Who Gets to the Top?
Generalists versus Specialists in Organizations*

Daniel Ferreira  
London School of Economics and CEPR

Raaj K. Sah  
University of Chicago

January 24, 2009

Abstract

We propose a model of communication in organizations in which the quality of communication depends on the match between senders and receivers of messages. The model allows for two dimensions of knowledge: breadth and depth. Specialists have deep knowledge of few areas while generalists have superficial knowledge of many areas. Generalists are useful because they can communicate with many different specialists. As a consequence, optimal organizational structures are such that generalists are at the top and specialists are at the bottom. Our model has implications for organization design, the optimal degree of centralization of decision making, and the depth of hierarchies. In particular, we show that an increase in the complexity of the environment together with improvements in communication technology lead to a decrease in specialization at the top.

*We thank Pierre-André Chiappori, Luis Garicano, Canice Prendergast, and many seminar and conference participants for their comments. E-mail addresses: d.ferreira@lse.ac.uk, sah@uchicago.edu.
1 Introduction

The purpose of this paper is to study imperfect communication as a source of coordination costs in organizations. There is compelling evidence from the psychology and organizational behavior literatures that individuals who have expertise in different areas find it difficult to communicate among themselves (see Heath and Staudenmayer, 2000, and the references therein). Accordingly, we model imperfect communication as arising from heterogeneity in the knowledge of members of an information-processing organization. We derive implications for the trade-off between specialization of knowledge and communication costs, the role of top and middle managers, and the optimal design of hierarchies.

The novelty in our approach is the idea that the quality of communication depends on the match between the sender and the receiver of a message. When individuals endowed with different types of knowledge communicate with each other, some misunderstandings are unavoidable. Thus, efficient information-processing organizations must care not only about the allocation of different individuals to different hierarchical ranks, but also about the match between superiors and subordinates.

In our model, knowledge has two dimensions: breadth and depth. Due to limited cognitive capabilities, if someone invests in acquiring deep knowledge about some things, he must sacrifice the breadth of his knowledge, and vice-versa. Therefore, the cognitive impossibility of knowing too much about too many things implies that some people choose to learn much about few things, while others learn a little about many things. We call the former a specialist and the latter a generalist.

Specialization is detrimental to communication because it increases the expertise heterogeneity across individuals. We show that the conflict between specialization and communication implies that there is economic value to employing generalist managers. As a consequence, information flows from specialists to generalists. In an optimal organizational structure, generalists are at the top of hierarchies while specialists are at the bottom. This intuitive result is in contrast with the theoretical literature on the assignment of heterogeneous individuals to different hierarchical levels, which emphasizes the role of talent or ability as the main determinant of rank (for example, see Rosen, 1982).
Casual and systematic evidence corroborates this implication. More MBAs head large business corporations than individuals with any other background: Bertrand and Schoar (2003) find that 40% of all CEOs in their sample have completed an MBA.\textsuperscript{1} This is so even for those companies that have relatively narrow lines of business, such as pharmaceuticals and steel.\textsuperscript{2} Those who have worked for many years as specialists (for example, as engineers and scientists) try, on their way to the top, to reeducate, or at least to repackage, themselves as generalists. Such patterns have become more pronounced since the rise of large business corporations, beginning roughly in the 18th century, and even more pronounced since the rise of multinational corporations, beginning roughly in the 19th century.

Our model implies a sharp trade-off between centralized and decentralized decision-making processes. In centralized organizations, the top decision maker has coarse information about many activities, while in decentralized ones the decision maker has precise information about few activities, but no information about most of them. As a consequence of this trade-off, we find that decreasing specialization at the top is a consequence of increasing environmental complexity and improvements in communication technology.

Our model endogenously generates pyramid-shaped 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-level managers and then report it to higher-level managers. Middle managers are semi-specialized workers who act as translators of information sent by production workers to top decision makers. Therefore, the communication technology determines the optimal number of layers in a hierarchy.

We are among the first to formalize the phenomenon of communication across individuals in which the heterogeneity of individuals is explicit.\textsuperscript{3} This is the main innovation in this paper. While information processing in organizations is a topic that has already been studied at some length by organizational economists, the process of communicating information, and

\textsuperscript{1}These numbers are even higher for younger CEOs.

\textsuperscript{2}There are some nearly consistent exceptions in public organizations, such as heads of geological survey organizations and heads of weather bureaus.

\textsuperscript{3}We recognize that this formalization could be made more comprehensive in many ways, as it will be seen later. For a recent model of social learning in which the quality of communication also depends on the match between heterogeneous senders and receivers, see Niehaus (2008).
in particular its relationship to knowledge heterogeneity, has not received as much attention.

In order to understand the effects of imperfect communication, we should first isolate them 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., Holmström, 1979) focuses on imperfect preference alignment when information transmission is perfect. However, this separation is not without costs and the integration of the two approaches is still a promising topic for future research.

The structure of the paper is as follows. After discussing the related literature, we introduce the basic model, notation and assumptions in Section 2. Section 3 discusses the main results and provides some examples and applications. Section 4 concludes with a discussion of some possible extensions.

**Related literature**

The framework we present in this paper is a step towards better modeling and understanding imperfect communication in organizations. A recent paper by Cremer, Garicano, and Prat (2007) takes a similar approach. They too focus explicitly on issues related to imperfect communication in organizations, rather than on pure information processing problems. Their problem is to find the optimal organizational code (or language) that minimizes communication costs. As in this paper, they derive many implications for organizational design.

Our paper is different mainly due to our focus on activity-specific codes, rather than on organization-specific codes. In our paper, we take organizational codes as givens and assume that individuals specializing in different areas (engineering, accounting, etc.) speak different languages. The development of an organizational code (as in Cremer et al.) may help them better communicate with each other, but it is unlikely to eliminate communication costs completely.¹

Another important recent contribution is the work of Dessein and Santos (2006). In their

¹Although our paper is in many aspects very different from Cremer et al (2007), an interesting related result in both papers is the importance of higher-level managers in hierarchies as translators of information across individuals who speak different languages.
model, the optimal organization trades off coordination (i.e. exploiting synergies across tasks) versus adaptation (i.e. better use of local information). They highlight that improvements in the communication technology may lead to decreasing specialization because they facilitate the use of information in organizations. This result is similar to our finding that managerial specialization should decrease as communication technology improves. They also find that increases in environmental complexity reduce specialization. Thus, our paper shares some of their main results, but differs mainly due to two reasons: (i) in our model, the quality of communication varies according to the match between senders and receivers in terms of their knowledge specialization, and (ii) we focus on the design of hierarchies, i.e. the assignment of different types of workers and managers to different hierarchical levels. These modeling choices allow us to focus on different aspects of the trade-off between specialization and coordination.

