The Comparative Statics of Optimal Hierarchies*

CHENG CHEN
University of Hong Kong
WING SUEN
University of Hong Kong

May 16, 2017

Abstract. Several types of models of hierarchical decision making share two common properties: the characteristics at different levels of the hierarchy are complementary, but this complementary does not extend beyond adjacent levels. We propose a unified yet simple approach with few assumptions to study comparative statics of decision making in all these models. We use this new approach to study organizational decision making, and show that increased delay cost incentivizes the organization to empower lower level employees more than upper level employees.

Keywords. organizational decision making, knowledge hierarchy; monitoring hierarchy; complementarily; conditional separability; delay

JEL Classification. D21; D23; L22; L23

*We are grateful to Luis Garicano, Robert Gibbons, Ricard Gil, Hongyi Li, Andrew Newman, Dilip Mookherjee, Esteban Rossi-Hansberg, Zhigang Tao and Yanhui Wu for their valuable comments. Financial support from HKU is greatly appreciated.

1. Introduction

Starting with the work of Williamson (1967), economists and management scientists have studied organizational decision making in various types of hierarchical models (e.g., Beckmann 1977; Calvo and Wellisz 1978, 1979; Keren and Levhari 1979; Qian 1994; Garicano 2000; Beggs 2001; Meagher 2003; Garicano and Rossi-Hansberg 2004, 2006; Patacconi 2009; Caliendo and Rossi-Hansberg 2012; Wu 2015). In order to maximize total payoff of a hierarchical organization, the organization has to decide the number of employees in various hierarchical levels (i.e., layers) and the total number of layers of the organization. A common feature among many different versions of hierarchical decision models is that successive layers of the hierarchy exhibit complementarity. Take, for instance, the knowledge hierarchy model of Garicano (2000). When lower-level workers become more knowledgeable, fewer problems are passed on to their supervisors to solve. This reduces the cost of investing in knowledge among higher-level workers because fewer of them are needed. If the value of solving problems increases, a straightforward application of monotone comparative statics (Milgrom and Shannon 1994) would suggest that the firm will decide to hire more knowledgeable workers at all levels of the hierarchy, assuming that the total number of layers in the hierarchy remains fixed. But the total number of layers is endogenous in an optimal hierarchy, and it turns out that the firm's profit is not jointly supermodular (nor quasi-supermodular) in workers' knowledge and the number of layers. Unambiguous comparative statics results of the organizational decision making are therefore not easy to obtain, especially because the depth of a hierarchy, which is a key decision made by the organization, is discrete in nature.

Existing research in hierarchical decision making uses numerical analysis and makes functional form assumptions in order to tackle the problem, and the derivation of the results is usually quite tedious.[1] In this paper, we find that several types of hierarchy models share two common properties in their objective functions which are economically meaningful. We propose a unified yet simple approach with few assumptions to study comparative statics of decision making in all these models. We use this approach to address some classic questions in organizational decision making, such as how productivity or cost of delay affects the organization's decisions on the span of control and the depth of a hierarchy.

Importantly, our approach also offers new and testable implications of organiza-

[1] In knowledge hierarchy models (Garicano 2000; Garicano and Rossi-Hansberg 2004), the distribution of the difficulty (or frequency) of problems is usually assumed to be exponential. Some results obtained in this literature are based on numerical calculations (e.g., Garicano and Rossi-Hansberg 2006, Garicano and Rossi-Hansberg 2012). Chen (2017) adopts a linear function for the monitoring intensity in order to derive analytical results in the monitoring hierarchy model. Acemoglu and Newman (2002) assume an exogenous number of layers for simplicity in their monitoring hierarchy model. Mookherjee (2013) and Garicano and Van Zandt (2013) provide excellent surveys for these hierarchy models.

tional decision making previously not available in the theoretical literature. First, the firm only adds or drops one layer after a small shock to its productivity or delay cost, which we call a property of *one-step jump*. Second, when the firm deletes a layer, its characteristics at all existing layers increase after the reorganization. However, the new characteristic at layer i cannot exceed the old characteristic at layer $i+1$ (before the reorganization). We call this an *accordion property*. These two predictions generate bounds on the change of the number of layers and on the change of characteristics at existing layers that hold well in the data.[2]

As an important extension of our benchmark model, we incorporate delay cost into the knowledge hierarchy model to study how the urgency to produce products affects important organizational decisions such as the depth of the hierarchy and relative empowerment of workers across different layers inside the hierarchy. Business decisions are often time-critical. That is why floor traders are given a lot of leeway to make decisions on the spot. The importance of delay cost for organizational decision making is not confined to the finance sector. Galbraith (1977) found that "after 1964 the problem facing Boeing was not to establish a market but to meet the opportunities remaining as quickly as possible. Now a delay of a few months would result in canceled orders and fewer sales." He reported that "to respond to competitive time pressure from Douglas, Lockheed, and the British-French Concorde, Boeing was forced to drastically reduce the time devoted to product development and design." Whitney (1988) showed that how fast an organization makes their decisions substantially affects its profit and revenue. Furthermore, Rajan and Wulf (2006) explained that one major reason for increased decentralization inside U.S. firms in recent years was to enable "faster decision making and execution." In addition, Bloom, Van Reenen and Sadun (2010) argued that "tougher competition may make local manager's information more valuable, as delays to decisions become more costly." In short, existing evidence suggests that the cost of delay plays a key role in shaping organizational decisions of the firm.

Theoretical literature on how delay cost affects organizational decision making is scant. There is a literature starting from Radner (1992, 1993) that explores the optimal organizational structure that minimizes the processing time for a given task.[3] However, important questions such as how increased demand for fast decision making affects organizational decisions are left unanswered. This motivates us to study the impact of the cost of delay on organizational decisions. Specifically, we assume

[2] Caliendo, Monte and Rossi-Hansberg (2015) and Caliendo et al. (2015) show that most firms that change their number of layers do it by adding or dropping one layer. Furthermore, they show that in firms that remove one layer from the hierarchy, wages of employees at existing layers increase after the reorganization. However, their wages (after the reorganization) are still lower that what their direct supervisors earn before the reorganization. These empirical findings support our theoretical results of one-step jump and the accordion property.

