

Generalists versus Specialists in Organizations¹

Daniel Ferreira²

Raaj K. Sah³

July 3, 2001

Working Draft - please do not circulate or quote

¹We thank Pierre-André Chiappori, Luis Garicano and Canice Prendergast for their comments.

²Department of Economics, The University of Chicago.

³Harris Graduate School of Public Policy Studies, The University of Chicago.

Abstract

In this paper, we discuss the trade-off between specialization and coordination in an organizational design problem. Most papers on the assignment of heterogeneous managers to different hierarchic levels emphasize the role of talent: better managers should be on top of hierarchies. However, this requires talent to be measured on an one-dimensional scale. In this paper, we explore the implications of allowing talent to have two dimensions: breadth and depth. Specialists have deep knowledge of few areas while generalists have narrow knowledge of many areas. When perfect communication is impossible, hierarchies arise in which generalists are at the top and specialists are at the bottom. We propose a model of imperfect communication and discuss its implications for organizational design, the optimal degree of centralization and the depth of hierarchies. We show that our model also implies plausible organizational structures, like balanced hierarchies and pyramidal structures.

1 Introduction

It is widely believed that coordination limits specialization. However, this fact has not been fully explored in the economic literature. One exception is Becker and Murphy's (1992) paper. They view coordination costs as a much more important force limiting specialization than Adam Smith's extent of the market. In their model, coordination costs increase with specialization, eventually making it uneconomical. But why is it so? What is so special about specialization that makes coordination so difficult? This question cannot be answered with Becker and Murphy's framework. In their model, coordination costs are a black box, much like neoclassical firms. In this paper, we propose to open this box and analyze coordination costs in detail.

Why do we want to open this black box? We believe that coordination costs are at the heart of recent changes in the specialization of work and in the use of managers. For example, despite the secular increase in the division of labor, the recent increase in the use of job rotation and work teams is consistent with a decrease in specialization (see Möbius, 1999). If coordination costs are driving the recent changes in the degree of specialization, why are coordination costs increasing? Another example is the increased use of managers. Managers are coordinators, therefore the demand for managers should depend on the need for coordination. Evidence shows that the ratio of production to non-production workers has been steadily decreasing over time (See Radner, 1992), suggesting that coordinating production has become a more complex activity. But if we do not understand what coordination costs really are, we cannot understand this fairly robust evidence.

In this paper we study one type of coordination cost that is motivated by evidence from psychological and organizational behavior literature: imperfections in communication. We model imperfections in communication as arising from heterogeneity in knowledge. More specifically, individuals that have knowledge about different things will find it difficult to communicate among themselves. There is compelling evidence from psychology and orga-

nizational behavior literature that supports this assumption (see Heath and Staudenmayer, 2000). We take communication problems arising from differential knowledge as our main assumption, and then ask what are its implications for an information-processing organization. We discuss the trade-off between specialization of knowledge and communication costs, the role of top and middle managers, and the optimal design of hierarchies.

In order to understand the effects of imperfect communication, we should isolate it from other imperfections, therefore we ignore incentive alignment (or agency) problems. This separation between coordination and incentive issues is standard. The team theory approach to organizational problems (e.g., Marschak and Radner, 1972) focuses on imperfect information transmission when preferences are aligned, while the principal-agent approach (e.g., Holmstrom, 1979) focuses on imperfect preferences alignment when information transmission is perfect. This separation is not without costs, and the integration of the two approaches is a promising topic for future research.

Our first result, which follows almost immediately from the assumptions, is that the trade-off between specialization and communication implies that there is economic value to generalist workers. Generalists might not be as good as specialists in the acquisition of new information, but they are better in communicating it. We then show that the division of labor will imply that some agents will specialize in production while others specialize in the transmission of information. As a consequence, information flows from specialists to generalists. In terms of organization structure, *generalists are at the top of hierarchies while specialists are at the bottom*. This seemingly intuitive result is in sharp contrast with the theoretical literature on the assignment of heterogeneous agents to different hierarchical levels. The literature emphasizes the role of talent or ability as the main determinant of rank (see Rosen, 1982). We argue that this one-dimensional characterization of knowledge is too narrow and suggest a two-dimensional characterization instead: knowledge may differ in both breadth and depth. Due to limited cognitive capabilities, when persons invest in acquiring deeper knowledge about some things, they have to sacrifice the breadth of their

knowledge, and vice-versa. Therefore, the cognitive impossibility of knowing too much about too many things implies that some people will specialize in knowing too much while others will specialize in knowing too many things. There is no sense in which one is more talented than the other. Still, generalists are better in information processing and decision-making and they are naturally assigned to higher ranks.

This implication seems to be confirmed by casual evidence. In most organizations, persons with different types of knowledge are assigned to different hierarchical levels. For instance, it is common for the top management of a given firm to have an MBA-type of education, while at the lower level of the decision process one usually finds specialists in production or research. As a general trend, top managers have more general knowledge about the activities in the firm, in the sense that they know at least a little bit about each activity, while persons at the lower hierarchical levels have more specialized knowledge, that is, they have a deeper understanding of some few areas.

Our model also endogenously generates pyramid-like hierarchies with generalists at the top and specialists at the bottom. We show that the role of middle managers is to aggregate the information they receive from lower levels and then report it to the top management. Middle-managers are semi-specialized workers that function as translators of information sent by production workers to top decision makers. Therefore, the optimal number of layers in a hierarchy is determined by the available communication technology.

Our model implies a sharp trade-off between centralized and decentralized decision processes. In centralized organizations, the decision maker will have coarse information about many activities, while in decentralized ones the decision maker will have precise information about few activities, but no information about most of them. Which type of organization will prevail will depend then on how important the interaction among the activities is in determining the organization's payoff. Resulting from this trade-off, we find that centralization is more likely to occur the higher is the prior uncertainty about the activities and the better is the communication technology. These are, in principle, testable implications.

2 Literature Review

The pioneering work on the problem of the assignment of heterogeneous agents to different hierarchic levels is due to Rosen (1982). He developed a theory that explains the joint distribution of firm sizes and managerial compensation, based on the empirically observed fact that both are skewed to the right. The main idea in his paper is that slight improvements in upper levels decisions have an enormous influence on the productivity of subordinates. In his model, managers have two functions: supervision and management. Management involves choices and command, and has increasing returns to scale, since it does not depend on the number of subordinates. Supervision depends on the number of subordinates, so it does have decreasing returns to scale: it is the factor that prevents the firm from expanding indefinitely. Managers are embodied with different abilities. Higher quality managers both manage and supervise more efficiently. As an equilibrium result, better managers will be at the top, “Greater managerial talent commands greater resources” (p.317).

In a recent paper, Garicano (1999) addresses a similar question in a model that allows for managers to acquire different types of knowledge. As an equilibrium result, individuals at the bottom solve more common and easy problems, while individuals at the top solve difficult and rare problems. Superiors only act when asked questions from subordinates. Knowledge does not overlap, so superiors do not know what subordinates know, but there is a clear ranking of managerial talent.

Many other papers that address the problem of organization design assume homogeneous agents (see for example Sah and Stiglitz 1986, 1988, Radner 1992, 1993, Bolton and Dewatripont 1994, Van Zandt 1999).¹ They were not constructed to answer the questions we address in this paper, but mainly to stress the advantages of parallel over sequential

¹Sah and Stiglitz (1986,1988) do work with heterogeneous agents in an ex post sense, since in their model different agents end up making different decisions. However, this heterogeneity is not known ex ante, so it cannot be used to assign persons to different positions in the organization.

processing.² An important exception is the model of Prat (1997). Prat allows for different processors to have different capacities. As a result, higher capacity processors are assigned to higher hierarchic levels. A processor with a higher capacity is a metaphor for an abler manager. Therefore, again the equilibrium result implies abler managers assigned to higher levels.

In these models, a similar result arises, which says that individuals at higher hierarchic levels are abler or more “talented” (Rosen, Prat) or “solve more difficult problems” (Garicano). These results do have empirical content, but they require knowledge to be measured in a one-dimensional scale, so one can say which manager has more talent, or which problems are more difficult to solve. In this paper, we develop a model in which knowledge has two dimensions, and there is a trade-off between them. Therefore, there is no clear sense in which a person can be said to be more talented or better at solving difficult problems than another one.³

Harris and Raviv (1999) recently developed a model to explain the choice between hierarchies and matrix forms that has some of the elements of our analysis. They assume that each manager is capable of detecting and coordinating interactions only within his limited area of expertise, which is similar to our assumptions. However, even though middle managers have different coordination expertise, they assume that the CEO can coordinate any interaction. In this sense, CEO is still more talented than middle managers. It is the variable cost of the CEO that prevents using him all the time.