Many other papers that model information-processing organizations assume homogeneous individuals (see for example Sah and Stiglitz, 1986 and 1988; Radner, 1992 and 1993; Bolton and Dewatripont, 1994; Van Zandt, 1999). Prat (1997) is an important exception. In his model, managers with higher information-processing capacities are assigned to higher hierarchical levels. In all these cases, however, the models were not constructed to answer the questions we address in this paper, but mainly to stress the advantages of parallel over sequential processing.5

Some papers on the assignment of heterogeneous individuals to different hierarchical levels conclude that individuals at higher levels are more talented (Rosen, 1982; Prat, 1997) or able to solve more difficult problems (Garicano, 2000). In our model, 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 than another one.

There are many other papers that deal with some related issues. Harris and Raviv (2002) develop a model to explain the choice between hierarchies and matrix structures, in which each manager is capable of detecting and coordinating interactions only within

---

5 Sah and Stiglitz are concerned with minimizing different types of errors in decision making, while Radner, Prat and Van Zandt are concerned with reducing delay in information processing. In Bolton and Dewatripont’s (1994) paper, specialization makes individuals better processors of information.
his limited area of expertise. In Vayanos’s (2003) model of information processing with synergies or interactions between areas, top managers process more aggregated information, while managers at the bottom process only local information. Geneakoplos and Milgrom (1991) develop a theory of organizations based on managers’ limited attention. For models that rely on the “authority approach,” meaning that subordinates only act when coordinators tell them to, see Hart and Moore (2005) and Castanheira and Leppämäki (2004). For theories of the trade-off between communication and delegation based on agency considerations, see Jensen and Meckling (1992), Aghion and Tirole (1997), Baker, Gibbons and Murphy (1999), Stein (2002), and Dessein (2002).

2 Setup

2.1 Project Attributes and Areas of Expertise

We consider an organization that accepts or rejects projects (i.e. ideas and proposals). A project has \( n \) attributes. Attributes are represented by i.i.d. random variables \( X_1, X_2, ..., X_n \) that are normally distributed with mean \( \mu \) and variance \( \sigma^2 \). We denote the realization of a random variable \( X_i \) by its lowercase counterpart \( x_i \).

The independence assumption implies that there are no informational synergies between two attributes, i.e. knowledge of one attribute is not directly useful for another attribute. Thus, synergies can only arise due to technological reasons, i.e. how two attributes are combined to generate profits.

Each attribute is associated with one area of expertise. Denote the set of all areas of expertise by \( N = \{1, 2, ..., n\} \). Let \( A \) be a non-empty subset of \( N \). We call an individual whose expertise set is \( A \) a generalist-\( A \). If \( A = N \), we call him a supergeneralist. If \( A \) is a singleton, i.e. \( A = \{i\} \), a generalist-\( A \) is a specialist in \( i \), in which case we call him a superspecialist-\( i \). In between these two extremes, there are \( \sum_{j=2}^{n} \binom{n}{j} \) types of individuals representing various other levels of expertise specialization. We denote by \( \#A \) the cardinality of a generalist-\( A \), which is the number of elements in \( A \). In some cases, it will be convenient to characterize an individual by its cardinality. For example, we will use the term generalist-3
to refer to any individual whose expertise set has cardinality 3.

2.2 Communication Technology

Each individual in this organization may send and receive reports about different attributes. We assume that communication is imperfect. As Arrow (1982) 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 reports depends on the expertise gap between the sender and the receiver. If two individuals with the same expertise set communicate with each other, no communication error is introduced. However, if someone receives a report from a person with a narrower expertise set, the former can only get a noisy signal of the message sent.\(^6\)

There are arguably two types of errors introduced when two individuals communicate with each other: one is generated by the sender of the message and the other is generated by the receiver.\(^7\)

The receiver cannot fully understand the message because of his cognitive limitations. We assume that the more the receiver knows about the nature of the attribute being reported, the better he can understand the reports, therefore the smaller is the error he introduces in the communication process.

The error introduced by the sender arises from his inability to perceive the limitations of the receiver. Because the sender is uncertain about which type of information the receiver is capable of understanding, he might use ineffective communication strategies, such as using difficult concepts or lines of exposition that are not quite suited for the receiver.\(^8\) Thus, we

\(^6\)Our model is about imperfect communication rather than limited capacity of processing information, implying that one cannot alleviate miscommunication problems simply by using parallel processing or employing individuals with higher processing capabilities.

\(^7\)“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).

\(^8\)Heath and Staudenmayer (2000) call these problems “inadequate communication and insufficient translation”; that is, “people do not communicate well in general and they fail to realize the additional problems of translating across differentiated specialists.”
assume that the more “similar” (in a sense to be defined later) the sender and the receiver are, the smaller is the error introduced by the sender in the communication process.

We pool these two types of error into one by assuming that the communication error depends on the expertise gap between the sender and the receiver. To make this idea clear, we need to define a measure of distance in expertise. Consider two different individuals, A and B. We want to create a measure of distance in expertise with respect to attribute $i$. We assume that, if $i \in A \cap B$, distance is given by:

$$d_{iAB} = |\#A - \#B|,$$

where $\#X$ is the number of elements in $X$. If $i \notin A \cap B$, we assign an arbitrarily high value to $d_{iAB}$, which we informally denote by $d_{iAB} = \infty$.

According to the definition in (1), if both individuals have some expertise with respect to $i$ (i.e. if $i \in A \cap B$), distance increases with the difference in the degree of specialization between the two. If $i \notin A \cap B$, either the sender or the receiver (or both) are completely unable to process information about $i$.

This distance function corresponds to a situation in which all expertise areas $j \in N$ are independent in a strong sense. For example, let $A = \{1, 2\}$, $B = \{1, 3\}$ and $C = \{1, 2, 4, 5\}$. The distance in expertise with respect to attribute 1 is 2 both between $A$ and $C$ and between $B$ and $C$, despite the fact that $A$ has more in common with $C$ than $B$ does. This assumption implies that common knowledge of attribute 2 does not help $A$ and $C$ when they send or receive messages concerning attribute 1.

This strong independence assumption is realistic in many cases, especially when expertise areas are defined very broadly (e.g. “marketing” or “accounting,” rather than “ability to read balance sheets” or “understanding sales statistics”). At the other extreme, we could also consider a measure of distance in which expertise areas are complements in the communication process:

$$d_{i^cAB} = \max \{\#A, \#B\} - \#(A \cap B).$$

It is straightforward to check that function (2) is indeed a distance in the mathematical sense. In the same example as before, this alternative distance function implies that the distance in expertise with respect to attribute 1 is 2 between $A$ and $C$ and 3 between $B$ and
Because both cases appear extreme, we expect reality to lie somewhere in the middle. Fortunately, it does not matter for the results which distance function we use, thus we choose to work with the simpler version (1) in order to simplify the algebra and the proofs. After making the appropriate modifications, all results that follow remain intact if we use the alternative function (2) instead, although the proofs are longer and more tedious.

