

Who gets to the top? Generalists versus specialists in managerial organizations

Daniel Ferreira*

and

Raaj K. Sah**

We study organizations with individuals whose expertise differ in content and breadth. For example, specialists have deeper expertise than generalists, but in fewer areas. Difficulties in communication depend on who communicates with whom. Our analysis, which is consistent with several empirical findings, shows that: (i) an organization is more valuable and its leader has broader expertise if it is more complex, faces more unpredictability, or communication technologies improve; (ii) those higher in multilayered hierarchies have broader expertise; and (iii) any one-dimensional concept (e.g., talent) cannot explain the assignment of different individuals to different levels in hierarchies.

“You spoke. The only one to understand it was you. Pray tell what that understanding is?”¹

1. Introduction

■ One motivation for this article on certain aspects of managerial organizations is as follows. Casual and systematic evidence suggests that more MBAs head large business corporations than those with any other educational degree.² This is so for business corporations with diversified lines of business, and also for those with relatively narrow lines of business, such as generic pharmaceuticals and specialty steel. A key selling point of the high-end MBA programs is that their graduates will learn virtually all aspects of business corporations, and that this will help them in career paths leading to top management. The courses in such programs draw from disciplines (e.g., economics, operations research, psychology, and sociology) that are arguably more diverse than those underlying the educational training of, say, actuary, IT professional, or

*London School of Economics, ECGI, and CEPR; d.ferreira@lse.ac.uk.

**University of Chicago and Indian Institute of Management, Ahmedabad; sah@uchicago.edu.

We thank Beth Allen, Pierre-André Chiappori, Luis Garicano, Ranjay Gulati, Winston Koh, David Martimort, Mike Mei, Vipul Naik, Aaron Plavnick, Canice Prendergast, Tetsushi Sonobe, and many seminar and conference participants for their help. We thank two anonymous referees for their valuable suggestions.

¹ This quotation is from a verse satirizing the writings of Ghalib (1797–1869), a deeply adored poet from South Asia, penned by one of his contemporaries. The original, “*Agar apna kaha . . .*,” is in Urdu (see Kanda, 2004); its rendering in English is by the authors.

² For example, Bertrand and Schoar (2003) find that nearly 40% of all CEOs in their sample are MBAs. They also note that this fraction is larger for younger executives.

licensed accountant. Compared to MBAs, the latter kinds of individuals usually have deeper expertise in their respective sets of areas. Perhaps partly for this reason, those who have worked for many years as specialists (including, for example, as engineers and scientists) try, on their way to the top, to reeducate, or at least to repackage, themselves as generalists. Many business schools offer executive training programs, of various durations, to facilitate such transitions. The recent growth of such programs, and of regular MBA programs, has been rapid and worldwide.

Another motivation for this article comes from the empirical literature on organizational psychology. There is compelling evidence that individuals who have expertise in different areas find it difficult to communicate among themselves. For example, an article by Heath and Staudenmayer (2000), which also contains several relevant references, summarizes these phenomena as “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.”

We can therefore think of some individuals as having deeper expertise in a smaller number of areas, and of other individuals as having less deep expertise in a larger number of areas. For brevity, we refer to them, respectively, as *specialists* and *generalists*, although, in this article, we model many different kinds and breadths of “generalist-ness.”

The work of large business corporations usually requires specialists as well as generalists. For example, an integrated multinational energy company typically has experts in geology, drilling, offshore oil rigs, pipelines, shipping, refining, marketing, hedging and trading, and so on. There usually are further subdivisions of expertise within each of these areas. Such large business organizations typically have more than one subordinate reporting to one superior manager. There is at least some heterogeneity in the types of expertise that these subordinates have, and there is at least some commonality of expertise between the superior and each of his subordinates.

With these motivations, the primary focus of this article is on individuals who have various types of expertise, and on communications among them. Throughout the article, all relevant communications are imperfect. How these imperfections are amplified, when two individuals communicate with each other, depends on their respective types of expertise. We examine the consequences of these interactions (between expertise types, and communications among them) for the value of organizations, the role of top and middle managers, and the assignment of different types of individuals to different levels of organizational hierarchies.

Among our qualitative results are the following; later in the article, each of these is stated with greater precision. Organizations are more valuable, and the breadth of the expertise of the top manager is larger, if there is a greater unpredictability or complexity in the organization’s business, or if exogenous changes reduce the difficulties in communication. Our analysis is consistent with several empirical findings, including that exogenous improvements in communication technologies tend to flatten hierarchies.

Another of our qualitative results regards what kinds of individuals are assigned to different levels of managerial hierarchies. In Rosen (1982), individuals at higher levels are more talented. In Garicano (2000), they are more able to solve difficult problems.³ In Prat (1997), they have greater capacity for processing information. Our analysis generates multilayered hierarchies in which those at higher levels have larger breadth of expertise than those at lower levels. Broader expertise is not the same as greater talent or greater ability to solve difficult problems, or vice versa. For example, consider an extraordinarily talented individual who can quickly process and solve transfer-pricing problems that virtually no one can even understand. This individual might lead a specialized unit in a company such as the energy multinational described earlier. However,

³ Garicano (2000) is important for another reason. Consider a patient with serious but nonemergency medical symptoms. Then, a primary physician is typically the first person to examine the patient, successively followed by physicians who are increasingly more specialized. There are many contexts in which the decision making flows analogously, from more of a generalist to more of a specialist. Among other things, Garicano (2000) studies such contexts. Such contexts are beyond the scope of the present article. Instead, as stated and illustrated earlier, our motivations come from managerial organizations, where communication among many different areas of expertise is important.

he is unlikely to also have the portfolio of expertise to effectively oversee the company's numerous other activities. Accordingly, he will likely not head the company.

Another perspective that our analysis brings out is that any one-dimensional concept, no matter what its name or definition, cannot explain the assignment of different individuals to different levels in hierarchies. We concretely illustrate this later in Sections 4 and 5. We summarize here the general idea using talent (e.g., Rosen, 1982); the same observations apply to expertise breadth (as in the present article), problem-solving capability (e.g., Garicano, 2000), information-processing capacity (e.g., Prat, 1997), or anything else. A scalar metric is needed if the ranking of individuals were based solely on talent. Therefore, suppose that an individual's talent is a scalar function with many arguments; what these arguments are does not matter for our conclusions. Then the same function will not apply across the profuse heterogeneity that we see across individuals within and across organizations, including in their activities and contexts. In contrast, for example, a scalar function might work relatively better in ranking Olympic-style weightlifters, within a particular weight category, because what is being measured is deliberately designed to be relatively context free.

To our knowledge, this article is the first attempt in economics to model and study breadth of expertise, and the process of communications across different types of expertise. As part of this, a potential contribution of the article is to develop the beginnings of a taxonomy to articulate these issues. Separately, this article's bias is to try to obtain qualitative results, and not to focus solely on the generality of models. This is because, otherwise, it is difficult to incorporate the lessons available from the empirical literature. Given these considerations, we work with a simple model. If qualitative results are not available in such a model, they are unlikely to be available in more general ones. On the other hand, as illustrated in Section 5, results from a simple model may potentially give insights into some of the tradeoffs that might arise within more general models.

In order to understand the interactions between expertise and communication, we should first isolate them from other considerations. Thus, for example, we abstract from incentive alignment (or agency) problems. This separation between coordination and incentive issues is standard in the literature. The team-theory approach to organizational problems (e.g., Marschak and Radner, 1972) focuses on imperfect information transmission when preferences are aligned, whereas the principal-agent approach (e.g., Holmström, 1979) focuses on imperfect preference alignment when information transmission is perfect. This separation is not without costs, and an integration of these two approaches is a promising topic for future research.⁴

□ **Organization of the article.** In the rest of this section, we briefly discuss some of the literature related to the present article. In Section 2, we present the basic model. In Section 3, we analyze two-layered hierarchies. For these, we present several qualitative results, along with their intuitions. The analysis of these simple hierarchies also facilitates that in the rest of the article. In Section 4, we begin with an arbitrarily organized multilayered hierarchy, and then examine some consequences of optimality. This yields several results on the nature and role of middle managers. The article concludes with Section 5, which presents some extensions of, and remarks on, the preceding analysis.

□ **Related literature.** The importance of coordination costs for limiting the degree of specialization is highlighted by Becker and Murphy (1992). In their model, a team of specialized workers is formed to produce a single good. The optimal size of the team is determined by the tradeoff between specialization and coordination: a larger team size increases task specialization but makes the coordination of different tasks more difficult. Our model has similar features; specialization is beneficial but requires coordination.

The article that is closest to ours is Prasad (2009), who independently studies the role of task specialization in hierarchical assignments. He is interested, as we are, in developing a theory of

⁴ Dewatripont and Tirole (2005) have taken some important steps toward such an integration.

hierarchies in which abilities have multiple dimensions. His focus is on the incentive properties of specialization in a multitasking principal-agent model; we abstract from incentive issues and focus on communication. A different contribution, sharing similarities with our article, is by Niehaus (2011), who studies a model of social learning in which the quality of communication depends on the match between heterogeneous senders and recipients.

Cremer, Garicano, and Prat (2007) focus explicitly, as we do, on imperfect communication in organizations. They consider the problem of finding the optimal organization-wide code (or language) that improves intraorganizational communication. In both their article and ours, managers act as translators of information. We take the organization-wide code as given, and focus on the phenomenon that individuals with different types of expertise speak different languages. Such difficulties in communication are likely to be significant even in organizations that have a highly developed organization-wide code.

Another important contribution is the work of Dessein and Santos (2006). In their model, the optimal organization trades off coordination (i.e., exploiting synergies across tasks) and adaptation (i.e., better use of local information). They find, as we do, that improvements in the communication technology may lead to less specialization, and that increases in complexity reduce specialization. Our article studies different aspects. For example, the quality of communication depends on the expertise types of the communicators, and our study includes the design of hierarchies, for example, the assignment of different types of managers to different levels.

Dewatripont and Tirole (2005) model imperfect communication problems arising from preference heterogeneity and moral hazard. As in our model, information in their model is neither “hard” nor “soft”; there are varying degrees of information “softness.” The quality of communication depends both on exogenous factors, such as shared language, and on endogenous ones, such as effort. Our article differs in that it includes examining the selection of these factors (languages, functional backgrounds, etc.) and, in addition, analyzing the suitability of different bundles of factors to different organizational roles and ranks.

An interesting related question is whether heterogeneity in preferences or opinions has consequences for organizational design. Landier, Sraer, and Thesmar (2009) develop a model in which dissent in organizations may have costs and benefits. The division of tasks in organizations (in their case, selecting and implementing projects) implies that there is an optimal level of heterogeneity in organizations, and that agents with different characteristics should be assigned to different ranks. Our article differs due to our focus on differences in expertise rather than in preferences or opinions, although these aspects could be correlated to some degree.

Our article is also related to the literatures on information processing within organizations, although their concerns are different from ours. In the literature that includes Radner (1992, 1993) and Van Zandt (1999), the organization’s goal is to minimize the time that it takes for information to flow from the bottom to the top, given that each agent is endowed with a limited capacity to process information. Radner (1993) also defines certain subcategories of hierarchies, such as regular or strictly balanced. Earlier versions of the present article (e.g., Ferreira and Sah, 2002) have included some implications of our concerns within such subcategories. For brevity, we omit that analysis here and, instead, consider multilayered hierarchies without *a priori* restrictions on their structures. In Bolton and Dewatripont (1994), specialization makes individuals better processors of information. Sah (1991), Sah and Stiglitz (1986, 1991), and the related literature study the consequences of human fallibility (of which information processing is only one aspect) for alternative structures of organizations and societies.