Hart and Moore (1999) developed a rather different model of hierarchies, in which they

²Sah and Stiglitz are concerned with minimizing different types of errors in decision making, while Radner and Van Zandt are concerned with reducing delay in information processing. In Bolton and Dewatripont’s (1994) paper, specialization makes agents better processors of information.

³Garicano’s model is actually closer to ours than the others, in the sense that one of his results is that knowledge of two different classes of people does not overlap, so there are things that managers know that their superiors do not know. We get this last result as well, but we actually require knowledge to overlap at least to some extent, in order to make communication feasible.

use the “authority approach”, meaning that subordinates only act when coordinators tell them to. (In Garicano’s and Harris and Raviv’s papers, for example, the coordinator acts only when subordinates ask her to). Managers are assumed to have command over a given number of assets. If a manager has an idea about a way of coordinating different assets under her control, she can implement the idea. If she does not have an idea, she can always delegate the authority to implement ideas to a subordinate. Their critical assumption is that the probability of having an idea is decreasing in the set of assets being looked after. In their own words, “coordinators are not supermen”. This assumption is very similar to our trade-off between depth and breadth of knowledge. However, they assume that agents are homogeneous, but have control over different assets. Individuals are assigned to a set of assets, and then their main theorem says that individuals looking after more assets should be senior to the ones with less assets. But who should be assigned to which asset? This assignment is not related to any measure of previous knowledge that the agents might have. And this is actually the problem that this paper is trying to address.

Another recent paper that have some similarities to this one is Vayanos’s (1999). In his model, top managers process more aggregated information, while managers at the bottom process only local information. There are synergies or interactions between areas, which means that one agent’s information is relevant to agents in other parts of the organization. The two crucial assumptions are that aggregation results in loss of useful information and that agents are limited in processing information. Specifically, his problem is one of constructing a portfolio of assets, given that each agent can only analyze up to k assets (but it does not matter which assets). Communication can only occur in hierarchic channels, and only limited information about each agent’s portfolio can be transmitted to superiors. But individuals are homogenous ex ante, so one does not get implications for the assignment of people to different levels. When he allows for some agent to be less skilled, in the sense of being able to process fewer assets, his conclusions are either that it does not matter to each level the less talented agent is assigned or that the agent should be at the top. His implications for organization

design are also quite different from ours. First, he assumes hierarchic structures. Second, his main result is that every manager will have at most one subordinate. In this paper, we allow for non-hierarchic forms of organization and show that they might be optimal. Also, our typical optimal structure is usually pyramidal, with managers having more than one subordinate.

Aghion and Tirole (1997) and Baker, Gibbons and Murphy (1999) modeled the relation between information and allocation of decision rights. Similarly to our model, the more informed an agent is, the better she is in making decisions. Their main trade-off is that delegation of authority might induce people to better use their information, but it creates agency costs. In our model, there are no agency problems, but agents have different types of knowledge and cannot communicate perfectly. Therefore, in the assignment of decision making rights to managers, one should take that into account.

3 Framework

3.1 Knowledge

Suppose there are n activities, or areas of expertise. Activities are represented by i.i.d. random variables⁴ X_1, X_2, \dots, X_n that are normally distributed with mean μ and variance σ^2 . There are two categories of individuals: specialists and generalists. Roughly speaking, specialists have knowledge of only one given area of expertise, while generalists have knowledge of more than one area of expertise. Let us call the set of all possible activities by $N = \{1, 2, \dots, n\}$. Let A be a non-empty subset of N . We call an A -generalist an individual that has knowledge of the activities in A only. If A is a singleton, i.e. $A = \{i\}$, an A -generalist is a specialist in activity i , in which case we call her an i -specialist. Therefore, generalists can be of $\sum_{j=2}^n \binom{n}{j}$ types. We call $\#A$ the cardinality of an A -generalist, which is

⁴We denote the realization of a random variable X_i by its lowercase counterpart x_i .

the number of elements in A . In some cases, it will be convenient to characterize a generalist by its cardinality. For example, we will use the term 3-generalist to refer to any generalist with cardinality 3.

3.2 Communication Technology

Every individual in this economy can send and receive *reports* about each activity. The assumption is that communication is imperfect. As Arrow noticed, “if this were not so, there would be no reason not to transfer all information on the availability of the resources and the technology of production to one place and compute at one stroke the optimum allocation of resources” (Arrow, 1982, p.1). We model imperfect communication by assuming that the quality of the received report depends on the knowledge gap between the sender and the receiver. The idea is that when persons with the same level of expertise communicate among themselves, no noise is introduced. However, when someone gets a report from a person with a narrower knowledge of a given set of activities, she can only get a noisy signal of the message sent. Equivalently, one could say that generalists can only process aggregated information, while specialists can process disaggregated information about fewer activities.⁵

There are arguably two types of noise introduced when two persons communicate with each other: one is generated by the sender of the message and the other is generated by the receiver.⁶ The receiver cannot fully understand the message because of cognitive limitations. We assume that the more the receiver knows about the nature of the activity that she is getting reports on, the better she can understand the reports, therefore the smaller is the noise she introduces in the communication process. The noise introduced by the sender

⁵Information aggregation is the origin of imperfect communication in Vayanos’s (1999) paper. Differently from our model, however, the ability to process disaggregated information in his model does not depend on the agents’ type of knowledge.

⁶“The bounded rationality of economic agents means that there are limits on their ability to communicate, that is to formulate and send messages and to read and interpret messages” (Van Zandt, 1998).

arises from her inability to perceive the limitations of the receiver. Since the sender is uncertain about which type of information the receiver will be able to process, she might choose the “wrong” communication devices, like using difficult concepts or lines of exposition that are not quite suited for the receiver. It is important to know your audience in order to communicate well. For example, an economist can be sure that another economist with a similar background will understand most of the economic jargon that she uses. However, if the same economist is reporting to an MBA, she will be less certain about what type of exposition she will have to do. If she is reporting to an engineer instead, this uncertainty will be even higher. Thus, we assume that the more “similar” (in a sense to be defined later) the knowledge of the sender and the receiver are, the smaller is the noise introduced by the sender in the communication process. That is, alike people think alike. The more you know someone, the better you are able to adapt your reports in order to make it easier for the receiver to understand.

We can pool the two types of noise into one, by simply assuming that this noise depends on the knowledge gap between the sender and the receiver. Formally, let s_i^A be the report about activity i sent by a type A person. If a type B person is the receiver, her reading of the report will be⁷

$$r_i^{BA} = s_i^A + \alpha(d_i^{BA}) u_i \quad (1)$$

where $u_i \sim N(0, \sigma_u^2)$. $\alpha(\cdot)$ is a measure of the precision of the communication, which depends on the distance in knowledge between the sender and the receiver. We assume that individuals can only report the signals that they have received from other persons or from nature. That is, individuals cannot choose which signal they will send (this is just a simplification). Equation (1) can be easily interpreted in terms of aggregation of information. While agent A knows the two pieces of information, s_i^A and $\alpha(d_i^{BA}) u_i$, agent B can only process the aggregated signal r_i^{BA} , which is the sum of the information bits known by A .

⁷We will use R_i to differentiate the random variable from its realization r_i , when this distinction is important.

The extent to which aggregation distorts the original message is assumed to depend on the knowledge gap between the sender and the receiver, as argued above. We define a measure of distance in knowledge as following: when $i \in A \cap B$

$$d_i^{BA} = \max \{ \#A, \#B \} - \#(A \cap B) \quad (2)$$

where $\#X$ is the number of elements in X . If $i \notin A \cap B$, then we set d_i^{BA} to be arbitrarily high; for concreteness we assume $d_i^{BA} = \infty$.

The interpretation of equation (2) is easier than it might seem at first glance. If both agents have some knowledge of the considered activity (i.e., if $i \in A \cap B$), the distance in knowledge increases with the difference in the degree of specialization between the two (which could be measured by $|\#A - \#B|$), but it also considers how many expertises they have in common, $\#(A \cap B)$. If $i \notin A \cap B$, either the sender or the receiver (or both) are completely unable to process information about the activity, so we set the distance in knowledge to be infinite in this case. It is also important to notice that the error in communicating $\alpha(d_i^{BA}) u_i$ does not depend on anything specific to the sender nor the receiver, except for their types. Therefore, two different agents with the same type A will generate exactly the same error when reporting to B (note that u_i does not vary across agents, only across activities).