We now model the communication process. Formally, let $s^A_i$ be the report about attribute $i$ sent by a type-$A$ person. If a type-$B$ person is the receiver, his reading of the report is

$$r^A_{i} = s^A_i + \alpha \left( d^A_{i}u \right),$$

where $u_i \sim N(0, \sigma^2_u)$. We call the function $\alpha \left( d^A_{i} \right)$ the intensity of communication error. We assume that the intensity of communication error is positively related to the distance in expertise. Formally, we assume that $\alpha : \{0, 1, \ldots, \infty\} \rightarrow [0, \infty)$ is a strictly increasing function, with boundary conditions $\alpha (0) = 0$ and $\alpha (\infty) = \infty$. Notice that the match between the sender and the receiver matters: the function $\alpha \left( d^A_{i} \right)$ depends on the identity of the sender, $A$, and on the identity of the receiver, $B$.

Equation (3) can be easily interpreted in terms of aggregation of information. While individual $A$ knows two bits of information, $s^A_i$ and $\alpha \left( d^A_{i} \right) u_i$, individual $B$ can only process the aggregated signal $r^A_{i}$, which is the sum of the information bits known by $A$.

Notice that the communication error $\alpha \left( d^A_{i} \right) u_i$ does not depend on anything specific to the sender nor the receiver, except for their types. Therefore, two different type-$A$ individuals generate exactly the same error when reporting to $B$ (notice that $u_i$ does not vary across individuals, only across areas of expertise). Intuitively, individuals endowed with the same expertise set should agree among themselves. As a consequence, they should cause the same type of communication error.\footnote{A perhaps more realistic formulation would allow for the intensity of communication error to depend also on the direction of communication. For example, it is natural to assume that $\alpha \left( d^A_{i} \right) = 0$ if $\#A < \#B$. All results that follow remain unchanged under such an assumption.}

\footnote{Information aggregation is the origin of imperfect communication in Vayanos’s (2003) paper. Differently from our model, however, the ability to process disaggregated information in his model does not depend on the individuals’ expertise set.}

\footnote{Allowing communication errors to vary across individuals of the same type would amount to allowing}
Finally, when an individual acquires information about an attribute directly, we say that he receives a signal from nature. We assume that individuals can only report the signals that they have received either from other individuals or from nature. That is, individuals cannot choose which signals they send. This simplification is sensible; because we will assume that all individuals share the same objective function, there will be no meaningful strategic communication issues in our setup.\footnote{Individuals might intentionally distort messages in equilibrium if their preferences do not coincide (see Crawford and Sobel, 1982).}

### 2.3 Organization Design

We define an organization as a group of individuals who have access to the set of all \( n \) attributes of a given project. By access we mean the ability to perform of a task that is related to a subset of \( N \), such as sending or receiving reports about some of the project’s attributes. The organization must also make decisions that affect the utility of all members.

The organization design problem consists of deciding (i) how many individuals will join the organization, (ii) which types of expertise the members will have, (iii) who will report to whom, and (iv) who will have formal authority over decisions. We call each possible solution to the organization design problem a \textit{structure}.

**Definition 1** An organizational \textbf{structure}, denoted by \( s = (M, R, d) \), consists of

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

- A reporting correspondence \( R : M \rightarrow M \), such that \( R(m_i) \) is the set of members who receive reports from member \( m_i \) (if some member \( m_k \) does not send reports, \( R(m_k) = \emptyset \)).

- A decision maker \( d \in M \), which is the member who has formal authority over decisions.

This definition uses the fact that all individuals are completely characterized by their types, which are subsets of \( N \). We denote by \( S \) the set of all feasible organizational structures.
2.4 Organizational Goals

Our framework is quite general and can be applied to many different organizational problems. In order to highlight its usefulness while still keeping the model transparent, we focus on a single simple decision problem, in which the organization has to decide whether or not to undertake a single project.

The project yields direct costs. If the project is undertaken \((p = 1)\), there is a fixed cost of \(c \geq 0\). Otherwise \((p = 0)\) there are no costs.

Profits (or payoffs) depend on the value of all \(n\) attributes. We adopt a simple linear technology in which profits depend on the sum of the values of all the project’s attributes. The ex post value of profits is given by

\[
\pi(p) = \begin{cases} 
\sum_{i=1}^{n} x_i - c, & \text{if } p = 1 \\
0, & \text{if } p = 0 
\end{cases} 
\]  

(4)

Notice that information about attributes has a synergistic element: the more one knows about each attribute, the better informed the project implementation decision is.

We assume that there must be at least one superspecialist involved in the initial evaluation of attribute \(i\) in order to generate a reading of \(X_i\), and that generalists cannot get readings from nature. Although this assumption is not crucial, it simplifies the analysis by reducing the number of cases we need to consider. It is a natural assumption if specific knowledge is a by-product of the production process (see Jensen and Meckling, 1992).

2.5 Timing

The timing of the decision process is as follows. Each superspecialist-\(i\) observes the realized value \(x_i\) of the attribute \(i\). After that, they send reports to other members of the organization. Anyone who receives a report can also send reports to other members. After all reporting activity is complete, one member of the organization (the decision maker) decides whether or not to undertake the project.
3 Finding Optimal Structures

We now address the problem of designing an optimal organizational structure. We assume that every member acts in the interest of the team (no agency problems).

Let $I$ be the decision maker’s information set. The decision maker maximizes expected profits, conditional on his information:

$$\Pi(I) = \max_{p \in \{0, 1\}} E[\pi(p) | I].$$  \hfill (5)

The organization’s problem is to find a structure $s \in S$ that maximizes expected profits in the ex ante stage:

$$\max_{s \in S} E[\Pi(I) | s].$$  \hfill (6)

According to Definition 1, the organization design problem consists of deciding (i) how many individuals will join the organization, (ii) which types of knowledge the members will have, (iii) who will report to whom, and (iv) who will have the authority to decide. The goal is to maximize ex ante expected profits.

In order to compare different structures, we also 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 more members and some costs of reporting. However, for the sake of simplicity, we choose to ignore these costs, but in most cases we will assume away structures that have redundant members, i.e. members who do not convey useful information or do not improve upon the existing amount of information. This amounts to a lexicographic criterion: first find the structures that minimize communication costs and then, among them, pick the ones that minimize the number of members and reports.