Many other articles deal with 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 his limited area of expertise. In Vayanos’s (2003) model of information processing with synergies, top managers process more aggregated information, whereas managers at the bottom process only local information. Geneakoplos and Milgrom (1991) develop a theory of organizations based on managers’ limited attention. Castanheira and Leppämäki (2004), Hart and Moore (2005), and Van den Steen (2010)

are examples of models of authority, in which subordinates only act when coordinators tell them to. Among the studies that include the trade-off between communication and delegation, based on agency considerations, are Aghion and Tirole (1997), Baker, Gibbons, and Murphy (1999), Dessein (2002), Jensen and Meckling (1992), and Stein (2002). Lazear (1995) connects many of these issues to an overview of personnel economics.

2. A basic model

□ **Project attributes and types of expertise.** We consider an organization that accepts or rejects projects (i.e., investments, ideas, and proposals). A project has n attributes; $i = 1$ to n . To avoid unnecessary details, n is finite and $n \geq 2$. The attributes are represented by i.i.d. random variables X_1, X_2, \dots, X_n , each of which is normally distributed with mean μ and variance σ^2 . We denote the realized value of the random variable X_i by its lowercase counterpart x_i . The independence of X_i implies that the information on one attribute is not useful for the evaluation of other attributes. An interpretation of a larger n is that the organization's business is more *complex*. An interpretation of a larger σ is that the organization faces greater *business unpredictability*. The mean μ is finite, and it can be positive, zero, or negative. Unless stated otherwise, σ is finite and strictly positive.

Each attribute corresponds to a particular expertise. Each individual possesses expertise on one or more attributes. In our abstraction, an individual is defined completely by the attributes on which he has expertise. Thus, for example, let A denote a nonempty subset of the set of all attributes, $N \equiv \{1, 2, \dots, n\}$. Then, A is an *expertise type*, that is, a particular combination of expertise. We refer to an individual as having expertise type A if the subset A describes the attributes on which he has expertise. For brevity, we often refer to this individual simply as "individual A ." Note that we are using the same symbol A for an individual and also for the attributes on which he has expertise. This should cause no confusion because, as just stated, an individual is completely defined here by the attributes on which he has expertise. We likewise denote individuals with other expertise types by symbols B , C , and so forth.

If $A = N$, we refer to individual A as a *supergeneralist*. If A is a singleton for an individual, that is, $A = \{i\}$, we call him a *superspecialist- i* . In between these two extremes, there are $2^n - n - 2$ generalists with different expertise types.⁵

We often use a partial description of an individual, which we refer to as his *expertise breadth*. This is the number of attributes on which an individual has expertise. We denote the expertise breadth of person A as $\#A$, which is the number of elements in set A . The reason why an individual is only partially described by his expertise breadth is that individuals with different expertise types can potentially have the same expertise breadth. Specifically, as noted in footnote 5, there are $\binom{n}{j}$ different expertise types with the same expertise breadth j . Thus, the supergeneralist's expertise breadth is n ; that of any of the n types of superspecialists is one; and in between, there are $n - 2$ different possible expertise breadths ($2, \dots, n - 1$).

□ **Communication.** Each individual in this organization may send and receive reports on one or more attributes of the project. Throughout the article, all relevant communications contain errors. The aspects of imperfect communication that we emphasize are the following; a more precise description is presented below in this subsection. If two individuals with the same expertise type communicate with each other, then no additional errors are introduced; that is, the recipient receives the same report that the sender sends. However, if an individual of a particular expertise type receives a report from another individual of a different expertise type, then the former receives a version which, compared to the report sent, is noisier or, at best, no less noisy.

⁵ There are $\binom{n}{j}$ distinct subsets of N , each with j elements, for $j = 1$ to n . The supergeneralist corresponds to $j = n$. Hence, there is $\binom{n}{n} = 1$ supergeneralist. A superspecialist corresponds to $j = 1$. Hence, there are superspecialists with $\binom{n}{1} = n$ different expertise types. The case $j = 0$ corresponds to $\binom{n}{0} = 1$, and it is not relevant here. Further, $\sum_{j=0}^n \binom{n}{j} = 2^n$. Hence, besides the supergeneralist and the superspecialists, there are $2^n - n - 2$ expertise types.

There are at least two mutually interlinked sources of amplification of the communication errors when two individuals communicate with each other: one relates to the sender of the report, and the other to its recipient.⁶

The recipient cannot fully understand the report because of his various limitations. We assume that if the recipient has a deeper expertise on the attributes being reported, then he has a better understanding of a report on these attributes, and therefore he contributes less to the amplification of the errors in communication.

The sender is unable to comprehend fully the limitations of the recipient. Given this inability, the sender can at best use all of the approaches for communication that are available to him, such as using different concepts or lines of exposition, none of which are perfectly suited to the recipient. We therefore assume that the more “similar” (in a sense to be defined later) the sender and the recipient are, the smaller is the amplification of the errors in communication.

As a first step in formally describing the communication process, we define the *expertise gap* between the sender and the recipient. Let A send a report to B on attribute i . Then, one specification of the expertise gap is that, if $i \in A \cap B$, the gap is

$$d_i^{AB} \equiv |\#A - \#B|. \quad (1)$$

If $i \notin A \cap B$, we assign an arbitrarily large magnitude to the expertise gap, which we informally express as $d_i^{AB} = \infty$.

According to specification (1), if both individuals have expertise on an attribute i (i.e., if $i \in A \cap B$), the expertise gap between them is the difference in their expertise breadths. In the case $i \notin A \cap B$, either A or B , or both, has no expertise on attribute i . As we shall see below, in this case, they are completely unable to communicate on attribute i .

In specification (1), the expertise types are independent in the following sense. For example, let $A = \{1, 2\}$, $B = \{1, 3\}$, and $C = \{1, 2, 4, 5\}$. The expertise gap with respect to attribute 1 is 2 between A and C , and it is also 2 between B and C . This is so even though A has more in common with C than B does. In other words, the common expertise regarding attribute 2 does not additionally help A and C when they are communicating on attribute 1.

The above notion of independence is realistic in many situations. For example, if two individuals communicate on some specific aspect of nanotechnology, the quality of this communication will likely not be altered significantly by whether or not they share expertise on the Chinese business culture. At the other extreme, consider the following alternative specification of the expertise gap. When $i \in A \cap B$, the expertise gap between A and B is

$$\max\{\#A, \#B\} - \#(A \cap B). \quad (2)$$

Recall the earlier example that $A = \{1, 2\}$, $B = \{1, 3\}$, and $C = \{1, 2, 4, 5\}$. Consider the expertise gap concerning attribute 1. Now this gap is 2 between A and C , and it is 3 between B and C . Thus, under specification (2), the commonly shared expertise concerning attribute 2 that A and C have helps them even when they are communicating on attribute 1.

Both specifications (1) and (2) appear extreme. We expect the reality to lie somewhere in the middle. Fortunately, many of our results do not depend on whether specification (1) or (2) is employed.⁷ Hence, we work with specification (1).

For use immediately below, we define the intensity of errors in managerial communications. For brevity, we refer to it as *error intensity*, and denote it by the function α . This function is increasing in the expertise gap. Let the nonnegative integer d denote any arbitrary magnitude of

⁶ Our model is about imperfect communication rather than about limited human capacity for processing information. Thus, one cannot remove these imperfections, for example, by using mechanisms such as parallel communication channels, or by employing individuals with higher capabilities to process information.

⁷ If A reports to B , then a sufficient condition for (1) and (2) to be identical is that A be a strict subset of B . This is because, with $A \subset B$, (2) becomes $\#B - \#(A \cap B) = \#B - \#A$, which is the same as (1). As we will see later, the sufficient condition just noted is satisfied, for example, by all of the reporting relationships in Sections 3 and 5.

the expertise gap. We assume that the following properties hold:

$$\alpha(d') > \alpha(d) \text{ if } d' > d, \alpha(0) = 0, \text{ and } \alpha(\infty) = \infty. \tag{3}$$

We now model the communication process. Let g_i^A be the report on attribute i sent by person A to person B . Then, d_i^{AB} is the expertise gap between A and B . The report received by B is

$$r_i^{AB} \equiv g_i^A + u_i \alpha(d_i^{AB}), \tag{4}$$

where $u_i \sim \mathcal{N}(0, \theta^2)$. Throughout the article, \mathcal{N} denotes a normal distribution, defined by its mean and variance. Unless stated otherwise, θ is finite and strictly positive.

In (4) and elsewhere in the article, the stochastic variable u_i reflects the intrinsic difficulty in communication on attribute i , which plays a role regardless of the sequence of individuals through which the information on attribute i gets communicated. Thus, an interpretation of a larger θ is that the business of the organization is such that the communication is intrinsically more difficult. For example, in many ways, communication is intrinsically more difficult in a multinational corporation than, say, in a corporation of which all of the activities are confined within a particular city. For brevity, we refer to a larger θ as implying *noisier communication*.

When A sends a report to B on attribute i then, intuitively, the communication error should tend to be larger if: (i) the intrinsic difficulty in communication is greater, or (ii) A and B are more dissimilar. $u_i \alpha(d_i^{AB})$ in (4) is a simple and transparent formalization of the preceding two aspects. For example, a larger θ implies a larger probability that the value of u_i will exceed any given finite number. In turn, a larger u_i implies that the communication error in (4) is larger. This reflects the first aspect of communication error that we just noted. Now consider the second aspect. Recalling (1), a greater dissimilarity between A and B means a larger expertise gap d_i^{AB} , which, from (3), implies a larger error intensity, $\alpha(d_i^{AB})$. In turn, this increases the communication error in (4). Changes in these two aspects of errors can arise from many sources. For example, innovations in technologies for communication can reduce the intrinsic difficulty of communications, for example, because of the lowered costs of written, verbal, and video communication. In contrast, such innovations are likely not to have a direct influence on the errors that arise from the dissimilarities among individuals. These errors are plausibly affected by other forces; as an example, we later discuss the rise of business education.

Note that, when reporting to B , any two individuals with expertise type A generate the same error $u_i \alpha(d_i^{AB})$. Communication errors that vary across individuals of the same expertise type imply unobserved heterogeneity across individuals. An example of this is if d_i^{AB} is different for two different pairs of individuals, even though each pair consists of an individual with expertise type A reporting to an individual with expertise type B . We do not model such heterogeneity because it will complicate the analysis without adding to the insights, given that our focus is on communications across different expertise types.

When an individual acquires information on an attribute directly, we say that he obtains a reading from nature. It makes sense to posit that the superspecialist- i has the ability to obtain the initial reading on attribute i . We assume that the initial reading of an attribute obtained by the superspecialist of that attribute is free of errors. That is, the initial reading of the random variable X_i obtained by a superspecialist- i is x_i , which is the realized value of X_i . Further, it is reasonable to posit that the initial reading of a superspecialist on an attribute is of no use if he does not have expertise on that attribute. That is, the initial reading of attribute j by a superspecialist- i is of no value if $i \neq j$.

We assume for now that individuals who are not superspecialists cannot obtain initial readings or, equivalently, their initial readings are too uninformative to be meaningful. This simplifies the analysis by significantly reducing the number of combinations that we need to consider. Also, this assumption is natural if specific knowledge is a byproduct of the production process (see Jensen and Meckling, 1992). In Section 5, we illustrate some of the additional tradeoffs that arise when this assumption is relaxed.

We assume that individuals can only report the information that they have received from other individuals or from nature. That is, individuals have no discretion concerning the reports that they send. This simplification is reasonable because we assume that all individuals share the same objective function. Hence, there are no meaningful strategic communication issues in our model.⁸

□ **Organization design.** We define an organization as a group of individuals who have access to all of the n attributes of a given project. By access, we mean the ability to send or receive reports on one or more of the attributes of the project. The organization also makes decisions that affect the utilities of all of its members.

The organization design problem consists of deciding (i) how many individuals should be in the organization, (ii) which member of the organization has what expertise type, (iii) who reports to whom on which attributes, and (iv) who has the formal decision-making authority. We call each possible solution to the organization design problem a *structure*. We use the symbol s to denote a structure.