The intuition is that the greater the distance in knowledge between the sender and the receiver, the less precise is the report as read by the latter. We formalize this by making the crucial assumption that $\alpha : \{0, 1, \dots, n\} \cup \{\infty\} \rightarrow [0, \infty)$ is a strictly increasing function. We also assume some boundary conditions $\alpha(0) = 0$ and $\alpha(\infty) = \infty$.

3.3 The Organization's Objectives

The organization has to decide whether to undertake a project or not. If the project is undertaken, there is a fixed cost of c , otherwise there are no other costs. If the project is

undertaken, the ex post profits are given by

$$\pi = \sum_{i=1}^n x_i - c \quad (3)$$

The timing of the decision process is as following. There should be at least one specialist involved in activity i in order to generate a reading of X_i . By assumption, generalists cannot get readings from nature. Each i -specialist will observe the realization x_i . Then, they are free to report to any other members of the organization. Everyone that received a report can also send reports to other members. After all reporting is done, one member of the organization, that has formal authority over the project, will decide whether or not to undertake the project.

The organization design problem thus consists in (i) deciding how many individuals will join the organization, (ii) which types of knowledge the members will have, (iii) who reports to whom, and (iv) who has the authority to decide. The goal is to maximize ex ante expected profits. It is assumed that every member acts in the interest of the team (no agency problems), the costs of reporting and adding members is zero and there is no delay.

However, in the discussion that follows, in most cases we will be implicitly assuming away structures that have redundant members (that is, members that do not convey useful information or do not improve upon the existing amount of information).

4 Finding Optimal Structures

Here we address the question of organization design. First, we need to define what we mean by *structure*.

Definition 1 *An organizational **structure** consists of*

- *a set of m members $M = \{m_1, \dots, m_m\}$ in which every element of M is a subset of N ; i.e., $m_i \in M \Rightarrow m_i \subseteq N$.*

- a reporting correspondence $R : M \rightarrow M$, such that $R(m_i) \subseteq M$ is the set of members that receive reports from member $m_i \in M$ (if some member m_k does not send reports, $R(m_k) = \emptyset$).
- a decision maker $m^* \in M$, which is the member who has the formal authority over the project.

This definition uses the fact that all agents are completely characterized by their types, which are subsets of N .

In order to compare different structures, we need some criteria to choose among structures that lead to the same expected profits, but with different numbers of members and reports. The most natural way is to impose some costs of adding managers and some costs of reporting. However, it would be easier to ignore these costs to focus on structures that have the best information processing properties. Other types of costs that we are ignoring for now are the costs of acquiring knowledge and the costs of delay. We postpone to section 6 the discussion about the effects of allowing for some of these costs on our results. However, in the discussion that follows, in most cases we will be implicitly assuming away designs that have redundant members (that is, members that do not convey useful information).

4.1 The Flat Hierarchy

In this section, we restrict ourselves to a world in which only $n + 1$ types of knowledge are available: specialists in each activity and N -generalists (i.e., generalists that have some knowledge about all n activities). We then show that there is only one possible type of hierarchy in this world, the so-called *flat hierarchy*: all lower-level workers report directly to one single manager. In other words, there are no middle managers. We consider this case here first mainly for simplicity, and postpone the discussion of hierarchies with many layers to the next section. Much of the intuition can be gained by considering this simpler case, and the results in the following sections are simple generalizations of the results in this one.

The following lemmas are straightforward.⁸

Lemma 1 *Specialists should never receive reports.*

Lemma 2 *N-generalists should never send reports.*

Therefore, only communication between generalists and specialists will be considered. The distance in knowledge between any specialist and the N -generalist is $d_i^{Ni} = n - 1$. To simplify notation, we denote $\alpha(n - 1) = \alpha$.

Now there are four questions we want to address: (1) How many agents should the organization employ? (2) What are their types (generalists or specialists)? (3) What is the optimal reporting structure? (4) Who should be the decision maker?

Suppose first that a specialist is the decision maker. By Lemma 1, we know that in such case the decision maker will receive no reports and will have to decide whether to undertake the project based only on her own information. Therefore, if an i -specialist makes the decision, she will undertake the project if and only if

$$E[\pi | X_i = x_i] = x_i + (n - 1)\mu - c \geq 0 \quad (4)$$

That is, she will undertake the project when its expected profits are nonnegative, given her private information on X_i . Therefore, ex ante expected profits are

$$E(\pi) = [1 - \Phi(a)] \{E(X_i + (n - 1)\mu - c | X_i + (n - 1)\mu \geq c)\} \quad (5)$$

where $a = \frac{c - n\mu}{\sigma}$ and $\Phi(\cdot)$ is the standardized normal cdf.⁹

On the other hand, in an organization in which a generalist is the decision maker, getting reports from others will never make the decision maker less informed and it will sometimes make her better informed. Therefore, she should combine the signals (reports) that she

⁸All proofs of lemmas and propositions are in the Appendix.

⁹For details, see the proof of Proposition 1 in the appendix.

gets with her prior knowledge about the probability distributions in a Bayesian manner, in order to decide whether or not to undertake the project. Therefore, if an N -generalist is the decision maker, receiving reports on the n activities, then the project will be undertaken if and only if

$$E[\pi \mid (R_1, \dots, R_n) = (r_1, \dots, r_n)] = n\mu + \sum_{i=1}^n \beta_i (r_i - \mu) - c \geq 0 \quad (6)$$

where r_i is the reading the decision maker gets about activity i and β_i is the optimal weight that she will give to the report about activity i , as in standard signal-extraction problems. By Lemma 2, in an optimal structure the generalist can only receive reports from specialists.

Ex ante expected profits are

$$E(\pi) = [1 - \Phi(b)] E \left[\sum_{i=1}^n X_i - c \mid n\mu + \sum_{i=1}^n \beta_i (R_i - \mu) \geq c \right] \quad (7)$$

where $b = \frac{c - n\mu}{\beta \sqrt{n(\sigma^2 + \alpha^2 \sigma_u^2)}}$.¹⁰

Therefore, there could be only two possible structures. In a *specialist-managed organization (smo)*, expected profits are given by (5), one specialist is the decision maker, there are n specialists and no generalists, and no reporting. In a *generalist-managed organization (gmo)*, expected profits are given by (7), one generalist is the decision maker, there are n specialists and one generalist, and all specialists report to the generalist.¹¹ Now we shall compare the two possible structures.

Proposition 1 (*The Demand for Generalists*) *The generalist-managed organization is no worse (no better) than a specialist-managed organization if and only if $n \geq 1 + \frac{\alpha^2 \sigma_u^2}{\sigma^2}$ ($n \leq 1 + \frac{\alpha^2 \sigma_u^2}{\sigma^2}$).*

Then, the following results are straightforward.

¹⁰Again, see the appendix for details.

¹¹Here we are ignoring redundant members.

Corollary 1 *If the prior uncertainty about the activities (σ^2) is sufficiently high, it is optimal to have a generalist manager (everything else constant).*

The intuition behind this result is that, with more ex ante uncertainty, the less useful is the knowledge of the prior distribution of the activities. Therefore, Bayesian generalist managers will give less weight to the prior distribution and more weight to the reports they receive when constructing their posteriori distributions. But specialists do not get any reports (see Lemma 1), so they have to make decisions based only on the priors. Therefore, their decision rules will not adjust to reflect this increase in risk, while the decision rule of the generalist will optimally respond and give less weight to the prior.

Corollary 2 *If the communication technology is sufficiently precise (low $\alpha^2\sigma_w^2$), it is optimal to have a generalist manager (everything else constant).*

The intuition is that if communication is very precise, the solution with one generalist manager will be very close to the full information solution.

In what follows, we want to focus on generalist-managed organizations. Therefore we assume the following:

Assumption A.1 $n \geq 1 + \frac{\alpha^2\sigma_u^2}{\sigma^2}$

The following corollary is just a restatement of Proposition 1.

Corollary 3 (Optimality of Flat Hierarchies) *If Assumption A.1 holds, an optimal organizational structure will have $n + 1$ members such that*

1. *There are n specialists, one for each activity, and one generalist;*
2. *All specialists report to the generalist;*
3. *The generalist is the decision maker.*

4.2 Comparing Flat Hierarchies

In the previous section, we showed the conditions under which it would be optimal to have a flat hierarchy with an N -generalist as the top manager. In this section, we assume that all n types of knowledge are feasible. That is, individuals can be either specialists or one of the $\sum_{j=2}^n \binom{n}{j}$ types of generalists. The question now is How can we compare flat hierarchies with different types of generalists at the top?