We ignore other types of costs such as the costs of acquiring knowledge and the costs of delay. We postpone to section 4 a discussion about the effects of allowing for some of these costs on our results.
3.1 Flat Hierarchies

In this section, we restrict attention to a world in which only \( n + 1 \) types of expertise are available: superspecialists in each area of expertise and supergeneralists (i.e. generalists that have some expertise in each of the \( n \) areas). We consider this case first for simplicity and postpone the discussion of the general case to the next sections. Much of the intuition can be gained by considering this simpler case and the results in the following sections are straightforward generalizations of the results in this one.

In this world, members of the organization can be of two different kinds: superspecialists, who can receive reports both from nature and from other members, and supergeneralists (or managers), who receive reports only from other members. Accordingly, we define a managerial structure as a structure in which the decision maker only receives reports from other managers or specialists, but never from nature. A structure in which the decision maker is also a specialist (i.e. he receives reports from nature) is called an specialist-managed structure.

The following lemmas are straightforward. In the proofs, we adopt the lexicographic criterion discussed above and assume away redundant members and reports.

**Lemma 1**

_Superspecialists never receive reports._

**Proof.** If anyone reports to a superspecialist-\( j \) about attribute \( i \neq j \), he will receive a signal of infinite variance because \( \alpha(\infty) = \infty \). Therefore, this signal is useless for a superspecialist-\( j \), either for decision making or for further reporting. If a superspecialist-\( i \) reports to another superspecialist-\( i \), both will accurately know the realized state \( x_i \), but the same information could have been acquired from the initial evaluation of \( x_i \) without reporting, thus one of the specialists is redundant.

**Lemma 2**

_Supergeneralists never send reports._

**Proof.** Supergeneralists should not report to superspecialists, due to Lemma 1. Since supergeneralists cannot get readings from nature, they can only send reports if they receive reports from other members. Therefore, a supergeneralist who receives reports from another supergeneralist could have obtained the same information by receiving reports from the
latter’s sources, economizing the use of at least one redundant member, without loss of information. Thus, supergeneralists should not report to other supergeneralists. ■

These two lemmas jointly imply that we need to consider only communication between supergeneralists and superspecialists. The distance in expertise between any superspecialist and the supergeneralist is $d_i^{iN} = n - 1$. Because we will keep $n$ fixed throughout this section, in order to simplify notation we define $\alpha \equiv \alpha (n - 1)$.

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

Proposition 1 below and its corollaries summarize the answers to these questions. We sketch the main steps of the argument in the text and refer the reader interested in the more tedious algebra to the Appendix. Mainly for expositional simplicity, here we assume that $\mu = 0$ and $c = 0$. In the Appendix, we prove Proposition 1 under the more general assumptions of $\mu \in \mathbb{R}$ and $c \geq 0$, confirming that our simplified version in this section is without loss of generality.

Suppose first that a superspecialist is the decision maker. By Lemma 1, we know that in such a case the decision maker receives no reports and has to decide whether to undertake the project based only on the information he receives from nature.

Let $s_s$ denote the specialist-managed structure in which a superspecialist-$i$ is the decision maker. The decision maker will undertake the project if and only if (recall that $\mu = c = 0$)

$$E[\pi \mid X_i = x_i] = x_i \geq 0.$$  

That is, he will undertake the project when its expected profits are nonnegative, given his private information on $X_i$. The value function of his maximization problem is

$$\Pi(X_i) = \max \{X_i, 0\}.$$  

Therefore, the ex ante expected profit is (recall that attributes are normally distributed)

$$E[\Pi(X_i) \mid s_s] = \frac{1}{2}E[X_i \mid X_i \geq 0].$$  

Consider now an organization in which a supergeneralist is the decision maker (a managerial structure—$s_m$). In this case, receiving reports from others never makes the decision
maker less informed and sometimes makes him better informed. Therefore, he should combine the signals (reports) that he receives with his prior knowledge about the probability distributions in a Bayesian manner. If a supergeneralist is the decision maker, he receives reports on the \( n \) attributes and chooses to undertake the project if and only if

\[
E[\pi \mid (R_1, \ldots, R_n) = (r_1, \ldots, r_n)] = \beta \sum_{i=1}^{n} r_i \geq 0,
\]

(10)

where \( r_i \) is the reading the decision maker gets on attribute \( i \) and \( \beta = \frac{\sigma^2}{\sigma^2 + \alpha^2 \sigma_u^2} \) is the optimal weight that he assigns to the report about each attribute, as in standard signal-extraction problems.\(^{13}\) By Lemma 2, in an optimal structure the supergeneralist can only receive reports from specialists.

The ex ante expected profit is

\[
E[\Pi(R_1, \ldots, R_n) \mid s_m] = \frac{1}{2} E\left[ \sum_{i=1}^{n} X_i \mid \beta \sum_{i=1}^{n} R_i \geq 0 \right].
\]

(11)

Therefore, there can only be two possible structures. In a specialist-managed structure, expected profits are given by (9), one superspecialist is the decision maker, there are \( n \) superspecialists but no generalists, and there is no reporting. In a managerial structure, expected profits are given by (11), one supergeneralist is the decision maker, there are \( n \) superspecialists and one supergeneralist, and all superspecialists report to the supergeneralist. Now we shall compare the two possible structures.

Proposition 1 (The Demand for Managers) The managerial structure is superior to the specialist-managed structure if and only if

\[
1 + \frac{\alpha^2 \sigma_u^2}{\sigma^2} < n.
\]

(12)

Proof. See the Appendix. \( \blacksquare \)

This proposition highlights the main trade-off in our model. On the right-hand side of condition (12) we have \( n \), which is the number of different project attributes that influence profits. Having many attributes that influence profits means that there are many synergies among different activities, and that using all relevant information in order to exploit these

\(^{13}\)See the proof of Proposition 1 in the Appendix for the details.
synergies is very valuable. Therefore, the greater the number of attributes, the more important managed coordination is. Thus, the right-hand side of this condition represents the value of coordinating synergistic activities by generalist managers.

The left-hand side of (12) proxies for communication costs, which depend on the distance in expertise between supergeneralists and superspecialists ($\alpha$), on the communication noise ($\sigma^2_u$), and on the prior uncertainty or attribute complexity ($\sigma^2$). Condition (12) thus represents a trade-off between synergistic coordination and communication costs. Managerial structures (i.e., centralized decision-making by generalist managers) are optimal if communication costs are low and synergistic gains are high. On the other hand, specialist-managed structures are optimal if communication costs are high and synergistic gains are low.

The following results are straightforward.

**Corollary 1** If attribute complexity (i.e., prior uncertainty $\sigma^2$) is sufficiently high, it is optimal to have a supergeneralist manager (everything else constant).

The intuition behind this result is that, with more ex ante uncertainty, knowledge of the prior distribution of the attributes becomes less useful. Therefore, when constructing their posterior distributions, Bayesian generalist managers give less weight to the prior distribution and more weight to the reports that they receive. Superspecialists, however, do not receive any reports (see Lemma 1), so they have to make decisions based only on their priors. Therefore, superspecialists’ decision rules do not change after increases in prior uncertainty, while the decision rule of the supergeneralist optimally responds to such increases by giving less weight to the prior distribution.