Definition 1. An organizational structure consists of:

- (i) a set of members. Every element of this set is a subset of N ;
- (ii) a decision maker who has the formal authority over decisions; and
- (iii) a reporting sequence for each of the attributes of the project, $i = 1$ to n . For attribute i , this set describes the sequence of members through which the information on attribute i is communicated, originating from a superspecialist- i (who obtains the initial reading from nature), and eventually being received by the decision maker. If the initial reading is not obtained on an attribute, then the corresponding set is not defined.

□ **Organizational goals.** Our framework can potentially be applied to many different organizational problems. To highlight its usefulness, while still keeping the analysis transparent, we focus on a simple decision problem, in which the organization has to decide whether to accept or reject a single project.

If the project is accepted (i.e., if $p = 1$), the fixed cost is c , which is finite and nonnegative. Otherwise (i.e., if $p = 0$), there is no fixed cost. The *ex post* payoff depends on the values of all n attributes. We adopt a linear technology in which the payoff is the sum of the values of all of the attributes of the project. The *ex post* payoff is⁹

$$\pi(p) \equiv \begin{cases} -c + \sum_{i=1}^n x_i, & \text{if } p = 1. \\ 0, & \text{if } p = 0. \end{cases} \quad (5)$$

□ **Timing.** The timing of the decision process is as follows. There is no prior communication among the members of the organization. First, each superspecialist in the organization observes the realized value of the attribute on which he has expertise. He then sends reports to a subset of the members of the organization. The organizational structure determines who communicates to whom, of which an example is the subset just mentioned. An individual receiving reports from superspecialists may send reports to another subset of the members of the organization. A member of the latter subset may send reports to another subset of members, and so on. After

⁸ Individuals might intentionally distort messages in equilibrium if their preferences do not coincide; see Crawford and Sobel (1982).

⁹ In Section 5, we discuss some additional considerations that arise in a more general version of (5), in which different attributes have different values. Specifically, let v_i denote the value of attribute i . Then the summation in (5) becomes $\sum_{i=1}^n v_i x_i$.

all reporting activity is completed, the decision maker decides whether to accept or reject the project.¹⁰

□ **Optimal structures.** Consider the organizational structure s . The structure influences the information set that the decision maker has. Let $\rho(s)$ denote this information set after the decision maker has received all of the reports. Given his information set $\rho(s)$, the decision maker accepts or rejects the project by maximizing the expected *ex post* payoff. If the decision maker accepts the project, this payoff is

$$P(\rho(s)) \equiv E \left[-c + \sum_{i=1 \text{ to } n} x_i \mid \rho(s) \right] \tag{6}$$

We call $P(\rho(s))$ the *expected payoff*. The decision maker accepts the project only if $P(\rho(s)) > 0$.

We call the stage before reports are collected or sent to the decision maker the *ex ante stage*. From the *ex ante* perspective, ρ is stochastic. For a given structure s , the *ex ante* expected profit is

$$\Pi(s) \equiv E[P(\rho(s)) \mid P(\rho(s)) > 0] \cdot \Pr[P(\rho(s)) > 0] \tag{7}$$

In the expectation in the above right-hand side, ρ unconditionally takes all possible values. For brevity, we refer to $\Pi(s)$ in (7) as the *organizational profit* and, often, only as the *profit*. The optimization problem is to find a structure $s \in S$ that maximizes the profit, (7), where S is the set of all feasible organizational structures.

Recalling Definition 1, a part of the organization design problem is how many individuals should be in the organization, and who reports to whom. Hence, to compare different structures, we need some criteria to choose among those structures that lead to the same profit, but with different numbers of members and reports. A natural way to do this is to incorporate costs of adding members and that of additional reporting. For simplicity, we abstract from these costs. However, in many cases, we assume away structures that have redundant members, that is, members who do not convey useful information or who do not improve upon the existing communication of information. This roughly is like a *lexicographic criterion*. That is, first find the structures that maximize the profit, ignoring the number of members and reports. Then, among the preceding structures, have a bias toward those with fewer members and reports. We abstract from other types of costs, such as the costs of acquiring information and the costs of delay.

To avoid unnecessary repetition, we state here two immediate implications of the above lexicographic criterion. We use these two implications throughout the article.

- (i) If the decision maker does not receive a report on attribute i , then an initial reading is not obtained on this attribute.
- (ii) Any two individuals, A and B , communicate with each other on attribute i only if both have expertise on this attribute, that is, only if $i \in A \cap B$.

3. Some analysis of flat hierarchies

■ Recall that a primary objective of this article is to use simple and tractable models to highlight some of the economic tensions between expertise and communication. We begin with a highly stylized class of structures. There are two layers in such a structure. The upper layer consists only of the decision maker. The composition of the lower layer is determined endogenously, as we shall see below. We refer to any such structure as a *flat hierarchy*.¹¹

In the first subsection below, we describe the profit of a flat hierarchy headed by an individual who has an arbitrarily given expertise breadth. We then present several qualitative properties of this profit, including those with respect to the level of business unpredictability (σ) and with

¹⁰ There are presently no constraints on the number of reports that an individual can send or receive. We consider such constraints later in Section 5.

¹¹ Our terminology here is in line with the existing literature on hierarchies (see, e.g., Rajan and Wulf, 2006).

respect to the noisiness of communication (θ). In this subsection, we also present some selected steps in the calculation and analysis of the profit. We have selected these steps for their economic meaning. Additionally, they help bring out the intuitions of several of the results presented later. The proofs of the steps and the results presented in this subsection, and in the rest of the article, are given in the Appendix.

In the next subsection, we start by comparing two flat hierarchies. These two are headed by individuals with different expertise breadths. Then we present an analysis of the optimal flat hierarchy, and of some of its properties, especially with respect to organizational complexity (n). We conclude by discussing some of the implications of our analysis.

□ **Profit of a flat hierarchy.** Let A denote the decision maker. Let k denote his expertise breadth. That is, $\#A = k \leq n$. Unless stated otherwise, we assume that $k \geq 2$, which is required to ensure that there are two levels in the hierarchies under consideration here. Without loss of generality, we can state that the expertise type of A is $(1, \dots, k)$. Thus, the organizational structure of a flat hierarchy is described by its decision maker's expertise breadth k . Hence, for notational ease, we express $\rho(s)$ as $\rho(k)$.

Recall that the initial readings are done by superspecialists. Hence, superspecialists will constitute the lower level of the hierarchy under consideration here. Further, a report sent by a superspecialist- i is useful to A if and only if $i \in (1, \dots, k)$. Hence, the lower level of the hierarchy will consist of k superspecialists, who are superspecialist-1 to superspecialist- k . Thus, in this section, the set of feasible structures is $S = \{2, \dots, n\}$, where each element $k \in S$ describes a flat hierarchy headed by a generalist with expertise breadth k , which has k superspecialists at the lower level.

For brevity, define

$$\alpha_k \equiv \alpha(k - 1). \tag{9}$$

Hence, from (1), (4), and (9), the reading received by A from superspecialist- i , where $i = 1$ to k , is

$$r_i \equiv x_i + u_i \alpha_k. \tag{10}$$

A 's assessment of X_i , which is the random value of attribute i , is as follows. His prior is that $X_i \sim \mathcal{N}(\mu, \sigma^2)$. This prior, and his reading (10) of the report that he receives from the superspecialist- i , determine his posterior of X_i . The expected value of this posterior of X_i , given r_i , is

$$E[X_i | r_i] = \mu + \beta_k(r_i - \mu), \text{ where} \tag{11}$$

$$\beta_k \equiv \sigma^2 / (\sigma^2 + \theta^2 \alpha_k^2). \tag{12}$$

One can think of β_k as the weight that the decision maker attaches to the reading (10). A larger θ , implying noisier communication, leads to a lower weight. A larger α_k , implying a larger error intensity, also leads to a lower weight. Further, recall that a larger σ implies that the organization faces greater business unpredictability. In this case, the importance of the reading (10) increases. Accordingly, the decision maker attaches a larger weight, β_k , to this reading.

The information that A has at this stage is $\rho(k) \equiv (r_1, \dots, r_k)$, which is the set of reports that he has received from his subordinates. Now consider those attributes for which the decision maker does not receive any report. For these attributes, he uses his priors. The expected value that he imputes to each of these attributes is $E[X_i] = \mu$. We combine this with (11), and recall the definition (6) of the expected payoff, P . This yields

$$P(\rho(k)) \equiv -c + E \left[\sum_{i=1 \text{ to } n} X_i | \rho(k) \right] = n\mu - c + \beta_k \sum_{i=1 \text{ to } k} (r_i - \mu). \tag{13}$$

This is the expected payoff if the project is accepted, conditional upon the structure $s = k$, and on the information set ρ that the decision maker has at this stage.

The decision maker accepts the project if and only if $P(\rho(k))$ is positive. Let $\Psi [P(\rho(k))]$ denote the cumulative distribution function of $P(\rho(k))$, where ρ unconditionally takes all possible values. Then, the profit is

$$\Pi(k) \equiv \int_{P(\rho(k))>0} P(\rho(k))d\Psi [P(\rho(k))]. \tag{14}$$

Here, the argument k of Π is merely a shorthand to denote that this flat hierarchy is headed by a manager with expertise breadth k . Let $\phi(\zeta) \equiv \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\zeta^2}$ denote the probability density of a unit normal distribution. Let $\Phi(\zeta') \equiv \int_{\zeta \leq \zeta'} \phi(\zeta)d\zeta$ denote the corresponding cumulative distribution function. Then, as derived in the Appendix, an explicit expression for (14) is

$$\Pi(k) = (n\mu - c) \left[1 - \Phi \left(\frac{c - n\mu}{a(k)} \right) \right] + a(k)\phi \left(\frac{c - n\mu}{a(k)} \right), \text{ where} \tag{15}$$

$$a(k) \equiv \sigma \sqrt{k\beta_k} \tag{16}$$

is the standard deviation of $P(\rho(k))$, where ρ unconditionally takes all possible values. Expression (15) is superficially opaque. Yet, as we will see below, its properties are economically intuitive.

Organizational surplus. A useful notion to introduce here is that of organizational surplus. This is the extra profit produced by the organization, compared to the profit when there is no organization. Let Π^0 denote the profit in the hypothetical situation in which there is no organization. In this situation, the only information available is that $X_i \sim \mathcal{N}(\mu, \sigma^2)$, for $i = 1$ to n . If the project is undertaken, the profit is $n\mu - c$. Thus,

$$\Pi^0 \equiv \begin{cases} n\mu - c, & \text{if } n\mu - c > 0. \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

To avoid unnecessary details, we assume that $n\mu - c \neq 0$.

Accordingly, the organizational surplus is $\Delta(k) \equiv \Pi(k) - \Pi^0$, where $\Pi(k)$ is given by (15). Without an organization, there is no information collection or communication, and the decision is based on prior knowledge only. Thus, an organization can be viewed as a real option to accept or reject the project only after obtaining and processing information through the organization. The exercise price of this option is zero because there is no gain or loss if the project is not accepted. The value of such an option is nonnegative. Hence, we conjecture that the organizational surplus should be nonnegative. Further, for the organization to be meaningful (and hence, to be needed), the surplus should be strictly positive for at least one value of k . We address these conjectures at the end of this subsection.

Some properties of the organizational profit and of the organizational surplus. We now present and interpret some qualitative properties of the organizational profit, $\Pi(k)$. Each of these properties also holds for the organizational surplus, $\Delta(k)$.¹² Before presenting these properties, we go through two intermediate steps that potentially add to our economic understanding.

