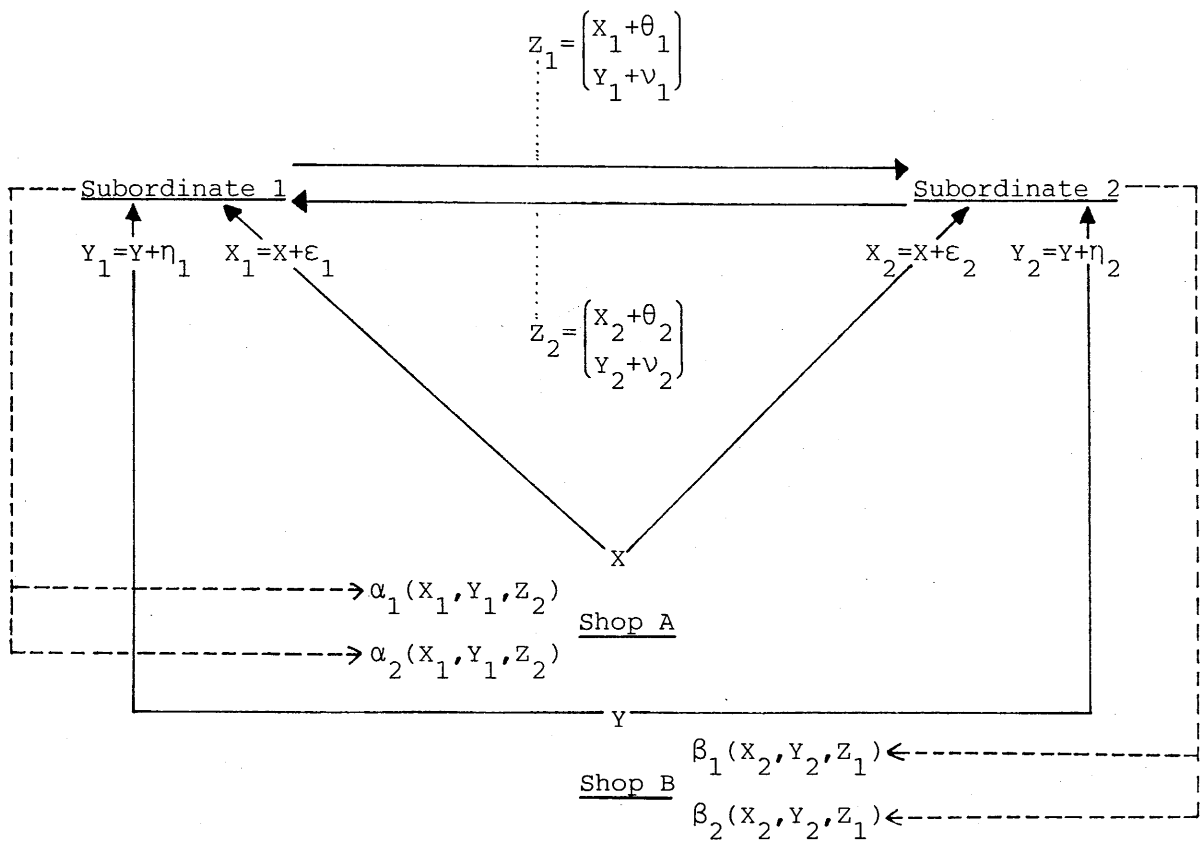

FIGURE 3. The optimal cost saving of the hierarchical system.