[3] Other papers include Bolton and Dewatripont (1994) and Patacconi (2009).

that workers acquire knowledge and communicate with supervisors in order to solve problems and produce output, which is the same assumption as in Garicano (2000). Different from Garicano (2000), we assume that if a problem is solved by agents higher up in the hierarchy, it causes a greater loss in revenue due to delay in delivering the good. Unsolved problems at the top layer generate losses as well, as the firm fails to honor the contract and deliver the good, which results in punishment fees due to non-performance or in loss of reputation of the firm. The rationale for our specification is that output or sales are affected by how fast an organization carries out the production. Naturally, the slower an organization makes decisions and deliver its products, the more losses it bears.

The key insight from our extension is that increased delay cost (i.e., increased demand for fast decision making) makes the firm less hierarchical and employees—especially lower-level employees—more empowered. In order to prevent costly delay, the firm recruits more knowledgeable workers at all levels so that there are fewer problems waiting to be solved by upper management. As a result, workers at all layers are able to solve more problems and empowered more. Importantly, since delay is cumulative, making agents at lower levels acquire knowledge and solve problems (i.e., empowering them) is relatively more effective in reducing delay than making middle or upper level workers do so. As a result, the firm invests disproportionately on knowledge acquisition by lower level workers, and the range of problems they can solve increases more compared to upper level workers. Therefore, lower level workers are more empowered in the production process. These predictions are starkly different from the effect of improved information and communication technology on relative empowerment and wages across layers.[4]

Our theoretical finding from the extended model helps explain the difference in organizational decision making between Japanese firms and U.S. firms. As Aoki (1989, 1990) documented, workers on the production floor in Japan have higher authority and deal with more complex problems compared with their U.S. counterparts. As a result, they are more empowered. He noted that one key element of Toyota's lean production is to fasten the firm's decision making process and make the firm respond to changes in market environment more rapidly. Our theory shows that if a firm cares more about delay in its production and decision-making process (like in Japan), it should make employees at lower hierarchical levels acquire more knowledge and empower them more (i.e., compared with middle- and high-level managers). Therefore, our model is useful for us understanding the phenomenon of high level of empowerment and authority for Japanese production workers that Aoki (1989, 1990) documented for Japanese firms.

[4] Improved information technology makes all agents learn more without generating an effect on relative empowerment across layers. Improved communication technology empowers agents at upper layers, as they become more effective in communicating with their subordinates.

The remainder of the paper is organized as follows. Section 2 proposes a general model of hierarchies. Section 3 uses a unified yet simple approach to implement comparative statics exercises in the general hierarchy model. Section 3 uses the same approach to derive new and testable implications previously not available in the theoretical literature. Section 5 extends our benchmark model of the knowledge hierarchy to study how increased cost of delay affects relative empowerment and relative wages of workers in different layers. Section 6 concludes.

2. A General Model of Hierarchies

We assume that the profits from a hierarchical firm with $L \geq 2$ layers can be written as:

$$\pi_L(z_1, \ldots, z_L) = g(z_1) + G(z_L) + \sum_{i=2}^{L} f(z_{i-1}, z_i), \tag{1}$$

with $\overline{Z} \geq z_L \geq \ldots \geq z_1 \geq 0$. For $L = 1$, let

$$\pi_1(z_1) = g(z_1) + G(z_1).$$

The maximum number of layers is $\overline{L}$, so the feasible set of L is $\{1, \ldots, \overline{L}\}$. The problem for the firm is

$$\max_{L} \left\{ \max_{z_1, \ldots, z_L} \pi_L(z_1, \ldots, z_L) \right\}. \tag{2}$$

The "characteristics" z_i of each layer i can be given different interpretations depending on the context of the application. The objective function (1) encompasses the leading models of managerial or production hierarchies in the literature. We illustrate the generality of our formulation via some examples.

Example 1. *In the knowledge hierarchy model (Garicano 2000), z_i is the knowledge of level-i workers (a smaller i refers to a lower layer). The wage of a worker with knowledge z_i is $\omega + cz_i$. The distribution of the difficulty of problems is $H(\cdot)$. For each production worker (i.e., level-1 worker), the number of level-i supervisors required to deal with problems not solvable by their subordinates is $\gamma(1 - H(z_{i-1}))$, where $\gamma < 1$ depends on the communications technology. Knowledge acquisition is assumed to be cumulative; so z_i has to be bigger than z_{i-1}. The profit per production worker is:*

$$AH(z_L) - (\omega + cz_1) - \sum_{i=2}^{L} \gamma(1 - H(z_{i-1}))(\omega + cz_i),$$

where A is the unit value of a solved problem (i.e., productivity).

We modify the above model to introduce an explicit delay cost.[5] The real-world ex-

[5] In the original Garicano model, there is an implicit delay cost of wages that need to be paid to each

ample we consider is that a manufacturing firm receives an order from a customer. The firm has to deliver the product to the customer by solving the problem that appears in the production process and doing it on time. If the problem is solved by employees in the bottom layer, the product is delivered on time and there is no delay. However, if the problem is solved by employees in non-bottom layers, there will a punishment fee for the late delivery as passing the problem to upper layers is time-consuming. In particular, we assume that the firm incurs an explicit delay cost of φ every time a problem has to go up by one level. If none of the employees can solve the problem, the firm cannot deliver the product (i.e., zero revenue) and incurs the highest punishment (i.e., $(L-1)\varphi$) due to maximum delay. Based on the above description, the expected profit per production worker is:

$$\pi_L(z_1,\ldots,z_L) = AH(z_L) - (\omega + cz_1) - \sum_{i=2}^{L} (1 - H(z_{i-1}))\,(\varphi + \gamma(\omega + cz_i))\,. \quad (3)$$

This objective function is a special case of (1), with $g(z_1) = -(\omega + cz_1)$, $G(z_L) = AH(z_L)$, and $f(z_{i-1}, z_i) = -(1 - H(z_{i-1}))(\varphi + \gamma(\omega + cz_i))$.

Example 2. *In the monitoring hierarchy model (Calvo and Wellisz 1978), z_i is be the number of workers at level i. Level L represents the production workers, and level 0 is the entrepreneur with $z_0 = 1$. (Note that we reverse the labeling of the levels here.) A worker at layer i loses wage w_i if shirking behavior is detected. This implies an incentive compatibility constraint:*

$$w_i - \psi \geq \left(1 - g\left(\frac{z_i}{z_{i-1}}\right)\right) w_i,$$

where ψ is the disutility of exerting effort, and $g(\cdot) \leq 1$ is the probability of detection, which decreases in the span of control z_i/z_{i-1}. Thus, the least costly incentive compatible wage equals $w_i = \psi m(z_i/z_{i-1}) \equiv \psi/g(z_i/z_{i-1})$. We assume that $m(\cdot)$ is strictly increasing and weakly convex. The profit of the firm is:

$$\pi_L(z_1,\ldots,z_L) = AQ(z_L) - \sum_{i=1}^{L} \psi z_i m\left(\frac{z_i}{z_{i-1}}\right), \quad (4)$$

where A is the unit value of output (i.e., productivity), and $Q(\cdot)$ is total output as a function of the number of production workers. In this example, $g(z_1) = -\psi z_1 m(z_1)$, $G(z_L) = AQ(z_L)$, and $f(z_{i-1}, z_i) = -\psi z_i m(z_i/z_{i-1})$.