The first intermediate step is: how is the profit affected by the volatility of the expected payoff? Recall from (16) that this volatility is represented by $a(k)$, which is the standard deviation of the expected payoff, P . We obtain: *the organizational profit is increasing in the volatility of the expected payoff. That is,*

$$\partial \Pi(k) / \partial a(k) > 0. \tag{18}$$

To interpret the above, recall the earlier observation that an organization can be viewed as a call option, of which the underlying random variable is the expected payoff, P . The *ex post*

¹² This is because the parameters (σ, θ, k) under consideration here do not affect Π^0 , which is defined in (17).

outcomes of this option are thus naturally convex in P . The expected value of these outcomes is the profit $\Pi(k)$. Hence, the profit increases with the volatility of P .

For the second intermediate step, recall from (12) that β_k denotes the weight that the decision maker attaches to the readings that he receives. For the sake of interpretation, we momentarily treat this weight as a parameter. Then, (16) and (18) yield the following intuitive implication: *The organizational profit is larger if the decision maker attaches a larger weight to the readings that he receives. That is,*

$$\partial \Pi(k) / \partial \beta_k > 0. \quad (19)$$

We now assess the effects of (σ, θ) on $\Pi(k)$. Expressions (12) and (19) yield:

Property 1. The organizational profit is larger if the business unpredictability is greater. That is,

$$\partial \Pi(k) / \partial \sigma > 0. \quad (20)$$

Property 2. The organizational profit is larger if the communication is less noisy. That is,

$$\partial \Pi(k) / \partial \theta < 0. \quad (21)$$

The above properties are quite intuitive. If the organization's business unpredictability is greater (i.e., σ is larger), then the values of the attributes are less known in advance. Hence, the organization contributes more, and its profit is larger. The same is true if there is less noise in communication; i.e., if θ is smaller. In this case, there is greater usefulness of communication between any two individuals with different expertise types.

We next consider how k (which is the expertise breadth of the decision maker) affects the organizational profit.

Property 3. The organizational profit increases with the decision maker's expertise breadth if the business unpredictability is extreme, or if the noise in communication is negligible. That is,

$$\text{If } k' > k, \text{ then } \Pi(k') > \Pi(k) \text{ if } \sigma \rightarrow \infty, \text{ or if } \theta \rightarrow 0. \quad (22)$$

This property can be interpreted as follows. If the organization's business unpredictability is extreme (i.e., $\sigma \rightarrow \infty$), then any additional information is valuable, provided that it is better than pure noise. Given this, having one more superspecialist at the lower level can only add to the profit of a flat hierarchy. Hence, in this extreme case, a flat hierarchy headed by a supergeneralist will dominate a flat hierarchy headed by a decision-maker with any other expertise breadth. The same conclusion holds if the noise in communication is negligible; i.e., if $\theta \rightarrow 0$. In this case, the error-intensity loses its significance for a different reason, namely, that there is no communication noise to begin with.

Some properties of the organizational surplus. The result below confirms our previously-stated conjectures concerning the organizational surplus.

Property 4. The organizational surplus approaches zero if the business unpredictability is negligible, or if the noise in communication is extreme. Otherwise, the organizational surplus is strictly positive. That is,

$$\Delta(k) \rightarrow 0 \text{ if } \sigma \rightarrow 0, \text{ or if } \theta \rightarrow \infty; \text{ otherwise } \Delta(k) > 0. \quad (23)$$

□ **Comparing flat hierarchies.** We thus far have considered a flat hierarchy headed by a decision maker with expertise breadth k . Now consider a different flat hierarchy headed by a decision maker with expertise breadth l . To avoid unnecessary details, $k \neq l$. Also, l is arbitrary, except that $2 \leq l \leq n$. The entire analysis contained in the previous section applies to the latter hierarchy, with k substituted by l . By comparing these two hierarchies, we obtain the proposition

stated below. This proposition serves as the basis for several of the results presented later in the article.

Proposition 1 (Comparison of flat hierarchies). Consider two flat hierarchies headed by decision makers with respective expertise breadths k and l . The former is better if and only if

$$\frac{k}{l} > \frac{\sigma^2 + \theta^2 \alpha_k^2}{\sigma^2 + \theta^2 \alpha_l^2} \tag{24}$$

If the inequality is in the opposite direction, then the hierarchy headed by the decision maker with expertise breadth l is better. If, instead, there is an equality in (24), then the relative performance of the two hierarchies is determined by a comparison of their respective number of members, namely $k + 1$ and $l + 1$. There are other results in the rest of the article that involve inequalities similar to that in (24). Depending on the context, analogous interpretations apply to those results.

Optimal flat hierarchies. The profit (15) is defined for $k = 2$ to n , where n is finite. Hence, there exists a value of k that maximizes (15). Let K denote the optimal value of k . We then have the following.

Corollary 1 (Optimal flat hierarchies). In an optimal flat hierarchy, the decision maker has expertise breadth K if and only if

$$K > k \frac{\sigma^2 + \theta^2 \alpha_k^2}{\sigma^2 + \theta^2 \alpha_K^2}, \text{ for } k = 2 \text{ to } n \text{ and } k \neq K. \tag{25}$$

We express the optimal value of k as $K(\sigma, \theta, n)$, to indicate its dependence on some of the parameters. The computation of this optimum will generally require a pairwise comparison of the optimand (15) for expertise breadths $k = 2$ to n . Without doing this, however, we obtain the following comparative statics results.

Proposition 2. The optimal expertise breadth of the decision maker is nondecreasing in business unpredictability and business complexity, and is non-increasing in the noisiness of communication. That is,

$$K(\sigma', \theta, n) \geq K(\sigma, \theta, n) \quad \text{if} \quad \sigma' > \sigma, \tag{26}$$

$$K(\sigma, \theta', n) \geq K(\sigma, \theta, n) \quad \text{if} \quad \theta' < \theta, \quad \text{and} \tag{27}$$

$$K(\sigma, \theta, n') \geq K(\sigma, \theta, n) \quad \text{if} \quad n' > n. \tag{28}$$

Note that Property 3 is a highly special case of (26) and (27). This property requires extreme values of parameters, specifically that $\sigma \rightarrow \infty$ or that $\theta \rightarrow 0$. Expressions (26) and (27) hold for all values of (σ, θ) . Further, in these expressions, the optimal K can be any number from 2 to n . Property 3, given the extreme values of (σ, θ) , implies that $K = n$.

Let $\Pi^*(\sigma, \theta, n)$ denote the profit of the optimal flat hierarchy. That is, $\Pi^*(\sigma, \theta, n)$ is $\Pi(k)$ when $k = K(\sigma, \theta, n)$. Then we have the following.

Property 5. The profit of an optimal flat hierarchy is nondecreasing in organizational complexity. That is,

$$\Pi^*(\sigma, \theta, n') \geq \Pi^*(\sigma, \theta, n) \text{ if } n' > n. \tag{29}$$

□ **Some implications.** We now briefly discuss how the results presented in the previous subsections relate to some stylized observations concerning organizations.

The rise of the modern corporation is closely linked to the increased importance of generalist managers as decision makers, as described in the seminal contributions of Berle and Means (1932) and Chandler (1962). Properties 1, 2, and 5 may be interpreted as saying that the value

of managerial organizations increases with innovations that increase business unpredictability (σ increases) or complexity (n increases), or improve communication technologies (θ decreases). Changes in these parameters may also explain the recent rise in demand for CEOs with generalist skills (Custodio, Ferreira, and Matos, forthcoming).

A useful, although imperfect, distinction in organizations is that between production workers and managers. In our model, an analogy of production workers is as superspecialists. This is because they obtain the initial readings from nature. If there were no initial readings, then there would be no justification for an organization. The “skill-biased technological change” literature has shown that the importance of nonproduction workers in manufacturing has been increasing over time.¹³ The increasing demand for nonproduction workers is compatible with technological changes that facilitate communication or increase business unpredictability or complexity. Such changes are thus compatible with a rise in managerialism.

Proposition 2 suggests that the decision maker’s expertise breadth increases with innovations that increase business unpredictability or complexity or improve communication technologies. There is some evidence that the degree of task specialization in some firms has been decreasing (seen partly in increased multitasking) and that this decrease is contemporaneous with improvements in communication technologies. For evidence of such trends, see Bresnahan, Brynjolfsson, and Hitt (2002) and Caroli and Van Reenen (2001).¹⁴

4. Middle managers

■ In this section, we study several aspects of middle managers. We begin by describing and analyzing a fairly large class of organizational structures. In these structures, any number and type of managers are organized, in virtually any manner, in any number of organizational layers. The second subsection presents some relationships between an individual’s expertise breadth and his rank in optimal multilayered organizational structures. In the third subsection, we examine some circumstances under which adding one or more middle managers is or is not useful. In the last subsection, we present some observations on how changes in exogenous parameters might affect the role of middle managers.

□ **Arbitrarily organized multilayered structures.** Here we consider all possible structures that satisfy two of the implications, stated in (8), of the lexicographic criterion. Thus, there are no preconditions the number of managers, the number hierarchical layers, or the reporting relationships, excluding only what is meaningless *a priori*. Let the set S denote all such structures.

Depending on the structure, its decision maker need not receive reports on all n attributes of the project. Consider a structure in which the decision maker receives reports on k attributes, where $2 \leq k \leq n$. Then, without loss of generality, we can state that he receives reports on attributes $(1, \dots, k)$. In this structure, a superspecialist-1 to a superspecialist- k will, respectively, obtain the initial readings on each of these attributes, which they will communicate to whoever are their immediate superiors. Let $S(k)$ denote the subset of S such that, in each structure within the subset $S(k)$, the decision maker receives reports on k attributes. Thus, $S = \{S(1), \dots, S(n)\}$. Let $s(k)$ denote an element of $S(k)$.

Now consider a structure $s(k)$. For brevity in the rest of this subsection, we mostly suppress the argument k of s . For $i = 1$ to k , let $M(s, i)$ denote the sequence of individuals through which the reading on attribute i gets communicated, originating from a superspecialist- i and eventually being received by the decision maker. Define

$$M(s, i) \equiv \{m_1(s, i), \dots, m_{J(s,i)}(s, i)\}, \text{ for } i = 1 \text{ to } k. \quad (30)$$

¹³ See Berman, Bound, and Griliches (1994), Autor, Katz, and Krueger (1998), Machin and Van Reenen (1998), and Berman, Bound, and Machin (1998).

¹⁴ See Möbius and Schoenle (2007) for an alternative theory of decreasing specialization.

In the above sequence, $m_1(s, i)$ is a superspecialist- i . The number $J(s, i)$ is the number of individuals in the sequence. This number will generally be different for different structures and attributes. The individual $m_{J(s,i)}(s, i)$ is the decision maker. This individual is the same for $i = 1$ to k . The number of middle managers who are intermediating the communication of information on attribute i is $J(s, i) - 2$.

With the above description, we can replicate the steps (10)–(16) for determining the profit of s . The reading on attribute i that the decision maker receives is

$$r(s, i) = x_i + u_i \alpha(s, i), \text{ where} \tag{31}$$

$$\alpha(s, i) \equiv \sum_{j=1 \text{ to } J(s,i)-1} \alpha(|\#m_{j+1}(s, i) - \#m_j(s, i)|). \tag{32}$$

For the decision maker, the expected value of attribute i is

$$E[X_i | r(s, i)] = \mu + (r(s, i) - \mu)b(s, i), \text{ where} \tag{33}$$

$$b(s, i) \equiv \sigma^2 / (\sigma^2 + \lambda(s, i)) \tag{34}$$

is the weight he attaches to his reading on attribute i , and

$$\lambda(s, i) \equiv \theta^2 \alpha(s, i)^2 \tag{35}$$

is the variance of the error in communication on attribute i . The profit of structure $s(k)$ is

$$\Pi(s(k)) = (n\mu - c) \left[1 - \Phi \left(\frac{c - n\mu}{a(s(k))} \right) \right] + a(s(k))\phi \left(\frac{c - n\mu}{a(s(k))} \right), \text{ where} \tag{36}$$

$$a(s(k)) \equiv \sigma \sqrt{\sum_{i=1 \text{ to } k} b(s(k), i)} \tag{37}$$

is the standard deviation of the expected payoff P .