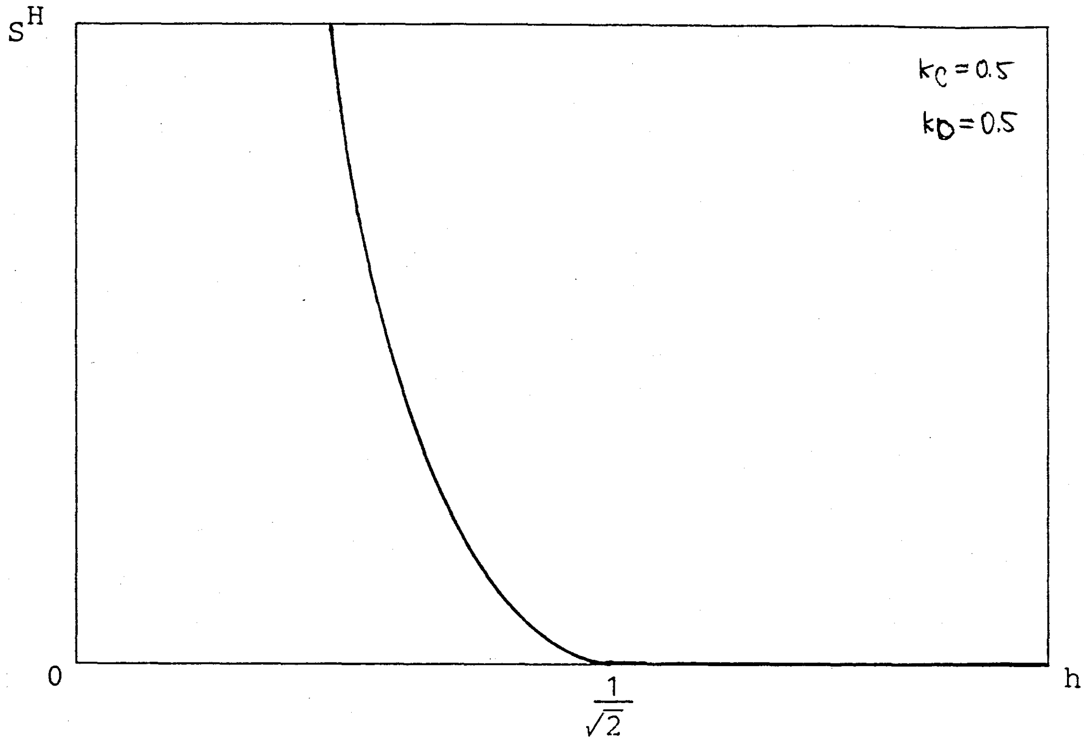

FIGURE 4. The optimal cost saving of the N-horizontal system.

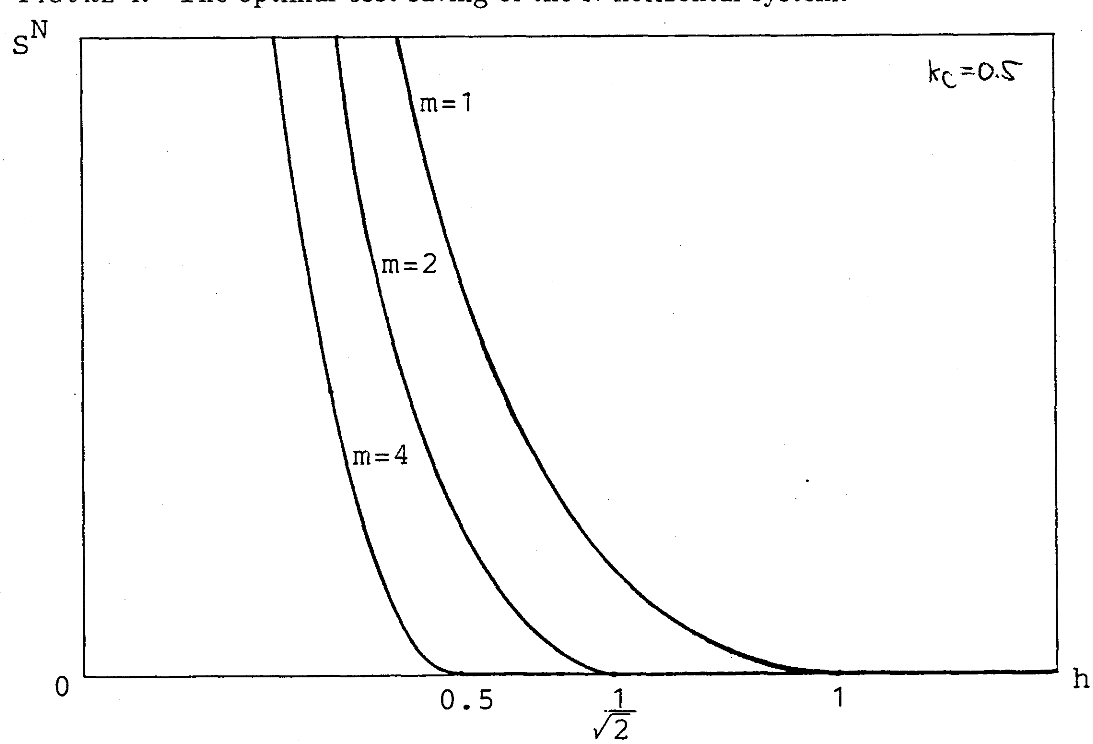

FIGURE 5. The optimal cost saving of the SD-horizontal system.