**Corollary 2** If the communication technology is sufficiently precise (low $\sigma^2_u$), it is optimal to have a supergeneralist 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.

These two corollaries may be interpreted as saying that any technological innovation that simultaneously increases production complexity or unpredictability ($\sigma^2$ increases) and improves the communication technology ($\sigma^2_u$ decreases) increases the demand for managers.
The “skill-biased technological change” literature has shown that the ratio of production to non-production workers in manufacturing has been decreasing over time.\textsuperscript{14} The increasing relative demand for non-production workers (which in our model are generalist managers) is compatible with technological changes that facilitate communication and increase complexity. Such technological changes are thus compatible with a rise in managerialism.

**Corollary 3** The number of project attributes ($n$) has an ambiguous effect on the demand for managers: On the one hand, it increases the value of coordinating synergistic activities ($n$ increases), on the other hand, it makes coordination more difficult due to increasing communication costs ($\alpha$ increases).

Coordination always become more valuable with more project attributes ($n$). Thus, at first glance it appears that an increase in $n$ should lead to the choice of a managerial structure. However, one must recall that with more attributes, the supergeneralist’s expertise is broader but more superficial, implying that his ability to communicate with different specialists is hampered. The intuition for this result is straightforward: imperfect communication limits the manager’s span of control. Notice also that, under our boundary conditions for $\alpha (.)$, if $n$ is large enough, the managerial structure is never optimal, which is consistent with the idea of a maximum span of control.

The following corollary is just a restatement of Proposition 1.

**Corollary 4** (Optimality of Flat Hierarchies) If $1 + \frac{\alpha^2 \sigma^2}{\sigma^2} \leq n$, an optimal organizational structure $s^* = \langle M^*, R^*, d^* \rangle$ has $n + 1$ members such that

1. There are $n$ superspecialists, one for each attribute, and one supergeneralist:

   \[
   M^* = \{m_1, \ldots, m_{n+1}\};
   \]
   \[
   m_i = \{i\}, \text{for all } i = 1, \ldots, n;
   \]
   \[
   m_{n+1} = N.
   \]

2. All superspecialists report to the supergeneralist:

\[ R^*(m_i) = \begin{cases} m_{n+1}, & \text{if } i = 1, \ldots, n; \\ \emptyset, & \text{otherwise.} \end{cases} \]

3. The supergeneralist is the decision maker:

\[ d^* = m_{n+1}. \]

We have then showed that there is only one possible type of hierarchy in this world, the so-called flat hierarchy: all workers report directly to a single manager. In other words, there are no middle managers.

3.2 Comparing Flat Hierarchies

In the previous section, we have characterized the conditions under which it would be optimal to have a flat hierarchy with a supergeneralist as the top manager. In this section, we assume that all types of expertise are feasible. That is, individuals can be either superspecialists or one of the \( \sum_{j=2}^{n} \binom{n}{j} \) types of generalists. Our goal now is to compare different flat hierarchies with different types of generalists at the top.

Let a generalist-\( A \) be the decision maker, receiving reports from \( n \) specialists. Define \( k_A \equiv \# A \leq n \). Without loss of generality, let \( A \) be equal to \( \{1, 2, \ldots, k_A\} \). Now, the problem is to find the degree of specialization of the top manager \( k_A \) that maximizes ex ante expected profits. It is important to understand the nature of the relevant trade-off here: broadening the knowledge of the top manager (increasing \( k_A \)) allows him to receive readings from a larger number of superspecialists, but at a cost of reducing the precision of his readings (because \( d_i^{k_A} \) increases, \( \forall i \in A \)). We state this result formally as a proposition.

**Proposition 2 (Comparison of Flat Hierarchies)** Let \( A \) and \( B \) be any two generalists with cardinalities \( k_A \) 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

\[
\frac{k_A}{k_B} < \frac{\sigma^2 + \alpha (k_A - 1)^2 \sigma_u^2}{\sigma^2 + \alpha (k_B - 1)^2 \sigma_u^2}.
\]

(13)
Proof. See the Appendix. ■

Notice that condition (13) is just a generalization of condition (12). To see this, notice that the cardinality of a supergeneralist is $n$, while the cardinality of a superspecialist is 1. Since $\alpha_i = 0$, substituting $n$ for $k_A$, 1 for $k_B$, $\alpha$ for $\alpha (k_A - 1)^2$ and 0 for $\alpha (k_B - 1)^2$, we arrive at condition (12).

Proposition 2 allows us to characterize the optimal degree of specialization at the top (when restricted to flat hierarchies):

Corollary 5 (Optimal Flat Hierarchies) In an optimal structure, the decision-maker has cardinality $k^* \in \{1, \ldots, n\}$ if and only if

$$k^* \geq \frac{\sigma^2 + \alpha (k^* - 1)^2 \sigma_u^2 k}{\sigma^2 + \alpha (k - 1)^2 \sigma_n^2}$$

(14)

for all $k \in \{1, \ldots, n\}$.

That a $k^*$ exists follows from the continuity of the objective function and the compactness of $\{1, \ldots, n\}$, and uniqueness of $k^*$ follows immediately from (14). Because the problem is neither concave nor differentiable, computing $k^*$ requires pair-wise comparisons of all elements in $\{1, \ldots, n\}$. However, unambiguous comparative statics are possible even without finding the optimal $k^*$.

Proposition 3 The optimal $k^*$ is non-decreasing in $\sigma^2$ and non-increasing in $\sigma_u^2$.

Proof. See the Appendix. ■

Proposition 3 generalizes the results in corollaries 1 and 2. It shows that the degree of specialization at the top decreases with innovations that increase production complexity and improve the communication technology. There is some evidence that the degree of task specialization in some firms has been decreasing (increasing multi-tasking) and that this decrease is linked to improvements in information and communication technologies. See Bresnahan, Brynjolfsson and Hitt (2002) and Caroli and Van Reenen (2000).15

15See Möbius and Schoenle (2007) for an alternative theory of decreasing specialization.


3.3 Middle-Managers

A natural question is whether the availability of middle managers improves upon the flat hierarchical structures we have studied so far. In this section, we start from a hierarchical structure in which (without loss of generality) the decision maker is a supergeneralist. When should (semi-)generalists of cardinality less than \( n \) be employed by the organization? It is clear that, if semi-generalists are to be used at all, they should be intermediaries between the top manager and the superspecialists. We state this formally as

**Proposition 4 (Knowledge and Rank)** Under an optimal organizational structure, a generalist-\( B \) receives reports about attribute \( i \) from a generalist-\( A \) only if \( i \in A \cap B \) and \( \# B > \# A \). In other words, individuals with more specialized knowledge are subordinates to individuals with less specialized knowledge.