It is readily seen that the flat hierarchies analyzed in Section 3 are highly special cases of the above structures. If the decision maker in a flat hierarchy has expertise breadth k , then $M(s, i)$ has only two elements. The first element, $m_1(s, i)$, is a superspecialist- i , implying that $\#m_1(s, i) = 1$. The second element, $m_2(s, i)$, is the decision maker. Hence, $\#m_2(s, i) = k$. Further, recall the definitions (9) and (12). Thus, (34) and (35) reduce to $b(s, i) = \beta_k$. Hence, recalling the definition (37), $a(s(k))$ equals $a(k)$, which is defined in (16). Thus, (36) reduces to (15).

We now return to the more general structures, of which the profit is (36). Recall the properties (18), (20), and (21) of the profit of a flat hierarchy. The corresponding expressions for (36) are

$$\partial \Pi(s(k)) / \partial a(s(k)) > 0, \partial \Pi(s(k)) / \partial \sigma > 0 \text{ and } \partial \Pi(s(k)) / \partial \theta < 0. \tag{38}$$

In Section 3, we had also presented intuitions underlying the properties (18), (20), and (21). These intuitions apply to (38) as well, even though the structures here have middle managers, organized in considerably arbitrary manners.

In the rest of this subsection, we examine some consequences of optimality on multilayered hierarchies. Here, optimality means a maximization of (36) across all possible structures. Our focus here is on some qualitative properties that optimality induces on multilayered hierarchies.

Let $\Pi^*(\sigma, \theta, n)$ denote the profit of an optimal multilayered hierarchy. This is the resulting profit when (36) is maximized over all $s(k) \in S(k)$ and over all $S(k) \in S$. Let $K(\sigma, \theta, n)$ denote the corresponding value of k . That is, $K(\sigma, \theta, n)$ is the expertise breadth of the decision maker in the optimal multilayered hierarchy just described. To avoid unnecessary details, we assume that K is unique. Also, depending on the context, we suppress one or more arguments of Π^* and K . Then,

$$K(n') \geq K(n) \text{ and } \Pi^*(n') \geq \Pi^*(n) \text{ if } n' > n. \tag{39}$$

That is, the profit of the multilayered hierarchy, and the expertise breadth of its decision maker, are nondecreasing in business complexity. Thus, (39) is the analog of (28) and (29) in the present context.

□ **Expertise and rank in optimal multilayered structures.** In this subsection, we examine some aspects of who will report to whom. In the introductory section, we highlighted that an important issue in organizational economics has been to understand what kinds of individuals will be assigned to the different levels of hierarchies. As will be seen below, our perspective on this is different from, but complementary to, those in the literature. Another purpose of this subsection is as follows. In the multilayered hierarchies described thus far, there are numerous possible structures. It is thus difficult to visualize any aspect of the optimal structure, because finding it requires pairwise comparisons of different structures. This task is simplified considerably by the following result.

Proposition 2. In an optimal structure, if A reports to B directly, or indirectly through an arbitrary sequence of managers, then

$$\#B > \#A. \quad (40)$$

A part of the above proposition is that if A reports directly to B , then B has a larger expertise breadth than A . It follows then that (i) *superspecialists do not receive any reports*, and (ii) *supergeneralists do not send any reports*.

Proposition 3 implies that the function α on the right-hand side of (32) now simplifies to $\alpha(\#m_{j+1}(s, i) - \#m_j(s, i))$. An economic implication of this proposition is that *regardless of the number or the nature of layers in an optimal hierarchy, subordinates have smaller expertise breadths than their respective superiors*. In other words, the breadth of expertise of managers decreases monotonically from the top to the bottom.

□ **Optimal induction of middle managers.** We now study another implication of optimality for middle managers. Consider a manager A located at any level of a hierarchy, except at its top. Let B be the manager, or any one of the managers, to whom A reports directly. Manager B can be, but need not be, the decision maker. In this subsection, we analyze some of the circumstances under which it is, or is not, desirable to induct a middle manager C in between A and B .

For use below, we define *marginal error intensity*. This is the increase in the error intensity, α , if there is a unit increase in the expertise gap between the recipient and the sender of information. Using (3), $\alpha(d+1) - \alpha(d) > 0$ is the marginal error intensity at any given level, d , of the expertise gap. The marginal error intensity is increasing in the expertise gap if the function α is convex, and it is decreasing if α is concave.¹⁵

Next, note from Proposition 3 that a necessary condition for C to be inducted between A and B is that $\#B > \#C > \#A$. For brevity, we do not repeat this condition in the rest of this subsection; we take this condition to be satisfied whenever C is inducted. We then have the following.

Proposition 4. Consider a manager A who reports directly to B . The induction of a middle manager C , in between A and B , is desirable if and only if the marginal error intensity is increasing in the expertise gap.

We now interpret the above result and briefly discuss some of its economic implications. Increasing marginal error intensity can equivalently be stated as

$$\alpha(\#B - \#C) + \alpha(\#C - \#A) < \alpha(\#B - \#A). \quad (41)$$

¹⁵ The analysis presented in this subsection can be further generalized by using the notion of superadditivity of α , instead of its convexity, and the notion of subadditivity of α , instead of its concavity. This distinction is not relevant in our specification, because $\alpha(0) = 0$.

The right-hand side of (41) is the error intensity if A reports directly to B . The left-hand side is the error intensity in the communication from A to B when C is inducted in between. In an economic sense, therefore, we can think of a middle manager partly as an intermediary, who receives information from his immediate subordinates and then translates it and communicates it to his immediate superiors. As is intuitive, and (41) shows, this translation is valuable if the marginal error intensity is increasing.

In contrast, if the marginal error intensity is decreasing, then the middle managers' translations worsen the communication. Concretely, Proposition 4 implies the following result.

Corollary 2. If the marginal error intensity is decreasing in the expertise gap, then a flat hierarchy is better than any multilevel hierarchy.

To our knowledge, the economic properties of error intensity have not been empirically studied. Our analysis suggests that some of these properties partly influence the presence and location of middle managers in multilayered hierarchies, as well as the number of layers in such hierarchies. We briefly discuss this possibility in the next subsection.

Recall our earlier observation that any one-dimensional concept, no matter what its name or definition, cannot explain the assignment of different individuals to different levels in hierarchies. The analysis presented in this section reiterates this concretely. For example, suppose that A reports to B , and that $\#B > \#A + 1$. Then, an individual C with expertise breadth in between those of A and B (i.e., $\#B > \#C > \#A$) will not necessarily be placed as a middle manager between A and B . In fact, C may not have any role at all in this hierarchy.

□ **Some remarks on the flattening of hierarchies.** A recent empirical literature shows that firm hierarchies are becoming flatter (e.g., Guadalupe and Wulf, 2010; Rajan and Wulf, 2006). Two trends have been noted: a broadening of the spans of control, and a reduction in the number of layers in hierarchies. Our analysis provides some perspectives on these phenomena.

First consider the span of control. The flat hierarchies, discussed in Section 3, become flatter if there is an increase of the decision maker's optimal expertise breadth K . Proposition 2 thus implies such flattening, and an increase in the span of control, if there is greater business unpredictability or complexity, or if the communication is less noisy. Rajan and Wulf (2006) consider advances in information technology (along with increased competition and improvements in corporate governance) as one potential explanation for the flattening of hierarchies. Our analysis is consistent with this explanation, if advances in information technology make communication less noisy. To our knowledge, the results that greater business unpredictability, or greater complexity, increases the span of control are unavailable in the literature.¹⁶

Next, consider the possibility that less noisy communication (i.e., a smaller θ) flattens a multilayered hierarchy. Our analysis yields some limited intuition that supports this possibility. Suppose that θ decreases to $\theta \rightarrow 0$ whereas, earlier, it was nontrivially positive. Then, the superspecialist can send error-free reports directly to the decision maker. Hence, from the lexicographic criterion, the middle managers will get eliminated. Thus, as θ approaches zero, hierarchies will have fewer layers.

We now present a different perspective on the flattening of multilayered hierarchies. Suppose that exogenous changes, such as improved business education and training, make generalist managers better at understanding more specialized information. One interpretation of this is that the marginal error intensity of such managers rises less rapidly with the expertise gap. As the analysis in the previous subsection suggests, this will have a tendency to reduce the role of middle managers and, hence, to reduce the numbers of layers in a multilayered hierarchy. Now recall the historical growth in business education that we noted in the introductory section. Our analysis

¹⁶ The latter result also applies, as the first part of (39) shows, to the span of control of the decision maker in multi layered hierarchies.

thus suggests a possible link between this growth and the empirically observed flattening of hierarchies.

5. Some extensions and final remarks

■ As we emphasized in the introductory section, we have thus far worked with a simple model. Our motivations for this include: (i) to highlight what is new in this article; (ii) to focus on some specific questions concerning managerial organizations, including hierarchical assignments of heterogeneous individuals; and (iii) to obtain some qualitative results that connect, or can potentially connect, to the empirical literature. Among our abstractions, the one that appears to us as particularly strong is the emphasis on the influence of the individuals' relative expertise breadths on the quality of communications among them. In the last three subsections below, we briefly summarize three other aspects of individuals that affect this quality. In the first subsection, we add an additional dimension to the heterogeneity of organizations.

We recognize that there are other potentially important generalizations, and also that we have not touched upon many research topics that naturally arise from our perspective.¹⁷ We hope that future research will shed light on these matters.¹⁸

□ **Another dimension of heterogeneity across organizations.** Here we generalize the preceding analysis in that different organizations deal with projects that differ in the nature and values of their attributes. Consider a particular organization, and replace the first part of the right-hand side of (5) by

$$\pi(p) \equiv -c + \sum_{i=1 \text{ to } n} v_i x_i \quad \text{if } p = 1, \quad (42)$$

where v_i is the value of attribute i . We first briefly summarize some immediate implications of (42) using flat hierarchies. We then highlight some of its broader aspects.

The flat hierarchies in Section 3 entail a special case of (42) where $v_i = 1$ for all i . Hence, if the decision maker's expertise breadth is $k \leq n - 1$, it did not matter there which k attributes these are. This will now no longer be the case. Without loss of generality, we arrange the index i such that $v_1 \geq v_2 \geq \dots \geq v_n$. For brevity, we assume that all of the preceding $n - 1$ inequalities are strict. Let A have expertise type $\{1, \dots, k\}$, where $k \leq n - 1$. Let B have any other expertise type but the same expertise breadth k . Then, A will be better as a decision maker than B . In other words, suppose that the flat hierarchy under consideration here is not headed by a superspecialist. Then, optimality requires that he should have expertise on the attributes that have larger values compared to those of the other attributes of the project.

Recall Proposition 1. Its analog, for the flat hierarchy under consideration here, is to compare two decision makers with respective expertise breadths l and k ; the same reinterpretation applies to the rest of this paragraph. Define $V_k \equiv \sum_{i=1 \text{ to } k} v_i^2$ as a metric of the valuation of the expertise breadth k . Then, Proposition 1 holds, with (24) replaced by $V_k/V_l > (\sigma^2 + \theta^2 \alpha_k^2)/(\sigma^2 + \theta^2 \alpha_l^2)$. On the preceding left-hand side, the valuation of the decision maker's expertise breadth now plays the role that was earlier played by his expertise breadth. Corollary 1 changes analogously. Proposition 2 remains unchanged. Thus, the "generalist-ness" is now being additionally traded off against the relative values of different attributes.