Let an A -generalist be the decision maker, receiving reports from n specialists. Without loss of generality, if $\#A = k_A \leq n$, we define A to be $A = \{1, 2, \dots, k_A\}$. Then the project will be undertaken if and only if

$$E[\pi \mid (R_1, \dots, R_{k_A}) = (r_1, \dots, r_{k_A})] = n\mu + \sum_{i=1}^{k_A} \beta_A (r_i - \mu) - c \geq 0 \quad (8)$$

where $\beta_A = \frac{\sigma^2}{\sigma^2 + \alpha_A^2 \sigma_u^2}$, $\alpha_A = \alpha(d_i^{A_i})$, $\forall i \in A$. Thus, the probability of undertaking the project is¹²

$$1 - \Phi(b_A) \quad (9)$$

where $b_A = \frac{c - n\mu}{\sqrt{k_A \beta_A \sigma}}$.

The ex ante expected profits in this case will be

$$[1 - \Phi(b_A)](n\mu - c) + \sqrt{k_A \beta_A \sigma} \phi(b_A) \quad (10)$$

Now the problem is to find the degree of specialization of the top manager k that maximizes ex ante expected profits. It is important to realize the nature of the relevant trade-off here: broadening the knowledge of the top manager (increasing k_A) will allow her to get readings from a larger set of activities, but at a cost of reducing the precision of her readings (since $d_i^{A_i}$ will increase, $\forall i \in A$). We state that formally as

¹²The derivations of equations 9 and 10 follow the same steps as in the proof of Proposition 1, and therefore are omitted.

Proposition 2 (Comparison of Flat Hierarchies) *Let A and B be any two generalists with cardinality k_A and k_B , respectively. A flat hierarchy with A at the top is preferable to a flat hierarchy with B at the top if and only if*

$$\frac{k_A}{k_B} \geq \frac{\sigma^2 + \alpha_A^2 \sigma_u^2}{\sigma^2 + \alpha_B^2 \sigma_u^2} \quad (11)$$

Notice that condition (11) is just a generalization of Assumption A.1. To see this, notice that the cardinality of an N -generalist is n , while the cardinality of a specialist is 1. Since $\alpha_i = 0$, substituting n for k_A , 1 for k_B , α for α_A and 0 for α_B , we get Assumption A.1.

Given (A.1), the solution of choosing k_A that maximizes (10) will be k^* such that $N \geq k^* > 1$. If k^* is strictly less than N , we have that the top manager does not get any readings from a set of $N - k^*$ activities. Instead of imposing a series of conditions like (11), we can always redefine $N' = k^*$ and $c' = c - (N - k^*)\mu$, so the top manager will always be an N' -generalist, being N' the number of all activities in the organization that are accountable to the top manager. In this case, we can safely ignore a set of $N - N'$ activities that are out of the control of the top manager. Thus, we can assume without loss of generality that the top manager will always have knowledge about all *accountable* activities in the organization; that is, all activities that actually report to someone. Given this result, in what follows we will always assume that $N' = N$.

4.3 Multiple Levels of Managers

Keeping assumption A.1, it is clear by Proposition 1 that a flat hierarchy (one in which every specialist reports directly to the same manager, which is an N -generalist) is better than no hierarchy at all. The question now is whether the availability of semi-generalists can improve upon the flat hierarchic structure. It is clear that, if semi-generalists are to be used at all, they should be intermediaries between the top manager (the N -specialists) and the specialists. We state that formally as

Proposition 3 (*Knowledge and Rank*) *A B-generalist receives reports about activity i from A-generalists only if $i \in A \cap B$ and $\#B > \#A$. In words, **managers with more specialized knowledge are subordinates to agents with less specialized knowledge.***

Then it is clear that the role of the middle managers is to reduce noise in communication. Let an A -generalist of cardinality k be an intermediary between the k specialists and the N -generalist. The report that N will get from A about activity i is (recall the assumption that individuals can only report the signals that they receive)

$$r_i^{NA} = s_i^A + \alpha(d_i^{NA}) u_i = x_i + \alpha(k-1) u_i + \alpha(n-k) u_i \quad (12)$$

If the top manager gets her report directly from the i -specialist, her reading will be

$$r_i^{Ni} = s_i^i + \alpha(d_i^{Ni}) u_i = x_i + \alpha(n-1) u_i \quad (13)$$

Since the middle manager is a means through which information flows from bottom to top, introducing the middle manager of type A to receive a report about i from the i -specialist and then send it to the top manager is better than the flat hierarchy if and only if

$$\alpha(k-1) + \alpha(n-k) < \alpha(n-1) \quad (14)$$

We generalize the previous argument in the following proposition.

Proposition 4 (*Middle Management*) *Take any given organizational structure. Say that manager A reports to manager B . If $\#B - \#A = 1$, than no middle manager should be introduced between the two (see Proposition 3). If $\#B - \#A > 1$, a middle manager C with $\#B > \#C > \#A$ should be introduced between the two if and only if there is at least one i such that $i \in A \cap B \cap C$ and $\alpha(\#B - \#C) + \alpha(\#C - \#A) < \alpha(\#B - \#A)$.*

It is clear that the main determinant of middle management is the shape of the “precision” function $\alpha(\cdot)$. The economic reasoning behind this result is easily understood. Consider

equation (14). It says that if two persons with a big distance in knowledge between them communicate with each other, the noise introduced is greater than the one generated when someone is introduced between them to “filter” the information. Of course, the middle manager cannot be so far away from the two extremes (in terms of knowledge) in order for this filtering to occur.

When Assumption A.1 holds, the following results are immediate from Proposition 4.

Corollary 4 *If the precision function is **subadditive**, then the optimal structure is the flat hierarchy.*

Corollary 5 *If the precision function is **strictly superadditive**, then the optimal structure has n layers of management.*

5 Characteristics of Optimal Structures

In this section, we discuss in more detail some additional properties of optimal structures. First, we need some few definitions. In what follows, we would like to be as close as possible to the definitions encountered in the literature, in order to facilitate comparisons with our results. Unfortunately, there is no generally agreed terminology for describing the characteristics of organizational structures. We take as our benchmark the work of Radner (1993), both because of its rigorously defined concepts and because of its influence on other works. We also explain the differences between our definitions and the ones found in the literature, whenever those differences are significant enough.

Definition 2 *A structure is **hierarchical** if the reporting correspondence $R : M \rightarrow M$ is a function. In words, in a hierarchy, every subordinate reports to at most one manager.*

Non-hierarchic forms occur when at least one of the managers report to more than one superiors, like in matrix forms. This definition is implicitly adopted by most papers on

organization design.¹³

Definition 3 *We say that a structure has l_i levels of managers in activity i if information about i has to pass through $l_i - 1$ managers before it reaches the decision maker.*

Definition 4 *A structure is **strictly balanced** if all activities have the same number of levels of managers l and all managers of the same level have the same number of immediate subordinates.*

Since Radner was concerned only with hierarchic structures, he only defined strictly balanced *hierarchies*.¹⁴ However, since we want to consider the possibility of non-hierarchic structures, we extended his definition of strictly balanced structures to include non-hierarchic structures as well. There is no necessary relation between hierarchic and strictly balanced structures. Figure 1 shows a hierarchic but not strictly balanced structure, while Figure 2 shows a strictly balanced but not hierarchic structure.

Now we are ready to show the following result.

Proposition 5 (*Optimality of Strictly Balanced Hierarchies*) *There always exists a strictly balanced hierarchy that maximizes expected profits.*

It is important to stress that this result does not say that other types of structures are not optimal. It is possible for matrix structures or structures that have “skip-level” reporting to be optimal. What Proposition 5 does say is that, as long as only the information processing properties of structures are important, no improvement can be achieved by deviating from a strictly balanced hierarchy.

Some classical papers on organizations have assumed strictly balanced hierarchies without comparing them with other alternatives, like in Beckmann (1960) and Keren and Levhari (1983). Proposition 5 suggests that this approach may be a reasonable one.

¹³Radner (1993) and Harris & Raviv (1999) explicitly define hierarchies in this way.

¹⁴In Radner’s 1992 paper, he used the term *regular* instead of *strictly balanced*.

As we will see later in this section and in Section 6, introducing costs of hiring managers or acquiring knowledge may make non-balanced and non-hierarchic structures strictly better than strictly balanced hierarchies.