**Proof.** If \( i \notin B \), the generalist-\( B \) receives an uninformative report, because \( \alpha \left( d_{iB}^{AB} \right) = \alpha (\infty) = \infty \). If \( i \notin A \), the only information that the generalist-\( A \) has is the prior distribution of \( i \), which is assumed to be known by everyone. Thus, meaningful communication is only possible if \( i \in A \cap B \).

Suppose now that \( i \in A \cap B \), but \( \# B \leq \# A \). Notice that optimality requires minimizing the communication error. Let \( \{A_1, \ldots, A_K\} \) be the sequence of types through which information about \( i \) flows until it reaches \( A \). Pick the last \( A_k \) in that sequence such that \( \# A_k \leq \# B \) (it exists because it must be the case that \( A_1 = \{i\} \)). By construction, the communication error between such an \( A_k \) and \( A_{k+1} \) must be no lower than the one between \( A_k \) and \( B \), because \( \# A_{k+1} - \# A_k \geq \# B - \# A_k \). Therefore, \( B \) could get a signal directly from \( A_k \) that would be at least as precise as the one \( A \) can get directly from \( A_k \), or indirectly through \( (A_{k+1}, \ldots, A_K) \), because errors are additive. Thus, direct communication from \( A_k \) to \( B \) yields with no loss of information, implying that \( A \) is redundant. ■

The role of middle managers is to reduce communication errors. Let a generalist-\( A \) of cardinality \( k \) be an intermediary between \( k \) specialists and the supergeneralist. The report that \( N \) receives from \( A \) about attribute \( i \) is

\[
r_i^{AN} = s_i^A + \alpha \left( d_i^{AN} \right) u_i = x_i + \alpha (k - 1) u_i + \alpha (n - k) u_i.
\]  

(15)
If the top manager receives his report directly from the superspecialist-\(i\), his reading will be

\[
    r_i^{iN} = s_i^i + \alpha (d_i^{iN}) u_i = x_i + \alpha (n - 1) u_i. \tag{16}
\]

Because the middle manager is a means through which information flows from the bottom to the top, introducing a middle manager of type \(A\), who receives a report about \(i\) from the superspecialist-\(i\) and then sends it to the top manager, is better than the flat hierarchy if and only if

\[
    \alpha (k - 1) + \alpha (n - k) < \alpha (n - 1). \tag{17}
\]

We generalize the previous argument in the following proposition.

**Proposition 5 (Middle Management)** Consider an organizational structure in which member \(A\) reports to member \(B\) and \(\#B > \#A\). The introduction of a middle manager \(C\) between the two increases expected profits if and only if:

1. \(\#B > \#C > \#A\), and

2. there is at least one \(i\) such that \(i \in A \cap B \cap C\) and \(\alpha (\#B - \#C) + \alpha (\#C - \#A) < \alpha (\#B - \#A)\).

**Proof.** If \(\#B - \#A = 1\), introducing one more manager does not change expected profits. Thus, by the lexicographic criterion, it is better not to do so. This result and Proposition 4 jointly imply the necessity part of item 1.

Let \(\#B - \#A > 1\) and consider a middle manager \(C\) such that \(\#B > \#C > \#A\). Suppose there is an \(i\) such that \(i \in A \cap B \cap C\) and \(\alpha (\#B - \#C) + \alpha (\#C - \#A) < \alpha (\#B - \#A)\). Therefore, introducing \(C\) improves the precision of the signal received by \(B\), which also improves the precision of the signal received by the decision maker, thus increasing expected profits, proving the sufficient part of the proposition.

Suppose now that \(\alpha (\#B - \#C) + \alpha (\#C - \#A) > \alpha (\#B - \#A)\). Then, the opposite occurs: introducing \(C\) worsens 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, by the lexicographic criterion, 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 $C$. This proves the necessity part of item 2. ■

The main determinant of middle management is the shape of the intensity of communication error function $\alpha(\cdot)$. The economic reasoning behind this result is easily understood. Consider two individuals whose distance in expertise between them is large. When they communicate with each other, there is a large communication error. Suppose that we find a third individual whose expertise is in between those of the two original individuals. If equation (17) holds, this middle manager may be able to translate the information, thus increasing the precision of the communication process. Of course, the middle manager cannot be so far away from the two extremes (in terms of expertise) in order for this translation to be effective.

Under the assumption that a flat hierarchy is better than no hierarchy, the following results are immediate from Proposition 5.

**Corollary 6** If the intensity of communication error is a subadditive function, then the optimal structure is the flat hierarchy.

**Corollary 7** If the intensity of communication error is a superadditive function, then the optimal structure has $n$ layers of management.

These results show that the shape of the intensity of communication error determines the optimal depth of hierarchies. Structures with many layers can be optimal as long as middle managers are good translators across different specialized activities.

### 3.4 Some Examples and Applications

To offer some additional examples and results, we first define some useful concepts.

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

Non-hierarchic forms occur when at least one of the individuals report to more than one superior, as in matrix forms. This definition is implicitly adopted by most papers on
organization design.\footnote{Radner (1993) and Harris and Raviv (2002) explicitly define hierarchies in this way.}

**Definition 3** We say that a structure has $l_i$ **levels** of individuals involved with attribute $i$ if information about $i$ needs to pass through $l_i - 1$ individuals before it reaches the decision maker.

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

These definitions are similar to the ones in Radner (1993), with one small difference. Since Radner is concerned only with hierarchic structures, he only defines strictly balanced **hierarchies**.\footnote{In Radner’s 1992 paper, he uses the term *regular* instead of *strictly balanced*.} However, because we want to consider the possibility of non-hierarchic structures, we extend 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.

When middle management is an option, we can prove the following interesting result.

**Proposition 6 (Optimality of Strictly Balanced Hierarchies)** There always exists a strictly balanced hierarchy that maximizes expected profits.

**Proof.** See the Appendix. ■

This result does not imply that other types of structures are sub-optimal. It is possible for matrix structures or structures that have “skip-level” reporting to be optimal. Proposition 6 only implies that, as long as only the information processing properties of structures are important, there is no gain from deviating from a strictly balanced hierarchy.

Some important papers on organizations assume strictly balanced hierarchies without comparing them with other alternatives (Beckmann, 1960; Keren and Levhari, 1983). Proposition 6 illustrates that, under our assumptions, this approach may be a reasonable one.
For simplicity, we have assumed that there are no costs of adding managers. Introducing costs of hiring managers or acquiring expertise may make non-balanced and non-hierarchic structures strictly better than strictly balanced hierarchies, as we can see from the examples below.