Now consider the following heterogeneity across organizations. Let $h = 1$ to H denote different organizations. Let v_i^h denote the value of attribute i for organization h , where $i = 1$ to

¹⁷ An example of such topics is that, in large publicly listed, widely held companies, the current top managers often play a central role in selecting their potential and actual successors. The nature of expertise of these managers thus likely influences the companies' intertemporal trajectories and how these trajectories interact with the changing environment. Sah (1991) and Sah and Stiglitz (1991) have examined how human fallibility of heterogeneous individuals influences some of these dynamics, while abstracting from their expertise types.

¹⁸ We thank two anonymous referees for suggesting all of the topics analyzed in the subsections below.

I describes the union of all project attributes across organizations, and $v_i^h = 0$ if attribute i is not potentially relevant for organization h . This description incorporates many kinds of differences across organizations, for example, a seemingly similar attribute may have different values for two organizations located in different countries.

Recall our earlier observation that any one-dimensional concept cannot explain the assignment of different individuals to different levels in hierarchies. We illustrated this observation in Section 4 on middle managers. Here it gets illustrated in structures as simple as flat hierarchies. For example, consider two flat hierarchies, in each of which k is the optimal expertise breadth of the decision maker. Then, these two decision makers can have very different expertise types, depending on the values of v_i^h in these two hierarchies. Having illustrated this, we use (5) instead of (42) for brevity in the rest of this section.

□ **Initial readings.** In the previous sections, only superspecialist- i has the capability to obtain the initial reading of attribute i . A more general formulation is

- (i) an individual with any expertise type can potentially obtain the initial reading of attribute i , provided that he has expertise on attribute i , and
- (ii) the initial reading of attribute i is less accurate if the reader's expertise breadth is larger. (43)

This immediately introduces the tradeoff that a generalist's initial reading on an attribute is less accurate than that of a superspecialist but the error intensity in this generalist's communication (of his initial readings) to his superior may be smaller. As an illustration of this tradeoff, consider a flat hierarchy s headed by a supergeneralist. Hence, each of the n superspecialists in s obtains the initial reading on the attribute on which he has expertise, and then communicates it to the supergeneralist. Suppose that $n \geq 3$. Consider another structure s' that differs from s only in that a generalist with expertise on attributes $\{1, 2\}$ collects the initial readings on these two attributes and communicates them to the supergeneralist. Thus, in comparing s and s' , we can focus solely on the information that the supergeneralist receives concerning attributes $i = 1$ and 2 .

Let $x_i + w_i \gamma(k)$, with $w_i \sim \mathcal{N}(0, 1)$, be the (stochastic) initial reading on attribute i that an individual collects if his expertise breadth is k and if his expertise type includes attribute i . From (43), $\gamma(k') > \gamma(k)$ if $k' > k$. Our assumption concerning the superspecialist's initial readings implies that $\gamma(1) = 0$. Then, for attributes $i = 1$ and 2 , the reading that the supergeneralist in s' receives is $x_i + w_i \gamma(2) + u_i \alpha(n - 2)$. The corresponding reading in s is $x_i + u_i \alpha(n - 1)$. Thus, the variances of the readings received in s' and s , respectively, are $(\gamma(2))^2 + \theta^2(\alpha(n - 2))^2$ and $\theta^2(\alpha(n - 1))^2$. The variance in s' is larger because it contains $\gamma(2) > 0$, which captures the inaccuracies in the initial readings collected by a generalist. As a tradeoff, the variance in s' is smaller because, from (3), $\alpha(n - 2) < \alpha(n - 1)$. The structure with a smaller variance will have a larger profit.

□ **Span of communication.** For many reasons, such as constraints on time and attention, an individual may not be able to send or receive more than a limited number of reports. Let W denote this limit. That is,

$$\text{an individual can send or receive no more than } W \text{ reports.} \tag{44}$$

In flat hierarchies, no individual sends more than one report, and none receives more than n reports. Thus, (44) has no effect on the analysis in Section 3 if $W \geq n$. Hence, in this subsection, we assume that $W \leq n - 1$. For brevity, we also assume that $W \geq 3$.

Incorporating (44) does not alter our results in Section 3, with the modification that W now plays the role that n did earlier. A key new effect of (44) is that a flat hierarchy will now not collect or communicate any information on $n - W$ attributes. This consequence, in itself, suggests that (44) will enhance the potential usefulness of middle managers. There will, of course, be other

considerations that will determine the overall advantage of middle managers. We illustrate these observations with the following simple example, in which $n > 2W$.

Consider a flat hierarchy s in which the decision maker has expertise on attributes $(1, \dots, W)$. Consider a three-layered structure s' that is headed by a supergeneralist, has n superspecialists at the lowest level, and has two middle managers who have expertise, respectively, on attributes $(1, \dots, W)$ and $(W + 1, \dots, n)$. That is, s' contains s and, in addition, contains another flat hierarchy headed by a middle manager, and the two middle managers report to the supergeneralist. We noted earlier in this subsection that (44) adds to the potential usefulness of middle managers. As an illustration, consider $\theta \rightarrow 0$; that is, the noise in communication is negligible. Without (44), the lexicographic criterion will eliminate the middle managers. In contrast, given (44), the middle managers will strictly be preferred, to counter the constraint (44). Accordingly, s' will have a larger profit than s .

□ **Breadth versus depth.** We have thus far abstracted from the costs of expertise. These costs will depend upon, among other things, the nature of the managerial labor market. One objective of this subsection is to illustrate how our model can incorporate some of these costs. Another objective is to explore the idea of the depth of expertise. Let managers B and B' have expertise on the same set of attributes. Then, in the analysis presented thus far, B and B' are identical. We now consider an extension in which such two otherwise-identical individuals can have different depths of expertise. The aspect of depth that we highlight, using flat hierarchies, is that an individual with a greater depth contributes less to the error intensity when he communicates with others.

Let B denote the decision maker, where $\#B \geq 2$. Let $\delta > 0$ denote his depth of expertise. Let A denote one of the superspecialists in this flat hierarchy. Then, the error intensity generated when A reports to B on attribute i is

$$\alpha(d_i^{AB}/\delta), \quad (45)$$

where, as before, (1) describes the expertise gap, d_i^{AB} .

Expression (45) immediately suggests the need to consider the cost of expertise depth. This is because, if depth were costless, the decision maker would have infinite expertise depth; that is, $\delta \rightarrow \infty$. From (3), (4), and (45), this will imply perfect communication. We therefore posit that the cost of employing a decision maker, denoted by $f(k, \delta)$, is increasing in (k, δ) , that is, in his breadth as well as in his depth. Let $C_0 > 0$ denote the fixed budget that can be spent on employing the decision maker. Then, for any given k , the depth $\delta(k)$ is implicitly defined by $f(k, \delta) = C_0$, and $\delta(k)$ is decreasing in k . The slope of $\delta(k)$ is the technical rate of substitution between k and δ . Equivalently, the magnitude of this slope is the price of a unit increase in breadth, where the depth is the numeraire. For brevity, consider the linear approximation $\delta(k) = \delta_{\max} - \omega(k - 2)$, even though our qualitative conclusions do not depend on linearity.¹⁹

In the previous sections, an increase in the decision maker's expertise breadth from k to $k + 1$ has two opposite effects on the profit Π . The first effect is that the decision maker receives information on one more attribute, namely the $(k + 1)$ st attribute. This increases Π . The second effect is that the larger k increases the expertise gap between the decision maker and each of the $(k + 1)$ superspecialists. This reduces Π . In the present subsection, a third effect arises. An increase in k reduces the expertise depth of the decision maker. This reduces Π .

Finally, consider the interpretation of the slope parameter ω as the price of a unit increase in the decision maker's breadth. If ω is larger, then a unit increase in k increases Π by a smaller amount, or decreases it by a larger amount. Hence, as the tradeoff between the expertise breadth and depth becomes more costly, the organizational profit, as well as the expertise breadth of the decision maker, decrease.

¹⁹ The parameter $\delta_{\max} > 0$ is the maximum possible depth, because $k \geq 2$. The parameter $\omega > 0$ is the magnitude of the slope of $\delta(k)$. The parameters (δ_{\max}, ω) are such that $\delta(k) > 0$ for the relevant values of k . Also note that the analysis presented in the previous parts of this article can be viewed as based on a special case of the linear approximation, where $\delta_{\max} \rightarrow 1$ and $\omega \rightarrow 0$.

Appendix

This Appendix contains all proofs and derivations not included in the main text.

□ **Derivation of expressions and results in Section 3.**

Derivation of expressions (11)–(13). Without loss of generality, let $A \equiv (1, \dots, k)$. For $i \in A$, individual A receives one observation, (10), from $\mathcal{N}(x_i, \theta^2 \alpha_k^2)$, where x_i is the realization of X_i . Individual A knows $\theta^2 \alpha_k^2$, but does not know x_i . A 's prior is that $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Hence, from DeGroot (1970), $E[X_i | r_i] = (\frac{\mu}{\sigma^2} + \frac{r_i}{\theta^2 \alpha_k^2}) / (\frac{1}{\sigma^2} + \frac{1}{\theta^2 \alpha_k^2})$. This simplifies to (11). For $j \notin A$, $E[X_j] = \mu$. Combining this with (11), $P(\rho(k)) \equiv -c + E[\sum_{i=1}^n X_i | \rho(k)]$ yields (13).

Derivation of expression (15). We first calculate the cumulative distribution function of $P(\rho(k))$, where ρ unconditionally takes all values. Using (10), the unconditional value of r_i is $X_i + u_i \alpha_k$. Using this, define

$$Y_i \equiv r_i - \mu = X_i + u_i \alpha_k - \mu. \tag{A1}$$

This expression and (13) yield

$$\Pr[P(\rho(k)) \leq \eta] = \Pr\left[n\mu - c + \beta_k \sum_{i=1}^n Y_i \leq \eta\right] = \Pr\left[\frac{\beta_k \sum_{i=1}^n Y_i}{a} \leq \frac{\eta - a'}{a}\right], \tag{A2}$$

where $a \equiv \sigma \sqrt{k \beta_k}$ (for brevity, we suppress the argument k of a in (16)) and $a' \equiv n\mu - c$. To avoid unnecessary details, we assume $a' \neq 0$ throughout this appendix. The variable $\sum_{i=1}^n Y_i$ is a normal random variable because, from (A1), it is a sum of normal random variables. Because (A1) yields $E[Y_i] = 0$, we have that $E[\sum_{i=1}^n Y_i] = 0$. Definition (A1) also implies that the variance of Y_i is $V[Y_i] = \sigma^2 + \theta^2 \alpha_k^2$, which, using (12), is σ^2 / β_k . This yields $V[\sum_{i=1}^n Y_i] = k\sigma^2 / \beta_k = a^2 / \beta_k^2 \Rightarrow V[(\beta_k \sum_{i=1}^n Y_i) / a] = 1$. Thus, we have that $[(\beta_k \sum_{i=1}^n Y_i) / a] \sim \mathcal{N}(0, 1)$. Hence, from (A2),

$$\Psi[P(\rho(k))] \equiv \Pr[P(\rho(k)) \leq \eta] = \int_{-\infty}^{\frac{\eta - a'}{a}} \phi(\zeta) d\zeta = \Phi\left(\frac{\eta - a'}{a}\right). \tag{A3}$$

Recall that Φ denotes the cumulative distribution function of a unit normal distribution and ϕ denotes its corresponding probability density. From (A3), $d\Psi[P(\rho(k))] = \frac{1}{a} \phi(\frac{\eta - a'}{a}) d\eta$. This and (14) yield

$$\Pi(k) = \frac{1}{a} \int_{\eta > 0} \eta \phi\left(\frac{\eta - a'}{a}\right) d\eta. \tag{A4}$$

We obtain the integral in (A4) by first defining $\eta' \equiv (\eta - a') / a$ and then integrating by substitution. This implies $\eta = a' + a\eta'$, and $d\eta = a d\eta'$. The lower limit $\eta > 0$ becomes $\eta' > -a' / a$. Thus, (A4) yields $\Pi(k) = \frac{1}{a} \int_{-a'/a}^{\infty} (a' + a\eta') \phi(\eta') a d\eta' = a' \int_{-a'/a}^{\infty} \phi(\eta') d\eta' + a \int_{-a'/a}^{\infty} \eta' \phi(\eta') d\eta'$, or

$$\Pi(k) = a' \left[1 - \Phi\left(-\frac{a'}{a}\right)\right] + a \int_{-a'/a}^{\infty} \eta' \phi(\eta') d\eta'. \tag{A5}$$

To integrate the second term in (A5), we use the fact that $\phi(\zeta) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\zeta^2}$ implies

$$\frac{d\phi(\zeta)}{d\zeta} = -\frac{\zeta}{\sqrt{2\pi}} e^{-\frac{1}{2}\zeta^2} = -\zeta \phi(\zeta), \tag{A6}$$

so that we can rewrite the second term in (A5) as

$$a \int_{-a'/a}^{\infty} \eta' \phi(\eta') d\eta' = -a \int_{-a'/a}^{\infty} \frac{d\phi(\eta')}{d\eta'} d\eta' = -a[\phi(\infty) - \phi(-a'/a)] = a\phi(-a'/a). \tag{A7}$$

This and (A5) yield

$$\Pi(k) = a' \left[1 - \Phi\left(-\frac{a'}{a}\right)\right] + a\phi\left(-\frac{a'}{a}\right), \tag{A8}$$

which establishes (15).