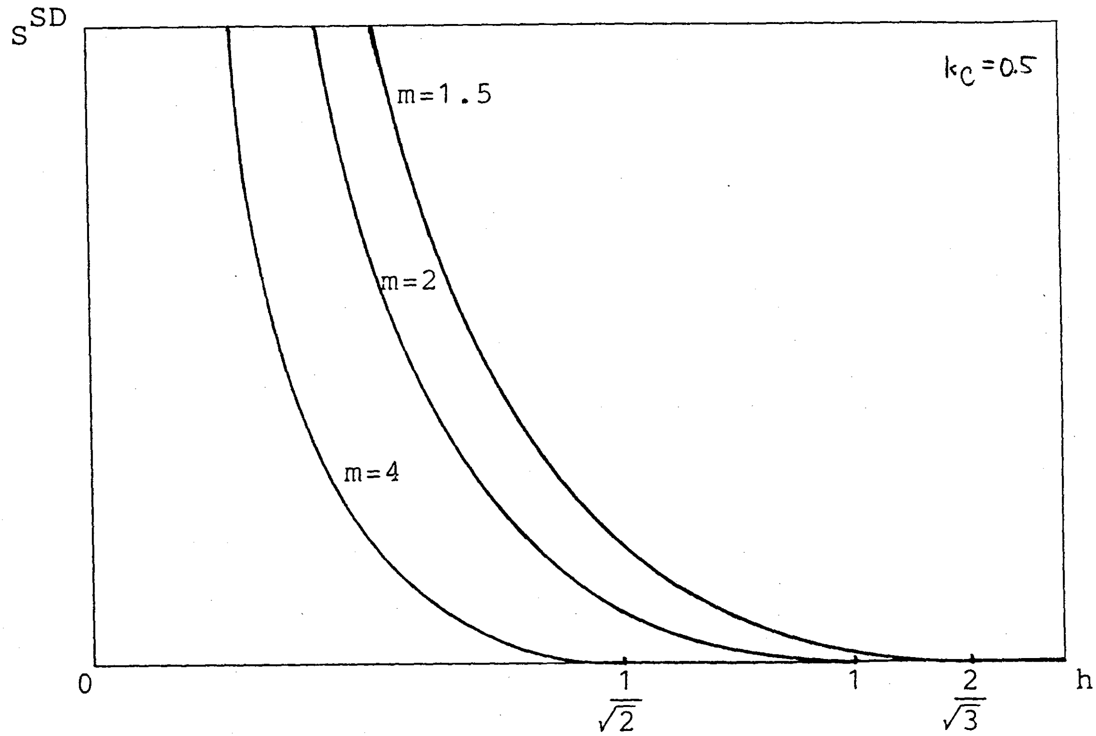

FIGURE 6. The optimal cost saving of the SS-horizontal system.