Example 1 Let $N = \{1, 2, 3, 4\}$, $\sigma^2 = \sigma_u^2 = 1$. Let the precision function be

$$\alpha(d_{BA}^i) = \left(\frac{d_{BA}^i}{3}\right)^2 \quad (15)$$

First, notice that Assumption A.1 holds, since $n = 4 > 2 = 1 + \frac{\alpha(3)^2 \sigma_u^2}{\sigma^2}$. Therefore, a flat hierarchy with a N -generalist at the top is better than the decentralized solution. To find the optimal structure, we can apply Proposition 4. First, should 2-generalists be introduced? Consider $A = \{1, 2\}$. We have that $d_{A1}^1 = d_{A2}^2 = 1$, therefore $\alpha(d_{A1}^1) = \alpha(d_{A2}^2) = \frac{1}{9}$. Also, $d_{NA}^1 = d_{NA}^2 = 2$, implying $\alpha(d_{NA}^1) = \alpha(d_{NA}^2) = \frac{4}{9}$. If specialists communicate directly with the N -generalist, then $d_{N1}^1 = d_{N2}^2 = 3$ and $\alpha(d_{N1}^1) = \alpha(d_{N2}^2) = 1$. Therefore,

$$\frac{1}{9} + \frac{5}{9} < 1 \quad (16)$$

Thus, A should be introduced. The same holds for $B = \{3, 4\}$. Now, let us check whether $C = \{1, 2, 3\}$ should be introduced. Applying the rule,

$$\alpha(1) + \alpha(1) = \frac{1}{9} + \frac{1}{9} < \frac{5}{9} = \alpha(2) \quad (17)$$

Thus, C should be introduced, as well as another 3-generalist that has some knowledge about $\{4\}$. Therefore, an optimal structure will have 4 levels of management, as shown in figure 3. We could have used Corollary 5 directly: since (15) is strictly superadditive, we would have four levels of management.

Notice that in this example, this optimal structure is a strictly balanced hierarchy. It is also interesting to notice that this structure minimizes the number of managers used. However, this is not a general property of strictly balanced structures, as we can see in the following example.

Example 2 Let $N = \{1, 2, 3, 4, 5, 6, 7\}$, $\sigma^2 = \sigma_u^2 = 1$. Let the precision function be

$$\alpha(d_{BA}^i) = \left(\frac{d_{BA}^i}{6}\right)^2 \quad (18)$$

This example is similar to the previous one: the precision function is strictly superadditive, therefore the optimal structure has 7 levels of managers. It is easy to see that any structure that minimizes the number of managers in this case has four 2-generalists at the second level and three 3-generalists at the third level. In order for this to happen, at least one 2-generalist has to report to two 3-generalists, as shown in Figure 4.

The previous example made clear that a hierarchy is not always the most economical way of using managers. We have been assuming so far that there are no costs of adding managers. But this assumption is admittedly false. Any organization should try to compare the costs of extra managers to the benefits they create. Thus, a natural question is: Which optimal structures minimize the number of managers?¹⁵

To address this question, we need a couple of additional definitions.

Definition 5 A structure is **balanced** if all activities have the same number of levels of managers l and, among the managers of the same level, at most one manager has fewer immediate subordinates than some other manager.

This definition is based on Radner (1993), but adapted to fit our model. Comparing Definitions 4 and 5, we can see that a strictly balanced structure is also balanced, as one would expect from the terminology we are using. The reason for defining balanced structures

¹⁵Here we are assuming a lexicographic ordering, in which we first look for structures that have the best information processing properties and then we choose among them the ones with the lowest number of managers. This procedure is a good approximation if the costs of managers are very low as compared to the costs of incomplete information. If the costs of managers are substantial, however, one should explicitly model this trade-off.

in this way is that sometimes it is not possible to have all managers at the same level with the same number of subordinates, due to integer constraints.

Consider a structure that has l levels in each activity. Let L_j be the set of all managers at level j (we count levels from bottom to top: 1 is the lowest level, where the specialists are, and j is the highest level, where we find the decision maker).

Definition 6 *A structure is **pyramidal** if it is balanced and $\#L_j \geq \#L_{j+1}$, for all $j = \{1, \dots, l - 1\}$.*

Not all balanced structures are pyramidal, as we can see from Figure 5. Hierarchies do not imply pyramids: see Figure 1. Also, a pyramid does not imply a hierarchy, as shown in Figure 6. However, we can show that

Lemma 3 *A balanced hierarchic structure is also pyramidal.*

Therefore, in a meaningful sense, a pyramid is a weaker concept than a balanced hierarchy.¹⁶ Trivially, from Proposition 5 and Lemma 3 we know that we can always find a pyramid that is an optimal structure. The next proposition shows another property of pyramidal structures.

Proposition 6 *(Pyramids minimize the use of managers) Among the class of optimal structures, there always exist a pyramidal structure that minimizes the total number of managers in the organization.*

¹⁶Bolton and Dewatripont's (1994) definition of pyramids is equivalent to our definition of hierarchy, therefore different from *our* definition of pyramids. Their definition of regular pyramids is virtually equivalent to our definition of strictly balanced hierarchies. Since their results always rule out non-hierarchic forms, all their definitions are related to hierarchies. We found it useful to disentangle the concepts of hierarchies and pyramids, being the latter a structure that resembles the geometric form of a triangle, despite of the structure of reporting. As we can see in Figure 6, even though the structure has a pyramidal shape, it is not hierarchic.

This proposition shows an important property of pyramidal structures. Proposition 6 says that, among structures that have the best information processing properties, pyramids are the ones that have the lowest costs of using managers. The same cannot be said about hierarchies, as we saw above.

Sometimes it is claimed that the span of control (defined as the number of subordinates of each manager at a given level) should be decreasing as one goes from lower to higher levels (see Keren and Levhari, 1983). This is not implied by Proposition 6. As a matter of fact, in an optimal structure that minimizes the number of managers, the span of control will usually be non-monotonic as one changes levels, as we can see from Figures 3 and 4.

6 Some Extensions

6.1 Restricting the Set of Available Knowledge

We have assumed so far a very simple knowledge technology: there is a trade-off between the number of activities one has knowledge about and the depth of her knowledge of each activity. We assumed that the costs of acquiring knowledge depend only on the number of activities, but not on the activities themselves. That is, acquiring knowledge of $\{1, 2\}$ is as costly as of $\{2, 3\}$ or $\{1, 3\}$. However, a more realistic assumption should allow for some set of activities to be easier to learn than others. For example, say that 1 is an “engineering” activity (e.g., product design) while 2 and 3 are “managing” activities (e.g., finance and marketing). Therefore, it seems natural to postulate that the knowledge set $\{2, 3\}$ is easier to acquire than $\{1, 2\}$ or $\{1, 3\}$.

As a simple way of imposing those different costs, we assume that only a subset of types $S \subset N$ is available. This restriction can be imposed to generate some realistic structures, as the following example illustrates.

Example 3 *Let $N = \{1, 2, 3, 4, 5, 6, 7, 8\}$, and let the precision function be strictly superad-*

ditive. Suppose that there are four divisions in the firm. Suppose that activity 1 is finance in division one, 2 is marketing in division one, 3 is finance in division two, 4 is marketing in division two, and so on. Suppose now that the only types of knowledge available are CEO-type of knowledge $N = \{1, 2, 3, 4, 5, 6, 7, 8\}$, financial knowledge $\{1, 3, 5, 7\}$, marketing knowledge $\{2, 4, 6, 8\}$, division-specific knowledge $[\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8\}]$ and specialized knowledge. Then, the optimal structure will have a mix of matrix and hierarchic features, as shown in Figure 7. Specialists in each division will report only to their division managers. Division managers will report both to the financial and the marketing manager. Financial and marketing managers will report only to the CEO.

6.2 Authority Delegation

We have assumed, for simplicity, that all activities have the same prior distribution and all agents know those priors exactly. This need not be the case. If i -specialists cannot understand reports about activity j , why should they be able to know anything at all about j ?

In this section, we drop the assumption that everyone knows all prior distributions. We rationalize that by assuming that those distributions keep changing over time, so i -specialists are more able to perceive changes in the distribution of i than other specialists or generalists. As a simplification, we assume that σ_i^2 , the variance of activity i , is the only parameter that changes. We assume that they can only assume two values

$$\sigma_H^2 > \sigma_L^2 \tag{19}$$

For simplicity, we treat only the case in which $N = \{1, 2\}$. The timing is as following. In the beginning of each period, each specialist can observe the variance of the distribution of their activities. Then, they can report it to the generalist. Generalists can perfectly read reports about both variances, while specialists cannot read reports about activities other

than their own.¹⁷ After all reporting about variances is done, the decision maker can decide whether or not to undertake the project, or to delegate the authority over the project to one of two specialists. If the top manager keeps the authority to herself, she gets reports about X_1 and X_2 from the specialists.