Example 3. *In a hierarchy model with multiple production stages, each successive layer of the firm improves the quality of its predecessor product from z_{i-1} to z_i at a strictly convex cost*

higher level employee who gets involved in solving a delayed problem.

$C(z_i - z_{i-1})$. (We define $z_0 = 0$ for the initial stage.) For concreteness, let

$$C(z_i - z_{i-1}) = b_0 + b_1(z_i - z_{i-1}) + b_2(z_i - z_{i-1})^2,$$

where $z_i \geq z_{i-1}$, and b_0, b_1 and b_2 are strictly positive. The firm receives payoff $AB(z_L)$ from selling the final product, where A indexes the productivity of the firm and $B(\cdot)$ is increasing and strictly concave. The profit of the firm is:

$$\pi_L(z_1, \ldots, z_L) = AB(z_L) - K(z_1) - \sum_{i=2}^{L} C(z_i - z_{i-1}), \tag{5}$$

where $K(\cdot)$ is an increasing and strictly convex function representing the cost of producing the initial product. Note that the firm will not make an infinite number of small adjustments since b_0 is strictly positive. It will not make the final product in just one step if b_2 is sufficiently large. In this example, $g(z_1) = -K(z_1)$, $G(z_L) = AB(z_L)$, and $f(z_{i-1}, z_i) = -C(z_i - z_{i-1})$.

In each of the examples above, there is complementarity between the characteristics of successive layers. In the knowledge hierarchy model, there is complementarity between workers' knowledge at level $i-1$ and at level i. When workers at the lower level can solve more problems, investing in knowledge of upper level workers is cheaper, since fewer of them are required. In the monitoring hierarchy model, we can show that $\partial^2 f/\partial z_{i-1}\partial z_i > 0$ if $m(\cdot)$ is weakly convex. When there are more supervisors, monitoring lower level employees is easier, which lowers the efficiency wage of these workers, making it cheaper to employ more lower level employees. Finally, in Example 3, a higher quality level chosen at the previous stage of production reduces the marginal cost of improving quality at the current stage whenever the adjustment cost function $C(\cdot)$ is weakly convex.

We make the following assumption for the general hierarchy model:

Assumption 1. *$f(z_{i-1}, z_i)$ is supermodular in (z_{i-1}, z_i).*

Assumption 1 implies that $\pi_L(\cdot)$ is supermodular in $(z_1, \ldots, z_L)$. Another key feature of the objective function (1) is that $\partial^2 \pi_L/\partial z_i \partial z_j = 0$ for $|i-j| > 1$. An implication of this observation is that the characteristics of the hierarchy above level j are related to the characteristics below level j only through z_j. Conditional on z_j, the optimal structure of the hierarchy above level j can be determined separately from that below level j. We call this property *conditional separability*.

Suppose a firm considers inserting M layers at the bottom of its existing hierarchy without changing the characteristics of the existing L layers. The maximum increase in profit from such a reorganization is:

$$\bar{\delta}^M(z_1) \equiv \max_{y_1,\ldots,y_M} \pi_{L+M}(y_1, \ldots, y_M, z_1, \ldots, z_L) - \pi_L(z_1, \ldots, z_L),$$

where $0 \leq y_1 \leq \ldots \leq y_M \leq z_1$. Note that $\overline{\delta}^M(z_1)$ depends only on the characteristics z_1 at the bottom level but not on $z_2, \ldots, z_L$, thanks to conditional separability. Similarly, suppose a firm adds M layers on top of its existing hierarchy without changing the characteristics of the existing L layers. The maximum increase in profit from such a reorganization is:

$$\overline{\Delta}^M(z_L) \equiv \max_{y_1,\ldots,y_M} \pi_{L+M}(z_1,\ldots,z_L,y_1,\ldots,y_M) - \pi_L(z_1,\ldots,z_L),$$

where $z_L \leq y_1 \leq \ldots \leq y_M \leq \overline{Z}$.

We assume that both $\overline{\delta}^M(\cdot)$ and $\overline{\Delta}^M(\cdot)$ are single-crossing. Although these assumptions are not stated in terms of the primitive functions g, G and f, they are easily verified in practical applications.

Assumption 2. *For any $M \leq \overline{L} - L$ and any $z_1' > z_1$, $\overline{\delta}^M(z_1) \geq 0$ implies $\overline{\delta}^M(z_1') > 0$.*

Assumption 3. *For any $M \leq \overline{L} - L$ and any $z_L' < z_L$, $\overline{\Delta}^M(z_L) \geq 0$ implies $\overline{\Delta}^M(z_L') > 0$.*

Observe that $\overline{\delta}^M(z_1) > 0$ is a sufficient (but not necessary) condition for inserting extra layers at the bottom of the hierarchy to be profitable. Take the knowledge hierarchy model with delay cost, for example. From equation (3), $\overline{\delta}^M(\cdot)$ is equal to the maximum value of

$$-cy_1 + cz_1 - (1 - H(y_M))(\varphi + \gamma(\omega + cz_1)) - \sum_{i=2}^{M} (1 - H(y_{i-1}))\,(\varphi + \gamma(\omega + cy_i))\,.$$

This expression is strictly increasing in z_1 because $(1 - H(y_M))\gamma < 1$. Hence $\overline{\delta}^M(\cdot)$ is single-crossing from below. Intuitively, as production-level workers become more knowledgeable, it is cheaper to elevate them into supervisors and replace them by less knowledgeable workers at the bottom (the communications technology $\gamma < 1$ implies that a firm needs fewer higher-level workers than lower-level ones). Similarly, it is straightforward to verify that Assumption 2 holds in Example 2 when $m(\cdot)$ is weakly convex. In the multi-stage production model of Example 3, we can use the first-order conditions for $y_1, \ldots, y_M$ to show that

$$\frac{\mathrm{d}\overline{\delta}^M(z_1)}{\mathrm{d}z_1} = K'(z_1) - K'(y_1).$$

Because $z_1 > y_1$ and $K(\cdot)$ is strictly convex, $\overline{\delta}^M(\cdot)$ is strictly increasing, which implies that it is single-crossing from below.