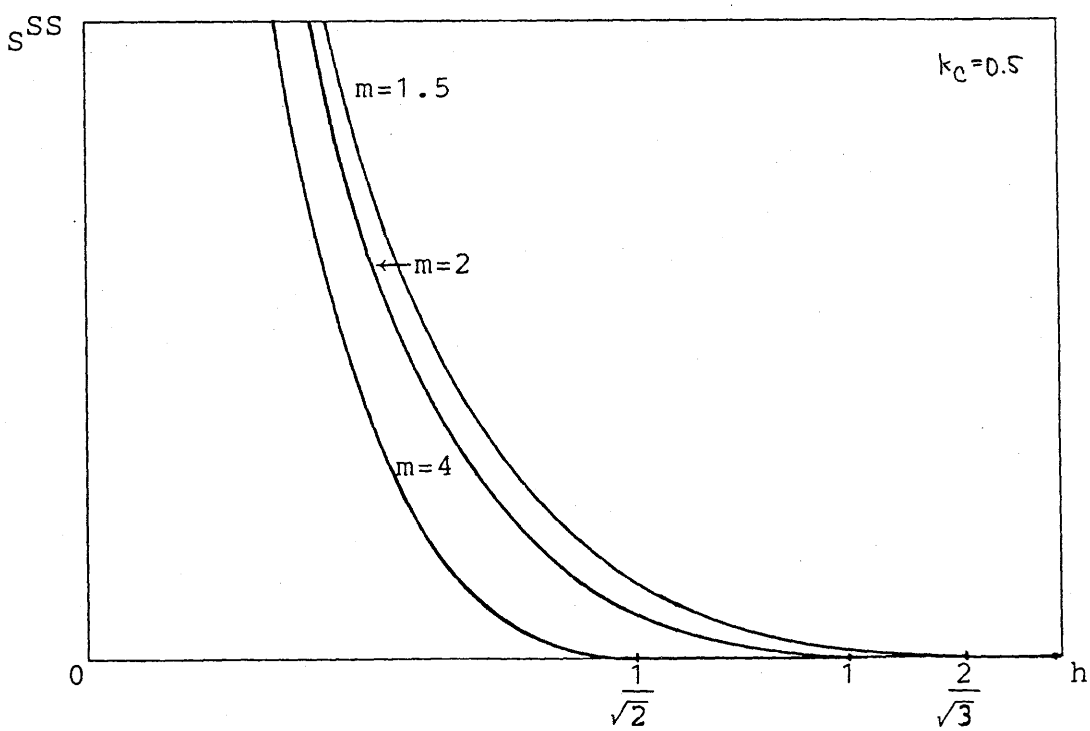

FIGURE 7. The determination of $N(\cdot)$.

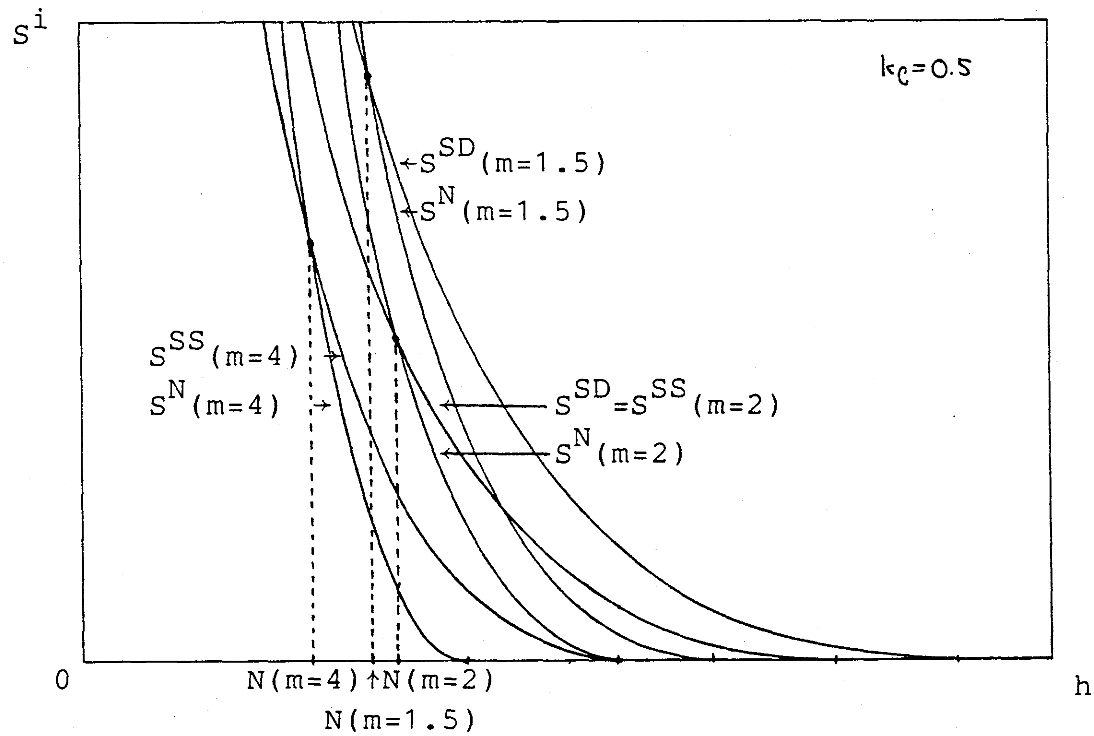

FIGURE 8. The determination of $H(\cdot)$.