Derivation of expressions (18)–(21). We use the standard properties of unit normal variates that $\partial\Phi(\zeta) / \partial\zeta = \phi(\zeta)$ and $\partial\phi(\zeta) / \partial\zeta = -\zeta \phi(\zeta)$. Thus, (A8) yields $\partial\Pi(k) / \partial a = -a'(\frac{a'}{a^2})\phi(-\frac{a'}{a}) + \phi(-\frac{a'}{a}) + a(\frac{a'}{a})(\frac{a'}{a^2})\phi(-\frac{a'}{a})$. Upon simplification,

$$\partial\Pi(k) / \partial a = \phi\left(-\frac{a'}{a}\right) > 0, \tag{A9}$$

which establishes (18). Expression (19) follows from (16) and (18). Expressions (20) and (21) follow from (12) and (19).

Derivation of expressions (22) and (23). Throughout this article, unless stated otherwise, $\sigma > 0, \theta > 0$, and σ and θ are both finite. To prove (22) and (23), we consider some extremes of σ and θ . For proving (22), we treat k as continuous even though it is discrete; this entails no loss of generality. From (12) and (16),

$$\sigma \rightarrow 0, \text{ or } \theta \rightarrow \infty, \text{ implies that } a \rightarrow 0. \tag{A10}$$

$$\sigma \rightarrow \infty, \text{ or } \theta \rightarrow 0, \text{ implies that } \partial a / \partial k = \frac{\sigma}{2} k^{-\frac{1}{2}} > 0. \tag{A11}$$

From (A9) and (A11), $\partial \Pi(k) / \partial k > 0$ if $\sigma \rightarrow \infty$ or if $\theta \rightarrow 0$. This yields (22). Next, we evaluate the two terms on the right-hand side of (A8), when $a \rightarrow 0$. Note that $a\phi(-a'/a) = 0$ if $a \rightarrow 0$. Next, if $a \rightarrow 0$ and $a' > 0$, then $a'[1 - \Phi(-a'/a)] = a'$. Likewise, if $a \rightarrow 0$ and $a' < 0$, then $a'[1 - \Phi(-a'/a)] = 0$. Further, $a'[1 - \Phi(-a'/a)] = 0$ if $a' = 0$, regardless of the value of a . Hence, from (A8), $a \rightarrow 0$ implies that $\Pi(k) = a'$ if $a' > 0$, and $\Pi(k) = 0$ if $a' \leq 0$. In turn, from (17), $\Pi(k) = \Pi^0$ if $a \rightarrow 0$. Thus, recalling (A10) and the definition of $\Delta(k)$, we have established that $\Delta(k) = 0$ if $\sigma \rightarrow 0$ or if $\theta \rightarrow \infty$. Next, if $\sigma > 0$ and if θ is finite, then $a > 0$, and the sign in (A9) holds. Accordingly, Π^0 is the smallest value of $\Pi(k)$, achieved if $\sigma \rightarrow 0$ or if $\theta \rightarrow \infty$; otherwise, $\Pi(k) > \Pi^0$. This yields (23). Q.E.D.

Proof of Proposition 1. From (13) and (18), $\Pi(k) \geq \Pi(l) \Leftrightarrow a(k) \geq a(l)$. This and (12) yield $\Pi(k) \geq \Pi(l) \Leftrightarrow \sigma \sqrt{k\beta_k} \geq \sigma \sqrt{l\beta_l} \Leftrightarrow k\beta_k \geq l\beta_l \Leftrightarrow k/l \geq \beta_l/\beta_k \Leftrightarrow k/l \geq (\sigma^2 + \theta^2\alpha_k^2)/(\sigma^2 + \theta^2\alpha_l^2)$. Q.E.D.

Proof of Corollary 1. It follows immediately from Proposition 1.

Proof of Proposition 2. Define

$$Q(k, l, \sigma, \theta) \equiv (\sigma^2 + \theta^2\alpha_k^2)/(\sigma^2 + \theta^2\alpha_l^2). \tag{A12}$$

For brevity, we suppress one or more arguments of Q , as needed. From (A12), $\partial Q / \partial \sigma = \{2\sigma\theta^2(\alpha_l^2 - \alpha_k^2)\}/\{(\sigma^2 + \theta^2\alpha_l^2)^2\}$. This yields $sign\{\partial Q / \partial \sigma\} = sign\{\alpha_l^2 - \alpha_k^2\}$. Using this and $\alpha_l^2 - \alpha_k^2 = (\alpha_l + \alpha_k)(\alpha_l - \alpha_k)$, we obtain $sign\{\partial Q / \partial \sigma\} = sign\{\alpha_l - \alpha_k\}$. Using this and the specification that $\alpha_k \equiv \alpha(k - 1)$ is strictly increasing in k , we obtain

$$sign \left\{ \frac{\partial Q(k, l, \sigma, \theta)}{\partial \sigma} \right\} = sign\{l - k\}. \tag{A13}$$

Analogously,

$$sign \left\{ \frac{\partial Q(k, l, \sigma, \theta)}{\partial \theta} \right\} = sign\{k - l\}. \tag{A14}$$

Next, using (A12), the condition (25) for the optimality of K , given (σ, θ) , is

$$K \geq kQ(K, k, \sigma, \theta). \tag{A15}$$

For brevity in the rest of this proof, we suppress the range of k , which is $k = 2$ to n , in (A15) and similar expressions. Let $\sigma' > \sigma$, and let K' denote the optimum corresponding to σ' . From (A13), $\partial Q(K, k, \sigma) / \partial \sigma < 0$ if $k < K$. Thus, $Q(K, k, \sigma) > Q(K, k, \sigma')$ for each $k < K$. This implies $kQ(K, k, \sigma) > kQ(K, k, \sigma')$ for each $k < K$. In turn, from (A15), $K > kQ(K, k, \sigma')$ for each $k < K$. Thus, a value of k smaller than K cannot be optimal with parameters (σ', θ) . Hence, $K' \geq K$. This yields (26). The proof of (27) is analogous.

To prove (28), define $K(n) \equiv \arg \max_k \Pi(k)$, where $k = 1$ to n . Now consider the same maximization but with $k = 1$ to n' , where $n' > n$. All of the previous choices of k are feasible in the new maximization. Hence, if $K(n') \leq n$, then $K(n') = K(n)$. Otherwise, $K(n') > n$, which implies that $K(n') > K(n)$, because $K(n) \leq n$. This yields (28). Q.E.D.

Proof of Property 5. Define $\Pi^*(n) \equiv \max_k \Pi(k)$, where $k = 1$ to n . When $k = 1$ to n' , with $n' > n$, the new feasible set of k is a strict superset of the previous one. Thus, $\Pi^*(n') \geq \Pi^*(n)$. Q.E.D.

Proof of expression (39). The first part of (39) follows from the arguments in the proof of expression (28) in Proposition 2. The second part of (39) follows from the arguments in the proof of Property 5. Q.E.D.

Proof of Proposition 3. We will show that A has no role if

$$\#B \leq \#A. \tag{A16}$$

Let

$$M \equiv \{m_1, \dots, m_J, A, m'_1, \dots, m'_J, B\} \tag{A17}$$

denote the sequence of individuals through which information on attribute i gets communicated, originating from a superspecialist- i (who is m_1), reaching to A , and then reaching to B . A special case of (A17) is if A reports directly to B ;

in this case, the subsequence $\{m'_1, \dots, m'_j\}$ is null. Because $\#m_1 = 1$, it must be that $\#B \geq 2$. This is because if $\#B = 1$, then either m_1 or B is needed; all others in (A17) are not. Using (A16), thus

$$\#m_1 = 1, \#B \geq 2, \text{ and } \#A \geq 2. \tag{A18}$$

Let m_j denote the last element in the subsequence $\{m_1, \dots, m_j\}$ of (A17), such that

$$\#m_j < \#B. \tag{A19}$$

Such an m_j must exist because of (A18); our proof does not change if m_j happens to be m_1 . Then, by definition,

$$\#m_{j+1} \geq \#B. \tag{A20}$$

Our proof does not change if m_{j+1} happens to be A .

Recall the definition (1) of the expertise gap. From (A19) and (A20), $\#m_{j+1} > \#m_j$. Hence, $d_i^{m_j m_{j+1}} = \#m_{j+1} - \#m_j$. Further, from (A19), $d_i^{m_j B} = \#B - \#m_j$. The preceding two expressions, combined with (3) and (A20), yield

$$\alpha(d_i^{m_j B}) \leq \alpha(d_i^{m_j m_{j+1}}). \tag{A21}$$

Thus, a direct communication from m_j to B is no worse than that from m_j to m_{j+1} . Recalling (A17), the subsequence of individuals between m_j and B is

$$\{m_{j+1}, \dots, m_j, A, m'_1, \dots, m'_j\}, \tag{A22}$$

which includes A . Suppose that the subsequence (A22) consisted solely of A , that is, the individual m_{j+1} is individual A , and A reports directly to B . Then, given (A16) and (A21), a direct communication from m_j to B is no worse than having the subsequence (A22) in between. Hence, the subsequence (A22) will be eliminated with or without the use of the lexicographic criterion. The preceding conclusion does not change if, besides A , there are one or more individuals in (A22). Hence, A has no role. *Q.E.D.*

Proof of Proposition 4. Consider the communication from A to B on attribute i . Expression (32) and Proposition 4 imply that $\alpha(s, i) = \sum_{j=1 \text{ to } J(s,i)-1} \alpha(\#m_{j+1}(s, i) - \#m_j(s, i))$. From (34), (35), (37), and the first part of (38), a smaller α increases Π . Hence, the induction of C is desirable if and only if it reduces α .

If A communicates directly to B , then A and B are two consecutive elements of $M(s, i)$. From (32), the contribution of this communication to α is $\alpha(\#B - \#A)$. The corresponding contribution, in the communication from A to B but through C , is $\alpha(\#B - \#C) + \alpha(\#C - \#A)$. Note that we are keeping unchanged all elements of $M(s, i)$ that precede A and those that follow B . Hence, recalling (41), Proposition 5 follows. *Q.E.D.*

Derivation of the results in the first subsection of Section 5. No new arguments are needed to establish (11) and (12); those presented earlier continue to hold. We next establish (15)–(18). For these and other results in the first subsection of Section 5, we take as the starting point the derivation (presented in the Appendix) of the results in Sections 2 and 3. We note here only the modifications that arise from using (42) instead of the first part of the right-hand side of (5).