Assumption 3 is related to a generalized version of diminishing returns. In Example 1, suppose top-level workers are not very knowledgeable, so that it is profitable to add extra layers to solve their problems (specifically, suppose $\overline{\Delta}^M(z_L) \geq 0$). Then, this strategy remains profitable when top-level workers become even less knowledgeable. In particular, $\overline{\Delta}^M(z_L)$ in the knowledge hierarchy model with delay cost is equal to the maximum value of

$$(1-H(z_L))\left[A - A\frac{1-H(y_M)}{1-H(z_L)} - (\varphi+\gamma(\omega+cy_1)) - \sum_{i=2}^{M}\frac{1-H(y_{i-1})}{1-H(z_L)}(\varphi+\gamma(\omega+cy_i))\right].$$

The expression in brackets is strictly decreasing in z_L. Hence the above expression as a whole is single-crossing from above.

In Example 2, we can use the first-order conditions for $y_1,\ldots,y_M$ to show that

$$\frac{\mathrm{d}\overline{\Delta}^M(z_L)}{\mathrm{d}z_L} = -AQ'(z_L) + A\frac{y_M}{z_L}Q'(y_M) - \frac{1}{z_L}\left(\psi y_1 m\left(\frac{y_1}{z_L}\right) + \sum_{i=2}^{M}\psi y_i m\left(\frac{y_i}{y_{i-1}}\right)\right).$$

When $\overline{\Delta}^M(z_L)=0$, the term in parentheses is equal to $AQ(y_M)-AQ(z_L)$. Thus, when $\overline{\Delta}^M(\cdot)$ crosses zero, we have

$$\frac{\mathrm{d}\overline{\Delta}^M(z_L)}{\mathrm{d}z_L} = \frac{A}{z_L}\left([Q(z_L)-z_LQ'(z_L)] - [Q(y_M)-y_MQ'(y_M)]\right),$$

which is negative because $y_M > z_L$ and the function $Q(z)-zQ'(z)$ is increasing in x whenever $Q(\cdot)$ is concave. The function $\overline{\Delta}^M(z_L)$ is indeed single-crossing from above.

Similarly, in Example 3, we can use the first-order conditions for $y_1,\ldots,y_M$ to show that

$$\frac{\mathrm{d}\overline{\Delta}^M(z_L)}{\mathrm{d}z_L} = -AB'(z_L) + AB'(y_M).$$

Assumption 3 holds as $y_M > z_L$ and $B(\cdot)$ is strictly concave.

3. Comparative Statics

Let there be some parameter t in the objective function (1) such that

$$\pi_L(z_1,\ldots,z_L;t) = G(z_L;t) + g(z_1) + \sum_{i=2}^{L} f(z_{i-1},z_i).$$

Since we are going to compare the characteristics at a layer before and after the reorganization, we need to be specific about the convention adopted to define "the same" layer. For example, layer 2 in a 3-level hierarchy becomes layer 3 in a 4-level hierarchy if the new layer is added from below, but layer 2 in a 3-level hierarchy remains layer 2 in a 4-level hierarchy if the new layer is added from above. More generally, suppose a firm changes from having L layers with characteristics $\mathbf{z}$ to having L' layers with characteristics $\mathbf{z}'$. We say that this firm becomes more hierarchical and its characteristics increase (*under the convention of adding layers from below*) if $L \geq L'$ and $(z'_{L-L'+1}, \ldots, z'_L) \geq (z_1, \ldots, z'_L)$.

Proposition 1. *Suppose Assumptions 1 and 2 hold, and $G(z_L; t)$ is supermodular in (z_L, t). Then a firm becomes more hierarchical and its characteristics increase (under the convention of adding layers from below) with increases in t.*

Proof. The assumptions imply that $\pi_L(z; t)$ is supermodular in (z, t). If the number of layers remains unchanged, then the result follows directly from Milgrom and Shannon (1994).

Next, suppose—contrary to the proposition—that the number of layers decreases from L to $L - j + 1$ ($j \geq 2$) when the parameter increases from t_1 to $t_2 > t_1$. By Assumption 2, $\overline{\delta}^{j-1}(z_j)$ is single-crossing from below in z_j.

Define $\pi^*_L(t) = \max_z \pi_L(z; t)$. Since both $\pi^*_L(\cdot)$ and $\pi^*_{L-j+1}(\cdot)$ are continuous, there exists a $t' \in [t_1, t_2]$ such that $\pi^*_L(t') = \pi^*_{L-j+1}(t')$. Let $\boldsymbol{z} = (z_1, \ldots, z_L)$ be the optimal vector of characteristics at t' when there are L layers. Similarly, let $\tilde{\boldsymbol{z}} = (\tilde{z}_j, \ldots, \tilde{z}_L)$ be the optimal vector of characteristics at t' when there are $L - j + 1$ layers. By conditional separability and revealed preference,

$$\begin{aligned}\pi_L(\boldsymbol{z}; t') &= \pi_{L-j+1}(z_j, \ldots, z_L; t') + \overline{\delta}^{j-1}(z_j; t') \\ &\leq \pi_{L-j+1}(\tilde{\boldsymbol{z}}; t') + \overline{\delta}^{j-1}(z_j; t').\end{aligned}$$

Thus, $\pi_L(\boldsymbol{z}; t') = \pi_{L-j+1}(\tilde{\boldsymbol{z}}; t')$ implies $\overline{\delta}^{j-1}(z_j; t') \geq 0$. Furthermore,

$$\begin{aligned}\pi_{L-j+1}(\tilde{\boldsymbol{z}}; t') &= \max_{y_1, \ldots, y_{j-1}} \pi_L(y_1, \ldots, y_{j-1}, \tilde{z}_j, \ldots, \tilde{z}_L; t') - \overline{\delta}^{j-1}(\tilde{z}_j; t') \\ &\leq \pi_L(\boldsymbol{z}; t') - \overline{\delta}^{j-1}(\tilde{z}_j; t').\end{aligned}$$

Thus, $\pi_{L-j+1}(\tilde{\boldsymbol{z}}; t') = \pi_L(\boldsymbol{z}; t')$ implies $\overline{\delta}^{j-1}(\tilde{z}_j; t') \leq 0$. Assumption 2 then requires that $\tilde{z}_j \leq z_j$.