Straightforward calculations show that the generalist will keep the authority to herself if and only if

$$h(\sigma_1^2, \sigma_2^2) \equiv \beta_1^2 (\sigma_1^2 + \alpha^2 \sigma_u^2) + \beta_2^2 (\sigma_2^2 + \alpha^2 \sigma_u^2) \geq \max \{\sigma_1^2, \sigma_2^2\} \quad (20)$$

Otherwise, she will delegate it to the specialist of the activity with the higher variance (if both have the same variance, she flips a coin).

Example 4 *Suppose $h(\sigma_H^2, \sigma_L^2) < \sigma_H^2$, $h(\sigma_H^2, \sigma_H^2) > \sigma_H^2$ and $h(\sigma_L^2, \sigma_L^2) > \sigma_L^2$. Therefore, the optimal allocation of authority is to keep the authority with the generalist in states $\{(\sigma_L^2, \sigma_L^2), (\sigma_H^2, \sigma_H^2)\}$, give the decision to the 1-specialist when (σ_H^2, σ_L^2) happens and to the 2-specialist when (σ_H^2, σ_L^2) happens.*

Notice that in the previous example it is important that formal authority is at the top. If some specialist, say 1, was responsible for delegating authority, she could not implement the optimal rule, since this requires knowing the realization of σ_2^2 .

7 Conclusions

[To be done]

¹⁷This is inessential. All that is needed is that the generalists' reading of σ_i^2 is less noisy than the j -specialists' reading of σ_i^2 .

Appendix

Proof of Lemma 1

Proof. If anyone reports to a j -specialist about activity $i \neq j$, she will receive a signal of infinite variance since $\alpha(\infty) = \infty$. Therefore, this signal is useless for the j -specialist, either for decision making or for further reporting. If an i -specialist reports to another i -specialist, both will accurately know the realized state x_i , but the same information could have been acquired from production of x_i without reporting, so one of the specialists is redundant. ■

Proof of Lemma 2

Proof. Generalists should not report to specialists, by Lemma 1. Since generalists can not get readings from nature, they can only send reports if they get reports from other agents. Therefore, an N -generalist that receive reports from another N -generalist could get the same information if she gets reports from the latter's sources, economizing the use of at least one redundant manager without loss of information. Thus, N -generalists should not report to other N -generalists. ■

Proof of Proposition 1

Proof. If a specialist takes the decision, she will undertake the project if and only if (see equation 4)

$$x_i + (n - 1) \mu \geq c \tag{21}$$

So the probability of undertaking the project is

$$\begin{aligned} \Pr(X_i \leq c - (n - 1) \mu) = \\ \Pr\left(\frac{X_i - \mu}{\sigma} \leq \frac{c - n\mu}{\sigma}\right) = 1 - \Phi(a) \end{aligned}$$

Where $a = \frac{c - n\mu}{\sigma}$ and $\Phi(\cdot)$ is the standardized normal cdf. Thus, the ex ante expected profits are given by

$$[1 - \Phi(a)] \{E(X_i + (n - 1) \mu - c \mid X_i + (n - 1) \mu \geq c)\} =$$

$$[1 - \Phi(a)] \{E(X_i | X_i \geq -(n-1)\mu + c) + (n-1)\mu - c\}$$

Using the formula for the expectation of truncated normal distributions, we get

$$[1 - \Phi(a)] \left\{ \mu + \sigma \frac{\phi(a)}{1 - \Phi(a)} + (n-1)\mu - c \right\} =$$

$$[1 - \Phi(a)] (n\mu - c) + \sigma\phi(a) = E_{smo}(\pi) \quad (22)$$

Where ϕ is the standardized normal density function. We denote by $E_{smo}(\pi)$ the ex ante expected profits of the specialist-managed organization.

On the other hand, if a N -generalist is the decision maker, receiving reports from n specialists, then she will Bayesian update her expectations over the activities. More precisely, given (1) we have

$$x_i = r_i - \alpha u_i$$

Therefore,

$$E[X_i | R_i] = \mu + \beta(R_i - \mu)$$

where

$$\beta = \frac{\text{cov}(R_i, X_i)}{\text{var}(R_i)} = \frac{\text{cov}(X_i + \alpha u_i, X_i)}{\text{var}(X_i + \alpha u_i)} = \frac{\sigma^2}{\sigma^2 + \alpha^2 \sigma_u^2}$$

Then the project will be undertaken if and only if (see equation 6)

$$n\mu + \sum_{i=1}^n \beta(r_i - \mu) \geq c$$

Thus, the probability of undertaking the project is

$$\Pr \left[n\mu + \sum_{i=1}^n \beta(R_i - \mu) \geq c \right] =$$

$$\Pr \left[n\mu + \sum_{i=1}^n \beta(X_i + \alpha u_i - \mu) \geq c \right] =$$

$$\Pr \left[\sum_{i=1}^n \beta(X_i + \alpha u_i) \geq c + n\mu(\beta - 1) \right] =$$

$$\Pr \left[\frac{\sum_{i=1}^n \beta (X_i + \alpha u_i) - \beta n \mu}{\beta \sqrt{n (\sigma^2 + \alpha^2 \sigma_u^2)}} \geq \frac{c + n \mu (\beta - 1) - \beta n \mu}{\beta \sqrt{n (\sigma^2 + \alpha^2 \sigma_u^2)}} \right] = 1 - \Phi(b)$$

where

$$b = \frac{c - n \mu}{\beta \sqrt{n (\sigma^2 + \alpha^2 \sigma_u^2)}} = \frac{a}{\sqrt{n \beta}}$$

Therefore,

$$a \geq b \Leftrightarrow \sqrt{n \beta} \geq 1 \Leftrightarrow \beta \geq \frac{1}{n} \Leftrightarrow \frac{\sigma^2 + \alpha^2 \sigma_u^2}{\sigma^2} \leq n \Leftrightarrow 1 + \frac{\alpha^2 \sigma_u^2}{\sigma^2} \leq n.$$

The ex ante expected profits in this case will be

$$\begin{aligned} & [1 - \Phi(b)] E \left[\sum_{i=1}^n X_i - c \mid n \mu + \sum_{i=1}^n \beta (R_i - \mu) \geq c \right] = \\ & [1 - \Phi(b)] E \left[n \mu + \sum_{i=1}^n \beta (R_i - \mu) - c \mid n \mu + \sum_{i=1}^n \beta (R_i - \mu) \geq c \right] = \\ & [1 - \Phi(b)] \left\{ n \mu (1 - \beta) - c + n \beta \mu + \beta \sqrt{n (\sigma^2 + \alpha^2 \sigma_u^2)} \frac{\phi(b)}{1 - \Phi(b)} \right\} = \\ & [1 - \Phi(b)] (n \mu - c) + \beta \sqrt{n (\sigma^2 + \alpha^2 \sigma_u^2)} \phi(b) = E_{gmo}(\pi) \end{aligned} \quad (23)$$

We denote by $E_{gmo}(\pi)$ the ex ante expected profits of the generalist-managed organization.

Since $\beta = \frac{\sigma^2}{\sigma^2 + \alpha^2 \sigma_u^2}$, we have that

$$\beta \sqrt{n (\sigma^2 + \alpha^2 \sigma_u^2)} = \sqrt{n \beta} \sigma$$

Now define the function

$$g(x) = \left[1 - \Phi \left(\frac{c - n \mu}{x \sigma} \right) \right] (n \mu - c) + x \sigma \phi \left(\frac{c - n \mu}{x \sigma} \right) \quad (24)$$

Take its derivative with respect to x to get

$$g'(x) = \sigma \phi \left(\frac{c - n \mu}{x \sigma} \right) > 0$$

Now let $x = \sqrt{n\beta}$. If $x \geq 1$, $E_{gmo}(\pi) = g(x) \geq g(1) = E_{smo}(\pi)$. But $\sqrt{n\beta} \geq 1 \Leftrightarrow \beta \geq \frac{1}{n}$
 $\Leftrightarrow \frac{\sigma^2 + \alpha^2 \sigma_u^2}{\sigma^2} \leq n \Leftrightarrow 1 + \frac{\alpha^2 \sigma_u^2}{\sigma^2} \leq n$ ■