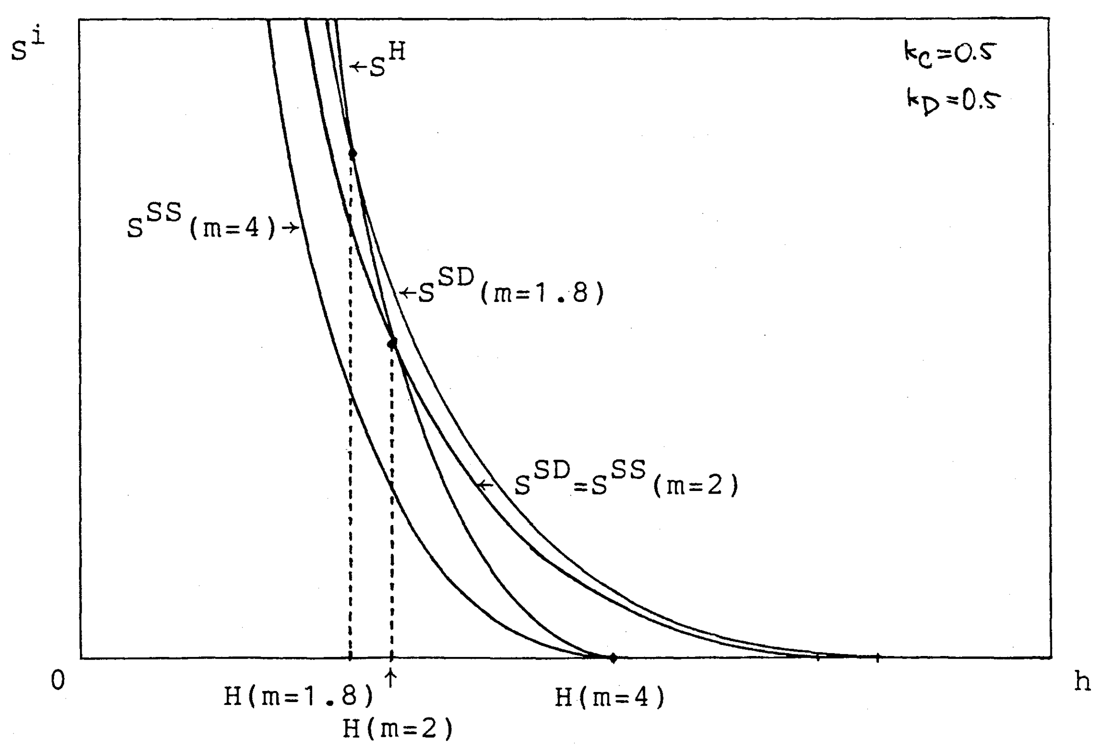

FIGURE 9. The optimal coordination system. (The case $k^2/(2k_C^2) < 1$.)