Because of conditional separability, $(z_{j+1}, \ldots, z_L)$ maximizes $\pi_{L-j+1}(\cdot; t')$ when the bottom characteristic is fixed at z_j, while $(\tilde{z}_{j+1}, \ldots, \tilde{z}_L)$ maximizes $\pi_{L-j+1}(\cdot; t')$ when the bottom characteristic is fixed at $\tilde{z}_j$. Supermodularity of $\pi_{L-j+1}(\cdot; t')$ and the fact that $\tilde{z}_j \leq z_j$ then imply that $\tilde{z}_i \leq z_i$ for $i = j + 1, \ldots, L$. We therefore have:

$$\frac{\mathrm{d}\pi^*_{L-j+1}(t')}{\mathrm{d}t} = \frac{\partial G(\tilde{z}_L; t')}{\partial t} \leq \frac{\partial G(z_L; t')}{\partial t} = \frac{\mathrm{d}\pi^*_L(t')}{\mathrm{d}t},$$

where the inequality follows because $G(\cdot)$ is supermodular. This inequality implies that $\pi^*_{L-j+1}(\cdot)$ must cut $\pi^*_L(\cdot)$ from above whenever t increases until they intersect.

However, by revealed preference, we have

$$\pi^*_{L-j+1}(t_1) - \pi^*_L(t_1) \leq 0,$$
$$\pi^*_{L-j+1}(t_2) - \pi^*_L(t_2) \geq 0.$$

This leads to the result that $t_1 \geq t' \geq t_2$ (a contradiction). Hence L must weakly increase when t increases. Moreover, we have established that $z_j \geq \tilde{z}_j$ at the point when the firm switches from $L - j + 1$ to L layers. Thus, the characteristics at the existing layers cannot decrease with t. ■

The above result helps us derive comparative statics in Examples 1–3. In particular, because the profit function $\pi_L(\cdot)$ is supermodular in (A, z_L), Proposition 1 applies if we take $t = A$. A positive productivity shock causes a firm to expand both vertically and horizontally. The firm changes its organizational decision by increasing the number of layers after the positive shock. Moreover, the knowledge of workers in each layer (in the knowledge hierarchy model) or the number of workers in each layer (in the monitoring hierarchy model) increases when A increases. As a result, workers are empowered more in the knowledge hierarchy model. In Example 3, a positive productivity shock causes the firm to adopt a more complex production process with more stages of production. Moreover, the quality of the intermediate product at each step, as well as quality of the final product at the end, increases when A increases. Our general treatment shows that the three versions of hierarchy models yield qualitatively the same comparative statics result with respect to a productivity shock.

For the next result, let

$$\pi_L(z_1, \ldots, z_L; t) = G(z_L) + g(z_1; t) + \sum_{i=2}^{L} f(z_{i-1}, z_i; t).$$

Suppose a firm changes from having L layers with characteristics $\boldsymbol{z}$ to having L' layers with characteristics $\boldsymbol{z}'$. We say that this firm becomes less hierarchical and its characteristics increase (under the convention of adding layers *from above*) if $L \geq L'$ and $(z'_1, \ldots, z'_{L'}) \geq (z_1, \ldots, z_{L'})$.

Proposition 2. *Suppose Assumptions 1 and 3 hold, $g(z_1; t)$ is supermodular in (z_1, t), and $f(z_{i-1}, z_i; t)$ is decreasing in t and supermodular in (z_{i-1}, z_i, t). Then a firm becomes less hierarchical and its characteristics increase (under the convention of adding layers from above) with increases in t.*

Proof. If the number of layers does not change, then by Milgrom and Shannon (1994), z_i rises with t. Now, assume—contrary to the proposition—that the number of layers increases from j to L when the parameter increases from t_1 to $t_2 > t_1$. Suppose $\pi^*_j(t') =$

$\pi_L^*(t')$ at some $t' \in [t_1, t_2]$. Let $\boldsymbol{z} = (z_1, \ldots, z_L)$ be the optimal vector of characteristics at t' when there are L layers. Similarly, let $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \ldots, \tilde{z}_j)$ be the optimal vector of characteristics at t' when there are j layers. By a similar argument as in the proof of Proposition 1, we can show that $\overline{\Delta}^{L-j}(z_j; t') \geq 0$ and $\overline{\Delta}^{L-j}(\tilde{z}_j; t') \leq 0$. By Assumption 3, we must have $\tilde{z}_j \geq z_j$. By the conditional separability property and supermodularity, this in turn implies that $\tilde{z}_i \geq z_i$ for $i = 1, \ldots, j-1$. We therefore have:

$$\frac{\mathrm{d}\pi_j^*(t')}{\mathrm{d}t} = \frac{\partial g(\tilde{z}_1; t')}{\partial t} + \sum_{i=2}^{j} \frac{\partial f(\tilde{z}_{i-1}, \tilde{z}_i; t')}{\partial t} \geq \frac{\partial g(z_1; t')}{\partial t} + \sum_{i=2}^{L} \frac{\partial f(z_{i-1}, z_i; t')}{\partial t} = \frac{\mathrm{d}\pi_L^*(t')}{\mathrm{d}t}.$$

This inequality implies that $\pi_j^*(\cdot)$ must cut $\pi_L^*(\cdot)$ from below whenever they intersect. However, by revealed preference, we have

$$\pi_j^*(t_1) - \pi_L^*(t_1) \geq 0,$$
$$\pi_j^*(t_2) - \pi_L^*(t_2) \leq 0.$$

This leads to the result that $t_1 \geq t' \geq t_2$ (a contradiction). Hence L must weakly decrease when t increases. Moreover, we have established that the characteristic at level i increases from z_i to $\tilde{z}_i$ at the point when the firm switches from L layers to j layers as t increases. Therefore, z_i cannot decrease with t. ■

The above result help us address a classic question in organizational decisions: how does increased delay cost affect the optimal depth of the organization (Radner 1992, 1993; Radner and Van Zandt 1992; Beggs 2001; Patacconi 2009)? In the knowledge hierarchy model of Example 1, the delay cost parameter φ satisfies the premise of Proposition 2. Hence, when delay cost increases, the organization responds by delayering, while increasing the knowledge of workers and empower them more at all remaining layers.[6] Economically speaking, when delay becomes costly, the firm decides to solve the problem faster. It achieves this goal by letting the problem pass through fewer layers and by having a higher probably of solving it at each layer.[7]

4. Accordion Property and One-Step Jump

Propositions 1 and 2 are silent about the characteristics of the "new" layers that are added to the firm hierarchy as parameters change. For example, suppose the number of layers decreases strictly from L to L' when delay cost rises infinitesimally from φ to φ'. By Proposition 2, the corresponding $z'_{L'}$ is greater than $z_{L'}$. But is it possible that

[6] We note that Propositions 1 and 2 are not trivial corollaries of Milgrom and Shannon (1994), as the overall maximization problem (2) is neither quasi-supermodular in $(z_1, \ldots, z_L, L)$ nor quasi-supermodular in $(z_1, \ldots, z_L, -L)$.