Proof of Proposition 2

Proof. Notice that $\beta_A \sqrt{k_A (\sigma^2 + \alpha_A^2 \sigma_u^2)} = \sqrt{k_A \beta_A} \sigma$ and $\beta_B \sqrt{k_B (\sigma^2 + \alpha_B^2 \sigma_u^2)} = \sqrt{k_B \beta_B} \sigma$. Recall the function $g(\cdot)$ as defined in (24). From (10), expected profits under A and B are given by $g(\sqrt{k_A \beta_A} \sigma)$ and $g(\sqrt{k_B \beta_B} \sigma)$. But since $g(\cdot)$ is an increasing function, $g(\sqrt{k_A \beta_A} \sigma) \leq g(\sqrt{k_B \beta_B} \sigma) \Leftrightarrow \sqrt{k_A \beta_A} \sigma \leq \sqrt{k_B \beta_B} \sigma \Leftrightarrow \sqrt{k_A \beta_A} \leq \sqrt{k_B \beta_B} \Leftrightarrow \frac{k_A}{k_B} \leq \frac{\beta_B}{\beta_A} \Leftrightarrow \frac{k_A}{k_B} \leq \frac{\sigma^2 + \alpha_A^2 \sigma_u^2}{\sigma^2 + \alpha_B^2 \sigma_u^2}$ ■

Proof of Proposition 3

Proof. If $i \notin B$, the B -generalist will receive a uninformative report, since $\alpha(d_i^{BA}) = \alpha(\infty) = \infty$. If $i \notin A$, the only information that the A -generalist knows is the prior distribution of i , which is assumed to be known by everyone. Suppose now that $i \in A \cap B$, but $\#B \leq \#A$, and C be the type from which A got her signal about i (C can be nature, in which case $\#C = 1$). Then, from 2, we see that $d_i^{BC} \leq d_i^{AC}$, therefore B could get a signal directly from C that would be at least as precise than the one A can get from C . Therefore, direct communication from C to B yields with no loss of information, so A is redundant. ■

Proof of Proposition 4

Proof. Let $\#B - \#A > 1$, and consider a middle manager C with $\#B > \#C > \#A$. Suppose there is one i such that $i \in A \cap B \cap C$ and $\alpha(\#B - \#C) + \alpha(\#C - \#A) < \alpha(\#B - \#A)$. Therefore, introducing C will improve the precision of the signal received by B , thus also improving the precision of the decision maker, which will increase expected profits. Suppose now that $\alpha(\#B - \#C) + \alpha(\#C - \#A) > \alpha(\#B - \#A)$. Then, the opposite will occur: introducing C will worsen the precision of the signal, therefore reducing expected profits. If $\alpha(\#B - \#C) + \alpha(\#C - \#A) = \alpha(\#B - \#A)$, introducing one more manager does not change expected profits, thus, from Definition 2, it is better not to do so. Finally, if

$i \notin A \cap B \cap C$, only signals with infinite variance can flow from A to B , passing through C . ■

Proof of Lemma 3

Proof. Let $R(L_j)$ be the set of superiors of the managers at level j . Since in a hierarchy every manager has at most one superior, then $\#R(L_j) \leq \#L_j$. But since the structure is balanced, $\#R(L_j) = \#L_{j+1}$. Therefore, $\#L_j \geq \#L_{j+1}$. ■

Proof of Proposition 5

Proof. The proof is by construction. Start with any given optimal structure, not necessarily strictly balanced nor hierarchic. Say that activity 1 has l_1 levels of managers. Let T_1^j be the type of the manager in the j level of activity 1. The report received by the decision maker is

$$r_1^* = x_1 + \sum_{j=1}^{l_1-1} \alpha \left(d_1^{T_1^{j-1}T_1^j} \right) u_1$$

Thus, the variance of the noise in communication is

$$\left[\sum_{j=1}^{l_1-1} \alpha \left(d_1^{T_1^{j-1}T_1^j} \right) \right]^2 \sigma_u^2$$

Say that activity 2 has $l_2 \neq l_1$ levels. Let T_i^j be the manager on the j -th level of activity i . Find a set of $l_1 - 1$ types of managers $\{A^1, \dots, A^{l_1-1}\}$ such that $\#A^1 = \#T_1^1, \dots, \#A^{l_1-1} = \#T_1^{l_1-1}$ and $\{2\} \in A^1 \cap \dots \cap A^{l_1-1}$. In words, that means finding a set of managers that can exactly replicate for activity 2 the reporting structure of activity 1. Now, replace the old reporting structure of activity 2 by $R(A^1) = R(A^2), \dots, R(A^j) = R(A^{j+1}), \dots, R(A^{l_1-1}) = R(m^*)$. It has to be true that

$$\sum_{j=1}^{l_1-1} \alpha \left(d_1^{T_1^{j-1}T_1^j} \right) = \sum_{j=1}^{l_1-1} \alpha \left(d_1^{A^{j-1}A^j} \right) \geq \sum_{j=1}^{l_2-1} \alpha \left(d_2^{T_2^{j-1}T_2^j} \right)$$

The first equality is true by construction. The inequality has to hold, because otherwise switching to $\{A^1, \dots, A^{l_1-1}\}$ in the second activity would reduce the variance in communication noise, increasing expected profits, which cannot be feasible if this is an optimal

structure. But our labeling of activities 1 and 2 is completely arbitrary and can be reversed, so we have established that

$$\sum_{j=1}^{l_1-1} \alpha \left(d_1^{T_1^{j-1} T_1^j} \right) = \sum_{j=1}^{l_2-1} \alpha \left(d_2^{T_2^{j-1} T_2^j} \right)$$

Therefore, substituting $\{A^1, \dots, A^{l_1-1}\}$ for the original managers in activity 2 will not reduce expected profits, so the resulting structure is still optimal. By applying the same procedure to all other activities $3, \dots, n$, we end up with an optimal structure that has l_1 levels of managers in each activity.

To transform the previous structure into a hierarchy, we use a simple procedure. Without loss of generality, suppose a given manager A reports activity 1 to one manager B and activity 2 to another C . Then, the information about the two activities will follow different paths until they reach the top. But then we can find a manager of type D such that $A \subset D$ and $\#D = \#B$ (this is always possible, since from Proposition 3 we know that in this case $\#B > \#A$), so now A can report only to D about activities 1 and 2. We then keep the path for activity 1 unchanged, while we change the reporting structure of 2 in the same way described above when we constructed the symmetric design. For the same reasons described above, this cannot affect expected profits. Then, one can repeatedly apply this procedure to all managers until everyone reports to only one superior, so one would have an optimal structure that is a hierarchy.

Now we have an optimal structure that is an hierarchy and has the same number of levels in each activity. Suppose now that in some level j not all managers have the same number of subordinates. By adding new managers to that level, one can always reduce the number of subordinates of the managers who have more of them until all managers are equalized (in the limit, one can add managers until all in that level have only one subordinate). That completes the proof. ■

References

- [1] **Aghion, P., and Jean Tirole, (1997)**, “Formal and Real Authority in Organizations”, *Journal of Political Economy*, 105, 1-29.
- [2] **Arrow, K.J., (1982)**, “Team Theory and Decentralized Resource Allocation: an Example”, technical report no. 371, *Institute for Mathematical Studies in the Social Sciences*, Stanford University.
- [3] **Baker, G., Gibbons, R., and Kevin J. Murphy, (1999)**, “Informal Authority in Organizations”, *Journal of Law, Economics and Organizations*, 15, 56-73.
- [4] **Becker, G., and Kevin M. Murphy, (1992)**, “The Division of Labor, Coordination Costs, and Knowledge ”, *Quarterly Journal of Economics*, v107, n4: 1137-60
- [5] **Beckmann, M.J., (1960)**, “Some Aspects of Returns to Scale in Business Administration”, *Quarterly Journal of Economics*, 464-471.
- [6] **Bolton, P., and Mathias Dewatripont, (1994)**, “The Firm as a Communication Network”, *Quarterly Journal of Economics*, 109, 809-839.
- [7] **Garicano, L., (2000)**, “Hierarchies and the Organization of Knowledge in Production”, *Journal of Political Economy*.
- [8] **Harris, M. and Artur Raviv, (1999)**, “Organization Design”, manuscript, The University of Chicago.
- [9] **Hart, O. and John Moore, (1999)**, “On the Design of Hierarchies: Coordination versus Specialization”, manuscript.
- [10] **Heath, C. and Nancy Staundenmayer, (2000)**, “Coordination Neglect: How Lay Theories of Organizing Complicate Coordination in Organizations ”, manuscript, Duke University.