**Example 1** Let \( N = \{1, 2, 3, 4\} \), \( \sigma^2 = \sigma_u^2 = 1 \). Let the intensity of communication error be

\[
\alpha (d_{i}^{AB}) = \left( \frac{d_{i}^{AB}}{3} \right)^{2}.
\]  (18)

First, notice that a flat hierarchy with a supergeneralist at the top is better than a specialist-managed structure, because \( n = 4 > 2 = 1 + \alpha(3) \frac{\sigma^2}{\sigma_u^2} \). In order to find the optimal structure, we apply Proposition 5. First, should we introduce generalists with cardinality 2? Consider \( A = \{1, 2\} \). We have that \( d_{1}^{IA} = d_{2}^{IA} = 1 \), therefore \( \alpha (d_{1}^{IA}) = \alpha (d_{2}^{IA}) = \frac{1}{9} \). Also, \( d_{1}^{AN} = d_{2}^{AN} = 2 \), implying \( \alpha (d_{1}^{AN}) = \alpha (d_{2}^{AN}) = \frac{4}{9} \). If superspecialists communicate directly with the supergeneralist, then \( d_{1}^{1N} = d_{2}^{2N} = 3 \) and \( \alpha (d_{1}^{1N}) = \alpha (d_{2}^{2N}) = 1 \). Therefore,

\[
\frac{1}{9} + \frac{5}{9} < 1.
\]

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,

\[
\alpha (1) + \alpha (1) = \frac{1}{9} + \frac{1}{9} < \frac{5}{9} = \alpha (2).
\]

Thus, \( C \) should be introduced, as well as another generalist-3 who has some expertise about \( \{4\} \). Therefore, an optimal structure will have 4 levels of management, as shown in Figure 3. We could also have used Corollary 7 directly: Since (18) is strictly superadditive, one needs four levels of management.

Notice that, in this example, the 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 from the following example.

**Example 2** Let \( N = \{1, 2, 3, 4, 5, 6, 7\} \), \( \sigma^2 = \sigma_u^2 = 1 \). Let the intensity of communication error be

\[
\alpha (d_{i}^{AB}) = \left( \frac{d_{i}^{AB}}{6} \right)^{2}.
\]  (21)
This example is similar to the previous one: The intensity of communication error is strictly superadditive, therefore the optimal structure has 7 levels of managers. Any structure that minimizes the number of managers in this case has four generalists-2 at the second level from the bottom and three generalists-3 at the third level. In order for this to happen, at least one generalist-2 must report to two different generalists-3, as shown in Figure 4.

4 Final Remarks

The model we present here is relatively straightforward and easy to manipulate. We have sacrificed generality in order to get a more manageable model. However, we believe that the model is still quite general and can be extended to analyze a number of different questions. Here, we briefly discuss some possible extensions of the model and its limitations.

We have assumed that there are no costs of adding managers (salaries). But this assumption is admittedly false. Any organization should try to compare the costs of extra managers to the benefits they create. We have showed by means of examples that allowing for costs of adding managers may make non-balanced and non-hierarchic structures strictly better than balanced hierarchies. With additional assumptions about the costs of different types of managers, one could design an algorithm that solves for the best organizational structure.

We have also assumed a very simple knowledge acquisition technology: there is a trade-off between the number of areas of expertise and the depth of knowledge in each area. In our setup, the (implicit) costs of acquiring knowledge depend only on the number of areas of expertise, but not on the nature of the areas themselves. However, a more realistic assumption would allow individuals to find it easier to acquire expertise in some set of areas than in others. In such a case, some other realistic structures might also be optimal. One example is the matrix structure.\textsuperscript{18}

Finally, we have also assumed that all project attributes have the same prior distribution and that all individuals know these priors. This need not be the case. If specialists cannot understand reports about an attribute outside their areas of expertise, why should they be

\textsuperscript{18}For theories of optimal matrix structures, see Harris and Raviv (2002) and Costa, Ferreira, and Moreira (2005).
able to know anything at all about this attribute (i.e. its prior distribution)? Dropping the assumption that everyone knows all prior distributions may lead to a situation in which it is never optimal to have a specialist decision-maker. This opens the door for authority delegation: the top manager might find it optimal to delegate the decision to a subordinate contingent on the information structure. These are all potential topics for further research.

A Appendix

Proof of Proposition 1

Proof. Unlike in the text, here we do not assume that $\mu$ and $c$ are zero.

If a superspecialist makes the decision, he will undertake the project if and only if

$$x_i + (n - 1) \mu \geq c. \quad (22)$$

The (ex ante) probability of undertaking the project is

$$\Pr (X_i \leq c - (n - 1) \mu) =$$

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

where $a = \frac{c-n\mu}{\sigma}$ and $\Phi(.)$ is the standardized normal cumulative distribution function. If we denote by $E_s(\pi)$ the ex ante expected profit under the specialist-managed structure, we have that:

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

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

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

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

$$[1 - \Phi (a)] (n\mu - c) + \sigma \phi (a). \quad (23)$$

where $\phi$ is the standardized normal density function.
On the other hand, if a supergeneralist is the decision maker, who receives reports from \( n \) superspecialists, then he will Bayesian update his expectations. More precisely, given (3) we have that
\[
x_i = r_i - \alpha u_i.
\]

Therefore,
\[
E [ X_i \mid R_i ] = \mu + \beta ( R_i - \mu ),
\]
where
\[
\beta = \frac{cov ( R_i, X_i )}{var ( R_i )} = \frac{cov ( X_i + \alpha u_i, X_i )}{var ( X_i + \alpha u_i )} = \frac{\sigma^2}{\sigma^2 + \alpha^2 \sigma_u^2}.
\]

Then, the project will be undertaken if and only if
\[
n\mu + \sum_{i=1}^{n} \beta ( r_i - \mu ) \geq c.
\]

Thus, the (ex ante) 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,
\[
a \geq b \iff \sqrt{n} \beta \geq 1 \iff \beta \geq \frac{1}{n} \iff \frac{\sigma^2 + \alpha^2 \sigma_u^2}{\sigma^2} \leq n \iff 1 + \frac{\alpha^2 \sigma_u^2}{\sigma^2} \leq n.
\]

If we denote by \( E_m ( \pi ) \) the ex ante expected profit under the managerial structure, we have that
\[
E_m ( \pi ) = [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] = \\
27
Because $\beta = \frac{\sigma^2}{\sigma^2 + \alpha^2 \sigma_u^2}$, we have that

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

Now, define the function

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

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 $x \geq 1$, $E_m(\pi) = g(x) \geq g(1) = E_s(\pi)$. But $\sqrt{n \beta} \geq 1 \iff \beta \geq \frac{1}{n}$

$$\iff \frac{\sigma^2 + \alpha^2 \sigma_u^2}{\sigma^2} \leq n \iff 1 + \frac{\alpha^2 \sigma_u^2}{\sigma^2} \leq n \tag{27}$$