[7] In the knowledge hierarchy model, the communication technology parameter γ does not satisfy the premise of Proposition 2. It is not possible to derive unambiguous comparative statics result for this parameter without imposing further functional form or parametric restrictions on the general model.

$z'_{L'} > z_L$, so that top-level workers after the reorganization become more knowledgeable and are empowered more than top-level workers before the reorganization? The accordion property established below shows that the answer is no.

Proposition 3. *Suppose Assumptions 1, 2 and 3 hold. If both L (with characteristics* $\tilde{\mathbf{z}}$*) and* $L+1$ *(with characteristics* $\mathbf{z}$*) are optimal at some t, and t enters the firm's objective function as defined in Propositions 1 or 2. Then for* $i = 2, \ldots, L$,

$$z_i \in [\tilde{z}_{i-1}, \tilde{z}_i],$$

and $z_1 \leq \tilde{z}_1$ *and* $z_{L+1} \geq \tilde{z}_L$.

Proof. Suppose t enters the firm's objective function as defined in Proposition 2. We first prove that $z_{L+1} \geq \tilde{z}_L$. Suppose the opposite is true, i.e., $z_{L+1} < \tilde{z}_L$. Then, either (a) there exists some $i < L$ such that $z_{i+1} \geq \tilde{z}_i$; or (b) $z_{i+1} < \tilde{z}_i$ for all $i < L$. In case (a), by conditional separability and supermodularity, $z_{i+1} \geq \tilde{z}_i$ implies $z_{L+1} \geq \tilde{z}_L$, a contradiction. In case (b), we have $z_2 < \tilde{z}_1$. Since L is optimal, it is unprofitable to add a layer below $\tilde{z}_1$, which implies $\overline{\delta}^{1}(\tilde{z}_1) \leq 0$. But this implies $\overline{\delta}^{1}(z_2) < 0$, which means profits can be increased by removing the bottom layer if the firm has $L+1$ layers, a contradiction.

Next, we prove that $z_i \geq \tilde{z}_{i-1}$ for all $i \leq L$. Suppose otherwise, and let $z_i < \tilde{z}_{i-1}$ for some $j \leq L$. But then conditional separability and the supermodularity of $\pi_{L-i+2}(\cdot)$ imply that $z_{L+1} < \tilde{z}_L$, which contradicts our first conclusion.

In the proof of Proposition 2, we have already established that $z_i \leq \tilde{z}_i$ for all $i \leq L$ when the firm is indifferent between $L+1$ and L. Thus, $z_i \in [\tilde{z}_{i-1}, \tilde{z}_i]$.

Suppose t enters the firm's objective function as defined in Proposition 1. Then we proceed in an analogous manner by first proving that $z_1 \leq \tilde{z}_1$. This implies that $z_i \leq \tilde{z}_i$ for all i. Finally, Proposition 1 already establishes that $z_i \geq \tilde{z}_{i-1}$. The proposition then follows. ∎

Figure 1 illustrates the accordion property. Suppose a firm switches from having 3 layers to 2 layers at some point t. Proposition 2 requires that $\tilde{z}_i \geq z_i$ for $i = 1, 2$, but imposes no upper bound on $\tilde{z}_i$. By Proposition 3, we establish that $\tilde{z}_i \leq z_{i+1}$. Thus, although the characteristics at any level may jump discontinuously as the parameter t changes, the accordion property imposes a limit on the size of the jump.

Proposition 4 below establishes that if both L and L' are optimal, then (generically) L' cannot differ from L by more than one. An implication of Proposition 4 is that the optimal L is a step function falling (or increasing) by one layer at each step when the parameter in interest changes continuously.

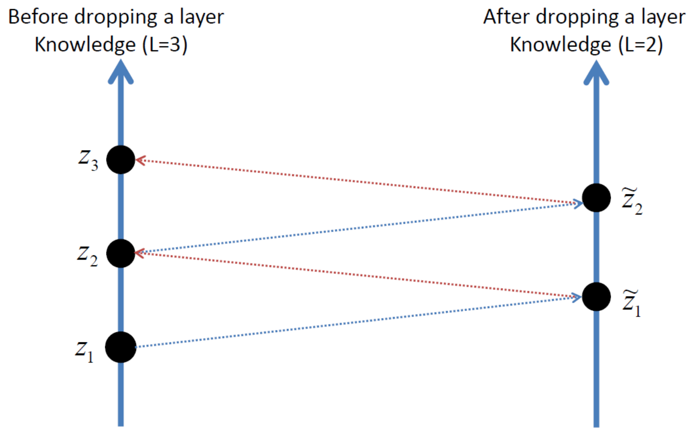

Figure 1. *The firm delayers from $L = 3$ to $L = 2$ at some φ as delay cost increases. Proposition 2 establishes that $\tilde{z}_i \geq z_i$ for $i = 1, 2$. Proposition 3 (the accordion property) establishes that $\tilde{z}_i$ is sandwiched between z_i and z_{i+1}.*

Proposition 4. *Suppose Assumptions 1, 2 and 3 hold. If both L and $L' > L$ are optimal at some t, then, generically, $L' = L + 1$.*

Proof. Let $\tilde{\boldsymbol{z}}$ represent the vector of optimal characteristics corresponding to L, and let $\boldsymbol{z}$ represent the vector of optimal characteristics corresponding to L'. Suppose to the contrary that $L' \geq L + 2$. Apart from the non-generic case of $L' = L + 2$ with $z_i = \tilde{z}_{i-1}$ for $i = 1, \ldots, L$, there are three possibilities.

Case (1). $z_{L'-1} > \tilde{z}_L$. Because L is optimal, we have $\overline{\Delta}^1(\tilde{z}_L) \leq 0$, which implies $\overline{\Delta}^1(z_{L'-1}) < 0$. This contradicts the optimality of L'.