Let B be a decision maker with $\#B = k$ and with an arbitrarily chosen expertise type. Let $2 \leq k \leq n$. Define

$$V_B \equiv \sum_{i \in B} v_i^2, \tag{A23}$$

$$a \equiv \sigma \sqrt{V_B} \beta_k, \quad \text{and} \quad a' \equiv -c + \mu \sum_{i=1 \text{ to } n} v_i. \tag{A24}$$

Note that (42) implies that now $P(\rho(k)) \equiv -c + E[\sum_{i \in B} v_i X_i \mid \rho(k)]$. Hence, the last part of (13) is replaced by $P(\rho(k)) = a' + \beta_k \sum_{i \in B} v_i (r_i - \mu)$. Using this, we reproduce the steps from (A1) to (A3). Thus, (A2) now becomes $\Pr[P(\rho(k)) \leq \eta] = \Pr[(\beta_k \sum_{i \in B} v_i Y_i)/a \leq (\eta - a')/a]$. Like before, $E[(\beta_k \sum_{i \in B} v_i Y_i)/a] = 0$, because $E[Y_i] = 0$. Next, $V[Y_i] = \sigma^2/\beta_k \Rightarrow V[\beta_k \sum_{i \in B} v_i Y_i] = \sigma^2 \beta_k \sum_{i \in B} v_i^2$. From (A24), this equals a^2 . Hence, $[(\beta_k \sum_{i \in B} v_i Y_i)/a] \sim \mathcal{N}(0, 1)$. Thus, there is no change in the steps from (A3) to (A7), which implies that the organizational profit is given by (A8) but with a and a' defined as in (A24). Condition (A9) also remains unchanged.

Let $A = k$. Choose any B such that $A \neq B$ and $\#B = k$. Recall that $v_1 > v_2 > \dots > v_n$. Hence, $V_A > V_B$ and . Thus, (A9), and the definitions of a in (A24), yield that Π is strictly larger for the hierarchy headed by A than that headed by any B .

Accordingly, in the remaining derivations for the first subsection of Section 5, if a hierarchy is headed by a decision maker with expertise breadth k , then his expertise type is k . Hence, it is no longer necessary to use the definition (A23), where the uppercase subscript in V_B denotes that v_i^2 are being summed over an arbitrary set. Instead, it is clearer to use the definition $V_k \equiv \sum_{i=1 \text{ to } k} v_i^2$, where the lowercase subscript in V_k denotes that v_i^2 are being summed from $i = 1$ to k . We therefore restate (A24) as

$$a \equiv \sigma \sqrt{V_k} \beta_k \text{ and } a' \equiv -c + \mu \sum_{i=1}^n v_i. \tag{A25}$$

Next, recall the derivation of Proposition 1 presented earlier in the Appendix. The only steps that change are that (A9) and (A25) imply that $\partial\Pi(k)/\partial(\sigma\sqrt{V_k\beta_k}) > 0 \Rightarrow \partial\Pi(k)/\partial(V_k\beta_k) > 0$. Thus, (12) and (A12) yield $\Pi(k) \geq \Pi(l) \Leftrightarrow V_k\beta_k \geq V_l\beta_l \Leftrightarrow V_k/V_l \geq Q(k, l)$. This yields $V_k/V_l > (\sigma^2 + \theta^2\alpha_k^2)/(\sigma^2 + \theta^2\alpha_l^2)$.

Finally, we prove Proposition 2. Note that $V_k \geq V_l \Leftrightarrow k \geq l$. We redo the steps beginning with (A15). The step (A15) now is $V_k \geq V_l Q(K, k, \sigma)$. Next, for each $k < K$, $Q(K, k, \sigma) > Q(K, k, \sigma') \Rightarrow V_k Q(K, k, \sigma) > V_k Q(K, k, \sigma') \Rightarrow V_k > V_l Q(K, k, \sigma')$. Hence, a value smaller than K cannot be optimal with parameters (σ', θ) . The rest of the proof of Proposition 2 is the same as that provided earlier. Q.E.D.

References

- AGHION, P. AND TIROLE, J. "Formal and Real Authority in Organizations." *Journal of Political Economy*, Vol. 105 (1997), pp. 1–29.
- AUTOR, D.H., KATZ, L.F., AND KRUEGER, A.B. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics*, Vol. 113 (1998), pp. 1169–1213.
- BAKER, G., GIBBONS, R., AND MURPHY, K.J. "Informal Authority in Organizations." *Journal of Law, Economics, and Organization*, Vol. 15 (1999), pp. 56–73.
- BECKER, G. AND MURPHY, K.M. "The Division of Labor, Coordination Costs, and Knowledge." *Quarterly Journal of Economics*, Vol. 107 (1992), pp. 1137–1160.
- BERLE, A.A. AND MEANS, G.C. *The Modern Corporation and Private Property*. New York: Harcourt, Brace and World, 1932.
- BERMAN, E., BOUND, J., AND GRILICHES, Z. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics*, Vol. 109 (1994), pp. 367–397.
- , —, AND MACHIN, S. "Implications of Skill-Biased Technological Change: International Evidence." *Quarterly Journal of Economics*, Vol. 113 (1998), pp. 1245–1279.
- BERTRAND, M. AND SCHOAR, A. "Managing with Style: The Effect of Managers on Corporate Policy." *Quarterly Journal of Economics*, Vol. 118 (2003), pp. 1169–1208.
- BOLTON, P. AND DEWATRIPONT, M. "The Firm as a Communication Network." *Quarterly Journal of Economics*, Vol. 109 (1994), pp. 809–839.
- BRESNAHAN, T.F., BRYNJOLFSSON, E., AND HITT, L.M. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics*, Vol. 117 (2002), pp. 339–376.
- CAROLI, E. AND VAN REENEN, J. "Skill Biased Organizational Change? Evidence from a Panel of British and French Establishments." *Quarterly Journal of Economics*, Vol. 116 (2001), pp. 1449–1492.
- CASTANHEIRA, M. AND LEPPÄMÄKI, M. "Optimal Information Management: Organizations versus Markets." Working Paper, ECARES and Helsinki School of Economics, 2004.
- CHANDLER, A.D., Jr. *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. Cambridge, Mass.: MIT Press, 1962.
- CRAWFORD, V.P. AND SOBEL, J. "Strategic Information Transmission." *Econometrica*, Vol. 50 (1982), pp. 1431–1451.
- CREMER, J., GARICANO, L., AND PRAT, A. "Language and the Theory of the Firm." *Quarterly Journal of Economics*, Vol. 122 (2007), pp. 373–408.
- CUSTODIO, C., FERREIRA, M.A., AND MATOS, P. "Generalists versus Specialists: Lifetime Work Experience and CEO Pay." *Journal of Financial Economics*, (forthcoming).
- DEGROOT, M.H. *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.
- DESSEIN, W. "Authority and Communication in Organizations." *Review of Economic Studies*, Vol. 69 (2002), pp. 811–838.
- , AND SANTOS, T. "Adaptive Organizations." *Journal of Political Economy*, Vol. 114 (2006), pp. 956–995.
- DEWATRIPONT, M. AND TIROLE, J. "Modes of Communication." *Journal of Political Economy*, Vol. 113 (2005), pp. 1217–1238.
- FERREIRA, D. AND SAH, R.K. "Generalists versus Specialists in Organizations." Working Paper, University of Chicago, 2002.
- GARICANO, L. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, Vol. 108 (2000), pp. 874–904.
- GENEAKOPOLOS, J. AND MILGROM, P. "A Theory of Hierarchies Based on Limited Managerial Attention." *Journal of the Japanese and International Economics*, Vol. 5 (1991), pp. 205–225.
- GUADALUPE, M. AND WULF, J. "The Flattening Firm and Product Market Competition: The Effect of Trade Liberalization on Corporate Policies." *American Economic Journal: Applied Economics*, Vol. 2 (2010), pp. 105–127.
- HARRIS, M. AND RAVIV, A. "Organization Design." *Management Science*, Vol. 48 (2002), pp. 852–865.
- HART, O. AND MOORE, J. "On the Design of Hierarchies: Coordination versus Specialization." *Journal of Political Economy*, Vol. 113 (2005), pp. 675–702.
- HEATH, C. AND STAUDENMAYER, N. "Coordination Neglect: How Lay Theories of Organizing Complicate Coordination in Organizations." In B.M. Staw and R.I. Sutton, eds., *Research in Organizational Behavior*. Greenwich, Conn.: JAI Press, 2000.

- HOLMSTRÖM, B. "Moral Hazard and Observability." *Bell Journal of Economics*, Vol. 10 (1979), 74–91.
- JENSEN, M.C. AND MECKLING, W.H. "Specific and General Knowledge and Organizational Structure." In L. Werin AND H. Wijkander, eds., *Contract Economics*. Cambridge, UK: Basil Blackwell, 1992.
- KANDA, K.C. *Mirza Ghalib*. Elgin, Ill.: New Dawn Press, 2004.
- LANDIER, A., SRAER, D., AND THESMAR, D. "Optimal Dissent in Organizations." *Review of Economic Studies*, Vol. 76 (2009), pp. 761–794.
- LAZEAR, E.P. *Personnel Economics*. Cambridge, Mass.: MIT Press, 1995.
- MACHIN, S. AND VAN REENEN, J. "Technology and Changes in Skill Structure: Evidence from Seven OECD Countries." *Quarterly Journal of Economics*, Vol. 113 (1998), pp. 1215–1244.
- MARSCHAK, J. AND RADNER, R. *Economic Theory of Teams*. New Haven, Conn.: Yale University Press, 1972.
- MÖBIUS, M. AND SCHOENLE, R. "The Evolution of Work." Working Paper, Harvard University, 2007.
- NIEHAUS, P. "Filtered Social Learning." *Journal of Political Economy*, Vol. 119 (2011), pp. 686–720.
- PRASAD, S. "Task Assignments and Incentives: Generalists versus Specialists." *RAND Journal of Economics*, Vol. 40 (2009), pp. 380–403.
- PRAT, A. "Hierarchies of Processors with Endogenous Capacity." *Journal of Economic Theory*, Vol. 77 (1997), pp. 214–222.
- RADNER, R. "Hierarchy: The Economics of Managing." *Journal of Economic Literature*, Vol. 30 (1992), pp. 1382–1415.
- . "The Organization of Decentralized Information Processing." *Econometrica*, Vol. 61 (1993), pp. 1109–1146.
- RAJAN, R.G. AND WULF, J. "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies." *Review of Economics and Statistics*, Vol. 88 (2006), pp. 759–773.
- ROSEN, S. "Authority, Control, and the Distribution of Earnings." *Bell Journal of Economics*, Vol. 13 (1982), pp. 311–323.
- SAH, R.K. "Fallibility in Human Organizations and Political Systems." *Journal of Economic Perspectives*, Vol. 5 (1991), pp. 67–88.
- AND STIGLITZ, J.E. "The Architecture of Economic Systems: Hierarchies and Polyarchies." *American Economic Review*, Vol. 76 (1986), pp. 716–727.
- AND ———. "The Quality of Managers in Centralized versus Decentralized Organizations." *Quarterly Journal of Economics*, Vol. 106 (1991), pp. 289–295.
- STEIN, J.C. "Information Production and Capital Allocation: Decentralized vs. Hierarchical Firms." *Journal of Finance*, Vol. 57 (2002), pp. 1891–1921.
- VAN DEN STEEN, E. "Interpersonal Authority in a Theory of the Firm." *American Economic Review*, Vol. 100 (2010), 466–490.
- VAN ZANDT, T. "Real Time Decentralized Information Processing as a Model of Organizations with Boundedly Rational Agents." *Review of Economic Studies*, Vol. 66 (1999), pp. 633–658.
- VAYANOS, D. "The Decentralization of Information Processing in the Presence of Interactions." *Review of Economic Studies*, Vol. 70 (2003), pp. 667–695.