**Proof of Proposition 2**

**Proof.** Let a generalist-$A$ be the decision maker, receiving reports from $n$ superspecialists. Define $k_A \equiv \#A \leq n$. Without loss of generality, let $A$ be equal to $\{1, 2, ..., k_A\}$. The generalist-$A$ undertakes the project if and only if

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

where $\beta_A = \frac{\alpha^2}{\sigma^2 + \alpha^2 \sigma_u^2}$, $\alpha_A = \alpha \left( d_i^{Ai} \right)$, $\forall i \in A$. Thus, the probability of undertaking the project is (the following steps are identical to the ones in the proof of Proposition 1):

$$1 - \Phi (b_A),$$

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

The ex ante expected profit is

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

(28)
Notice that $\beta_A \sqrt{k_A (\sigma^2 + \alpha_A^2 \sigma_u^2)} = \sigma \sqrt{k_A \beta_A}$ and $\beta_B \sqrt{k_B (\sigma^2 + \alpha_B^2 \sigma_u^2)} = \sigma \sqrt{k_B \beta_B}$. Recall function $g(.)$ as defined in (25). From (28), expected profits under $A$ and $B$ are given by $g(\sigma \sqrt{k_A \beta_A})$ and $g(\sigma \sqrt{k_B \beta_B})$. But since $g(.)$ is an increasing function, $g(\sigma \sqrt{k_A \beta_A}) \leq g(\sigma \sqrt{k_B \beta_B}) \Rightarrow \sqrt{k_A \beta_A} \leq \sqrt{k_B \beta_B} \Rightarrow k_A \beta_A \leq k_B \beta_B \Rightarrow k_A \beta_A \leq \frac{k_A}{k_B} \beta_A \leq \frac{\sigma^2 + \alpha_A^2 \sigma_u^2}{\sigma^2 + \alpha_B^2 \sigma_u^2}$.

Proof of Proposition 3

Let $k^*$ be the cardinality of the decision maker that maximizes expected profits (we know that $k^*$ is well-defined because the objective function is continuous and defined over a non-empty compact set). Define

$$C (k) = \frac{\sigma^2 + \alpha (k^* - 1)^2 \sigma_u^2}{\sigma^2 + \alpha (k - 1)^2 \sigma_u^2}.$$  

Optimality requires that the $n$ conditions below are met

$$k^* \geq C (k) k, \text{ for all } k \in \{1, \ldots, n\}.$$  

Suppose first that $\sigma^2$ changes. It is straightforward to check that

$$\text{sign} \left\{ \frac{\partial C (k)}{\partial \sigma^2} \right\} = \text{sign} \left\{ \alpha (k - 1)^2 - \alpha (k^* - 1)^2 \right\},$$  

which implies that $C (k)$ is decreasing in $\sigma^2$ for all $k < k^*$. Thus, for all $k < k^*$ condition $k^* \geq C (k) k$ is slacked by an increase in $\sigma^2$, implying that no $k < k^*$ can become the new optimal value after an increase in $\sigma^2$.

To prove the second part, it suffices to realize that

$$\text{sign} \left\{ \frac{\partial C (k)}{\partial \sigma_u^2} \right\} = -\text{sign} \left\{ \frac{\partial C (k)}{\partial \sigma^2} \right\}.$$  

Proof of Proposition 6

Proof. The proof is by construction. Start with any given optimal structure, not necessarily strictly balanced nor hierarchic. Say that attribute 1 has $l_1$ levels of managers. Let $T^j_1$ be the type of the manager at the $j$-th level of attribute 1. The report received by the decision maker is

$$r^*_1 = x_1 + \sum_{j=1}^{l_1 - 1} \alpha \left( d_{ij}^{j} \right) u_1.$$  

29
Thus, the variance of the communication error is
\[
\left[ \sum_{j=1}^{l_1-1} \alpha \left( d_{1T_i^jT_i^{j-1}} \right) \right]^2 \sigma_u^2.
\]

Say that attribute 2 has \(l_2 \neq l_1\) levels. Let \(T_i^j\) be the manager at the \(j\)-th level of attribute \(i\). Find a set of \(l_1 - 1\) types of managers \(\{A^1, \ldots, A^{l_1-1}\}\) such that \(#A^1 = #T_1^1, \ldots, #A^{l_1-1} = #T_1^{l_1-1}\) and \(\{2\} \in A^1 \cap \ldots \cap A^{l_1-1}\). In other words, this means finding a set of managers that can exactly replicate for attribute 2 the reporting structure of attribute 1. Now, replace the old reporting structure of attribute 2 by \(R(A^1) = R(A^2), \ldots, R(A^j) = R(A^{j+1}), \ldots, R(A^{l_1-1}) = R(m^*)\). It has to be true that
\[
\sum_{j=1}^{l_1-1} \alpha \left( d_{1T_i^jT_i^{j-1}} \right) = \sum_{j=1}^{l_1-1} \alpha \left( d_{1A^jA^{j-1}} \right) \geq \sum_{j=1}^{l_2-1} \alpha \left( d_{2T_2^jT_2^{j-1}} \right).
\]

The first equality is true by construction. The inequality has to hold, because otherwise switching to \(\{A^1, \ldots, A^{l_1-1}\}\) in the second attribute would reduce the variance of the communication error, increasing expected profits, which cannot be feasible if this is an optimal structure. But our labeling of attributes 1 and 2 is completely arbitrary and can be reversed, so we have established that
\[
\sum_{j=1}^{l_1-1} \alpha \left( d_{1T_i^jT_i^{j-1}} \right) = \sum_{j=1}^{l_2-1} \alpha \left( d_{2T_2^jT_2^{j-1}} \right).
\]

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

In order to transform the previous structure into a hierarchy, we use a simple procedure. Without loss of generality, suppose that a given manager \(A\) sends reports about attribute 1 to a manager \(B\) and about attribute 2 to another manager \(C\). Then, the information about the two attributes will follow different paths until it reaches the top. We can always find a manager of type \(D\) such that \(A \subset D\) and \(#D = #B\) (this is always possible, since Proposition 4 implies that \(#B > #A\), thus now \(A\) can send reports about both attributes to \(D\) only. We then keep the information path for attribute 1 unchanged, while we change
the reporting structure of 2 in the same way described above. For the same reasons described above, this cannot affect expected profits. Then, we repeatedly apply this procedure to all managers until everyone reports to only one superior, ending up with an optimal hierarchical structure.

Now, we have an optimal structure that is a hierarchy and has the same number of levels in each attribute. Suppose now that at 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 managers in each level have only one subordinate). This completes the proof. ■

References


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

Figure 4