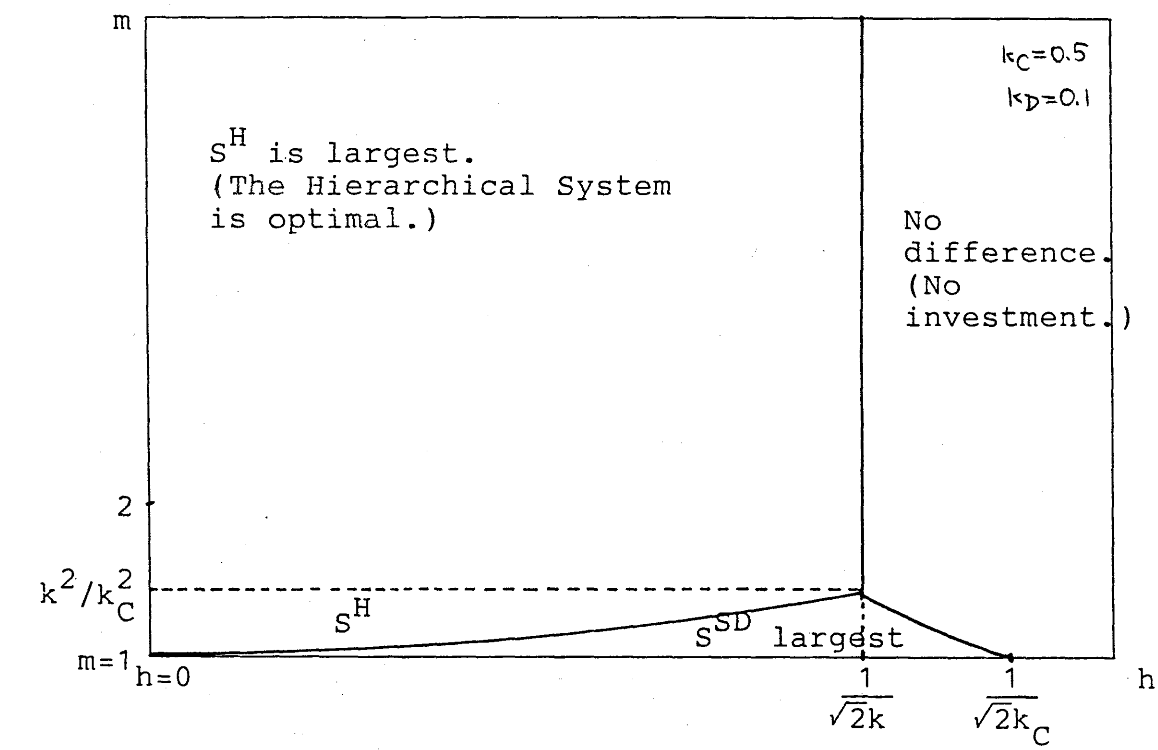

FIGURE 10. The optimal coordination system. (The case $1 < k^2/(2k_C^2) < 2$.)

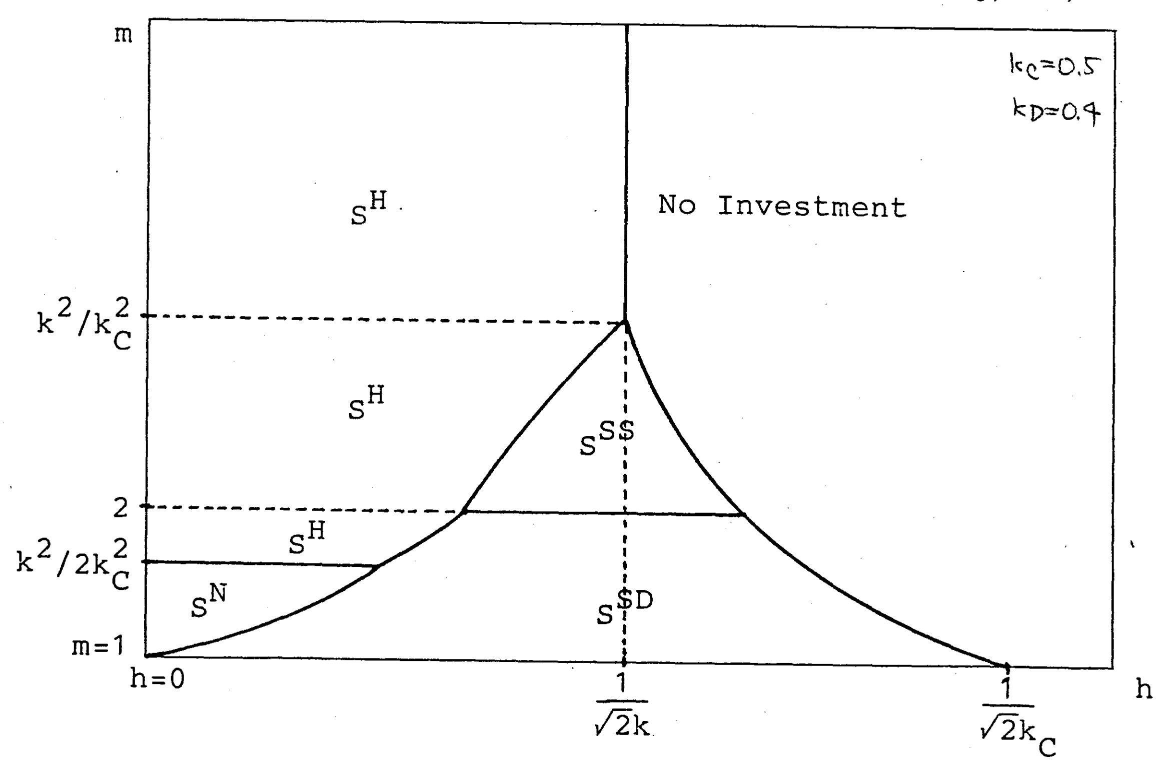

FIGURE 11. The optimal coordination system. (The case $k^2/(2k_C^2) > 2$)