Case (2). $z_2 < \tilde{z}_1$. Because L is optimal, we have $\overline{\delta}^1(\tilde{z}_1) \leq 0$, which implies $\overline{\delta}^1(z_2) < 0$. This contradicts the optimality of L'.

Case (3). Neither (1) nor (2) is true, i.e., $z_2 \geq \tilde{z}_1$ and $z_{L'-1} \leq \tilde{z}_L$. In this case, there are L layers between $\tilde{z}_1$ and $\tilde{z}_L$ under organization structure L, and there are $L' - 2$ layers between $z_{L'-1}$ and z_2 under organization structure L'. Because there are more layers between a narrower range of characteristics in organization L', there must exist $j \in \{2, \ldots, L\}$ such that z_j and z_{j+1} are both contained in the interval $[\tilde{z}_{j-1}, \tilde{z}_j]$. Define

$$D(\bar{z}) \equiv \max_{y_1,\ldots,y_{j-1}} \{\pi_j(y_1, \ldots, y_{j-1}, \bar{z})\} - \max_{y_1,\ldots,y_j} \{\pi_{j+1}(y_1, \ldots, y_j, \bar{z})\}$$

to be the difference in profits if the organization has j layers rather than $j + 1$ layers, conditional on the top layer's having knowledge level $\bar{z}$. Let $\tilde{y}_{j-1}(\bar{z})$ be the optimal characteristic of the level below $\bar{z}$ when the organization has j layers, and let $y_j(\bar{z})$ be the optimal characteristic of the level below $\bar{z}$ when the organization has $j + 1$ layers.

Supermodularity of $\pi_j(\cdot)$ implies that, for all $\bar{z} < \tilde{z}_j$,

$$\tilde{y}_{j-1}(\bar{z}) < \tilde{y}_{j-1}(\tilde{z}_j) = \tilde{z}_{j-1}.$$

Similarly, supermodularity of $\pi_{j+1}(\cdot)$ implies that, for all $\bar{z} > z_{j+1}$,

$$y_j(\bar{z}) > y_j(z_{j+1}) = z_j.$$

Combining these two inequalities gives $\tilde{y}_{j-1}(\bar{z}) < y_j(\bar{z})$ for all $\bar{z} \in [z_{j+1}, \tilde{z}_j]$. By the envelope theorem,

$$D'(\bar{z}) = f_2(\tilde{y}_{j-1}(\bar{z}), \bar{z}) - f_2(y_j(\bar{z}), \bar{z}).$$

Since $f(\cdot)$ is supermodular, $D(\bar{z})$ is strictly decreasing for $\bar{z} \in [z_{j+1}, \tilde{z}_j]$. Now, since organization structure L is optimal, profits cannot be increased by adding a layer below $\tilde{z}_j$ (while keeping everything above layer j fixed). This implies that $D(\tilde{z}_j) \geq 0$. Since $D(\cdot)$ is strictly decreasing in the region $[z_{j+1}, \tilde{z}_j]$, this implies $D(z_{j+1}) > 0$. But then the organization structure L' is not optimal, because profits could be increased by removing one layer below $j+1$ while keeping everything above $j+1$ fixed, which leads to a contradiction. ■

The above proposition shows that when the firm changes its organizational decision by delayering (due to a small shock to its cost), it only reduces the number of layers by one generically. This prediction receives support form data on knowledge hierarchies (Tag 2013; Frederic 2015; Caliendo, Monte and Rossi-Hansberg 2015; Caliendo et al. 2015).

5. Extension

In this section, we provide an extended discussion of two issues related to the knowledge hierarchy model in Example 1. First, we show that our unified approach can be used to solve the organizational design problem regardless of whether knowledge acquisition is assumed to be cumulative or non-cumulative. We emphasize that the comparative statics established before continue to hold even when knowledge acquisition is non-cumulative. Second, we show that the knowledge hierarchy model with an explicit delay cost defined in Example 1 yields a distinctive prediction on how increased delay cost affects relative knowledge acquisition (empowerment) of employees across layers, which has interesting implication for organizational decision making. Because this prediction requires additional assumptions beyond those specified in Assumptions 1–3, we leave the proof of this result to the Appendix.

There are two common versions of knowledge hierarchy models in the literature. In some papers (e.g., Garicano and Rossi-Hansberg 2006) and in Example 1 of our paper, knowledge acquisition is assumed to be cumulative: a supervisor has to invest in learning everything that his subordinates know. For example, if production (level-1) workers solve problems with difficulty $z \in [0, z_1]$ and their supervisors (level-2 workers) solve problems with difficulty $z \in (z_1, z_2]$, these supervisors need to learn everything from 0 to z_2. The wage of level-2 workers is therefore assumed to be an increasing function of z_2. In some other papers (e.g., Garicano 2000; Caliendo and Rossi-Hansberg 2012), knowledge acquisition is non-cumulative: a supervisor does not have to learn what his subordinates know. Using the same example above, level-2 workers only need to invest in learning problems with difficulty between z_1 and z_2, and their wages are an increasing function of $z_2 - z_1$. If we adopt the non-cumulative version of the knowledge hierarchy model, the objective function of the firm is written as:

$$AH(z_L) - (\omega + cz_1) - \sum_{i=2}^{L}(1 - H(z_{i-1}))[\varphi + \gamma(\omega + c(z_i - z_{i-1}))]. \tag{6}$$

The only difference between this equation and equation (3) is that the wage per employee at layer i is $\omega + c(z_i - z_{i-1})$ instead of $\omega + cz_i$, as employees at layer i only need to acquire the *incremental* knowledge stock between layer i and layer $i - 1$. It is easy to verify that this objective function satisfies Assumptions 1 to 3. Therefore, Propositions 1 to 4 continue to hold.

Next, we show the knowledge hierarchy model with explicit delay cost generates an unambiguous prediction on how increased delay cost affects empowerment of workers. In both the cumulative version and the non-cumulative version of the knowledge hierarchy model, workers at level i deal with problems with difficulty in the range $(z_{i-1}, z_i]$. We say that a worker is *more empowered* if the range of problem he deals with increases, i.e., if $z_i - z_{i-1}$ increases (for level-1 workers, we define $z_0 = 0$).