- [11] **Holmstrom, B., (1979)**, “Moral Hazard and Observability”, *Bell Journal of Economics*, v10, n1: 74-91.
- [12] **Keren, M., and David Levhari, (1983)**, “The Internal Organization of the Firm and the Shape of Average Costs”, *Bell Journal of Economics*, 474-486.
- [13] **Marschak, J. and Roy Radner, (1972)**, “The Economic Theory of Teams”.
- [14] **Möbius, M. (1999)**, “The Evolution of Work”, manuscript, MIT.
- [15] **Prat, A., (1997)**, “Hierarchies of Processors with Endogenous Capacity”, *Journal of Economic Theory*, 77, 214-222.
- [16] **Radner, R., (1992)**, “Hierarchy: The Economics of Managing”, *Journal of Economic Literature*, 30, 1382-1415.
- [17] **Radner, R., (1993)**, “The Organization of Decentralized Information Processing”, *Econometrica*, 61, 1109-1146.
- [18] **Rosen, S., (1982)**, “Authority, Control, and the Distribution of Earnings”, *Bell Journal of Economics*, 311-323.
- [19] **Sah, R.K., and Joseph E. Stiglitz, (1986)**, “The Architecture of Economic Systems: Hierarchies and Polyarchies”, *American Economic Review*, 76, 716-727.
- [20] **Sah, R.K., and Joseph E. Stiglitz, (1988)**, “Committees, Hierarchies and Polyarchies”, *Economic Journal*, 98, 451-470.
- [21] **Van Zandt, T., (1999a)**, “Decentralized Information Processing in the Theory of Organizations”, in Sertel, M.R., (ed) : *Economic Behaviour and Design, Contemporary Economic Issues*, vol.4, Proceedings of the Eleventh World Congress of the International Economic Association, 125-160.

- [22] **Van Zandt, T. , (1999b)**, “Real Time Decentralized Information Processing as a Model of Organizations with Boundedly Rational Agents”, *Review of Economic Studies*, 66, 633-658.
- [23] **Vayanos, D., (1999)**, “The Decentralization of Information Processing in the Presence of Synergies”, manuscript, MIT.

Figure 1

A hierarchic but not strictly balanced structure

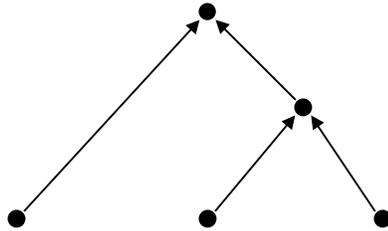


Figure 2

A strictly balanced but not hierarchic structure

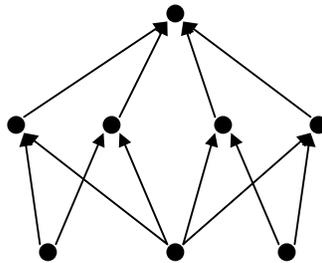


Figure 3

Example 1 - a strictly balanced hierarchy

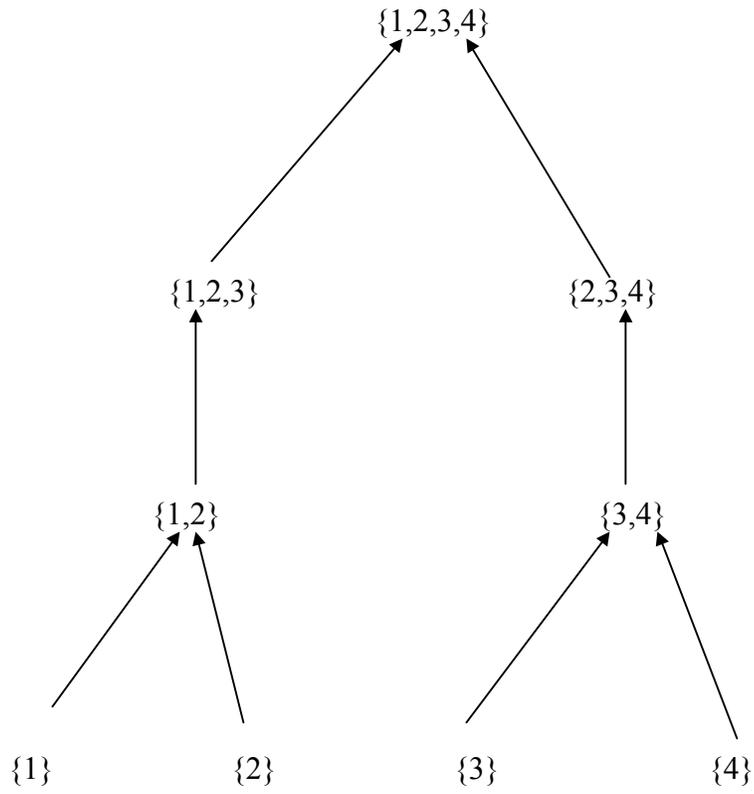


Figure 4

Example 2 - a non-hierarchic, non-strictly balanced optimal structure

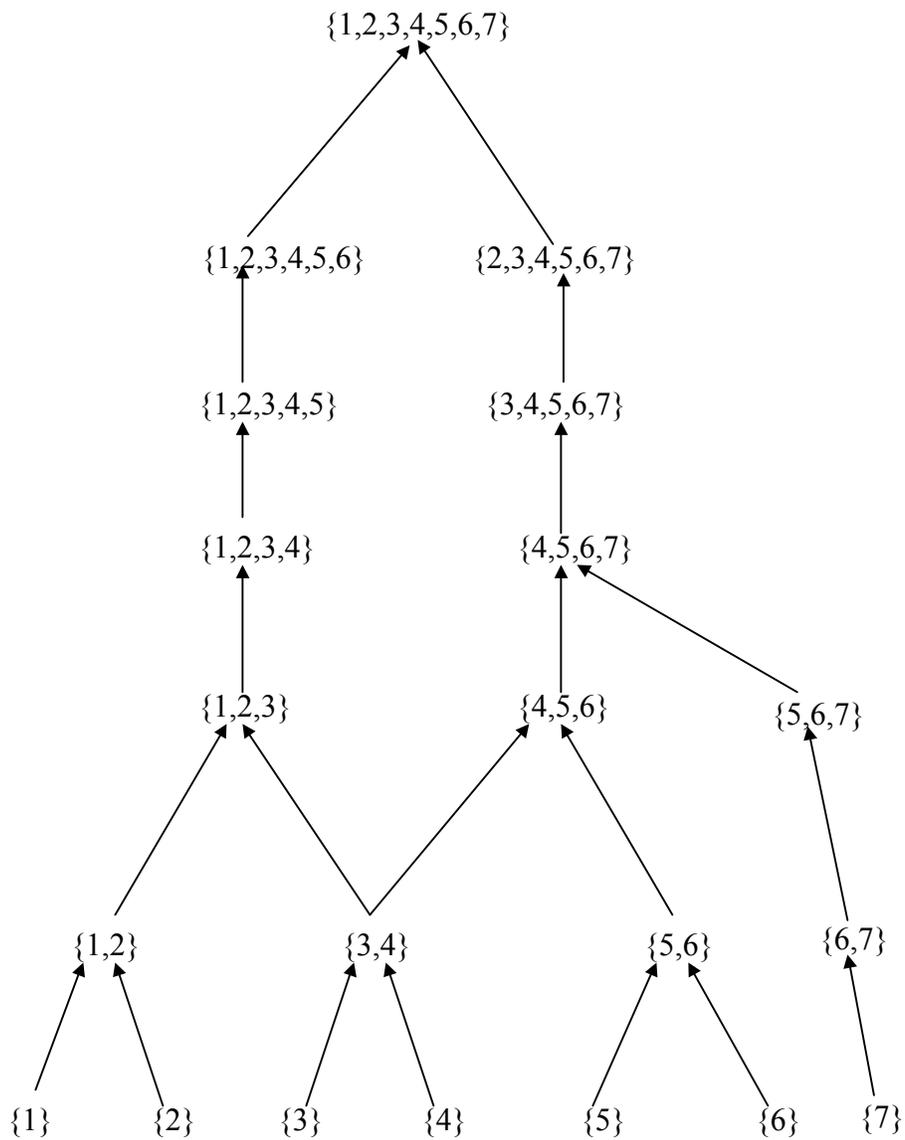


Figure 5

A strictly balanced but not pyramidal structure

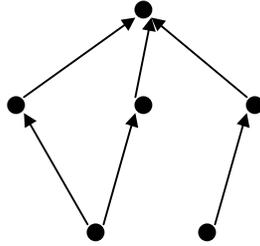


Figure 6

A pyramidal but not hierarchic structure

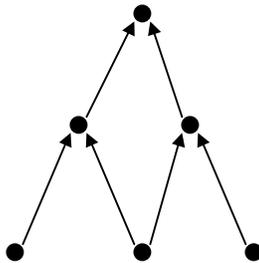


Figure 7
Example 3 - a matrix structure