Proposition 5. *Assume the distribution $H(\cdot)$ of the difficulty of problems is the exponential distribution. Suppose the marginal cost of delay φ increases and the firm does not adjust the number of layers. The, regardless of whether knowledge acquisition is cumulative or non-cumulative,*

$$\frac{\partial z_1}{\partial \varphi} > \frac{\partial(z_2 - z_1)}{\partial \varphi} > \ldots > \frac{\partial(z_{L-1} - z_{L-2})}{\partial \varphi} > \frac{\partial(z_L - z_{L-1})}{\partial \varphi} \geq 0.$$

Proof. See the Appendix. ∎

The key finding of Proposition 5 is a distributional effect: more costly delay causes the firm to disproportionately empower its lower-level employees. The firm becomes more knowledge-intensive at all levels of its hierarchy, but the increase in knowledge is relatively greater among lower levels than among higher levels. Furthermore, we show in appendix that $z_1 \leq z_2 - z_1 \leq \ldots \leq z_L - z_{L-1}$. I.e., incremental knowledge stock increases when we move up the hierarchical ladder. Therefore, the *percentage* increase in knowledge acquisition (in both the cumulative case and the non-cumulative case) is also greater among lower levels than among higher levels.

Although we make a distributional assumption to prove the above proposition, the basic intuition for the empowerment effect can be seen from the first-order conditions of the objective function (3) with respect to the characteristics at layer i and at a higher layer $j > i$ (for the version of cumulative knowledge acquisition). The first-order conditions imply that:

$$\frac{h(z_i)}{h(z_j)} \frac{\varphi + \gamma\omega + \gamma c z_{i+1}}{\varphi + \gamma\omega + \gamma c z_{j+1}} = \frac{1 - H(z_{i-1})}{1 - H(z_{j-1})}.$$

The left-hand-side of the above can be interpreted as the marginal rate of substitution of raising z_i relative to raising z_j. Holding knowledge levels constant, the left-hand-side of the above is increasing in φ because $z_{j+1} > z_{i+1}$. In a knowledge hierarchy, the benefit of empowering any given level of workers consists of two parts: (1) a reduction in cost of delay; and (2) a saving in wage cost for supervisors. Since low-level supervisors are cheaper than high-level supervisors, the first part (reduction in delay cost) figures more prominently for low level workers than for high level workers, and an increase in delay cost raises the advantage of empowering low-level workers relative to empowering high-level workers.

The effect of increased delay cost on relative knowledge acquisition and empowerment in our model is starkly different from the effect of improved information and communication technology on these variables in standard knowledge hierarchy models. In these standard models, improved information technology (i.e., a smaller c) makes all agents acquire more knowledge without generating an effect on relative empowerment, while improved communication technology (i.e., a smaller γ) empowers agents at upper layers more, as they become more effective in communicating with subordinates. The introduction of an explicit delay cost into standard knowledge hierarchy models yields economically important and empirically distinct predictions compared to standard models.

6. Conclusion

Our paper provides a stylized representation of hierarchies that is general enough to encompass several existing models of this organizational form. The model emphasizes two key features of hierarchies: complementarity and conditional separability. Complementarity in hierarchies has been pointed out by, among others, Rosen (1982) and Beggs (2001). Without the conditional separability property, however, complementarity itself does not yield unambiguous comparative statics when the number of layers is an endogenous variable. We then use a unified framework to implement comparative statics exercises to derive implications of increased delay cost for organizational decision making. In particular, we show that when the demand for fast decision making increases, the top manager should make the firm less hierarchical and employees should be more empowered. Importantly, employees at lower hierarchical levels should be empowered much more than employees at upper hierarchical levels, as this is the most effective to reduce delay cost. We also derive implications of an increase in firm productivity for organizational decision making such as the depth of the organization.

Garicano and Rossi-Hansberg (2006) exploit the Markovian property of the knowledge hierarchy model to solve the optimal knowledge acquisition inside the hierarchy.[8] This property is essentially the conditional separability property discussed in our paper. We improve on their exercise by showing that this property holds for a class of models and can be used to solve the optimal number of layers inside the hierarchy. Moreover, we utilize this property (and the complementarity property) to derive other comparative statics results of the hierarchy model (namely, one-step jump and the accordion property) which are consistent with the evidence but have not been explicitly proven in the literature before.

Some papers in the literature treat the number of layers as a continuous variable in order to simplify the problem (Keren and Levhari 1979; Qian 1994). However, this treatment yields qualitatively different predictions. See Chen (2017) for a discussion. Chen (2017) also explores different empirical predictions yielded by the knowledge hierarchy model and the monitoring hierarchy model.

[8] See the referral price defined in appendix 1.B of their paper.

Appendix: Proof for Proposition 5

For $i = 2, \ldots, L$, define $y_i \equiv z_i - z_{i-1}$ as the incremental knowledge stock acquired by employees at layer i. For $i = 1$, define $y_1 \equiv z_1$. Also assume that H is the exponential distribution with parameter λ. Note that the exponential distribution is (weakly) log-concave and entails a constant hazard rate, which greatly simplifies the analysis.

Part 1. Cumulative knowledge acquisition. With an exponential distribution, the first-order conditions can be written as:

$$\begin{aligned} \lambda(\varphi + \gamma\omega + \gamma c z_2) &= c e^{\lambda y_1}; \\ \lambda(\varphi + \gamma\omega + \gamma c z_{i+1}) &= \gamma c e^{\lambda y_i}, \quad i = 2, \ldots, L-1; \\ \lambda A &= \gamma c e^{\lambda y_L}. \end{aligned} \tag{7}$$

Since z_i increases in i and $\gamma \leq 1$, the first-order conditions (7) imply:

$$y_1 < y_2 < \ldots < y_{L-1}.$$

Lemma 1. *In an optimal organization structure with cumulative knowledge acquisition, $y_2 \geq 1/\lambda$.*

Proof. Let L be the optimal number of layers and $(z_1, \ldots, z_L)$ be the corresponding optimal knowledge levels. By definition, $\pi_L(z_1, z_2, \ldots, z_L) \geq \pi_{L-1}(z_2, \ldots, z_L)$, which implies that $\bar{\delta}^1(z_2) \geq 0$. Recall that

$$\bar{\delta}^1(z_2) = \max_{\zeta \leq z_2} c z_2 - c\zeta - e^{-\lambda\zeta}(\varphi + \gamma\omega + \gamma c z_2).$$

The optimal ζ^* satisfies the first-order condition:

$$e^{-\lambda\zeta^*}(\varphi + \gamma\omega + \gamma c z_2) = c/\lambda.$$

Thus, $\bar{\delta}^1(z_2) \geq 0$ implies $c z_2 - c\zeta^* - c/\lambda \geq 0$. Since $\zeta^* = z_1$, we have $y_2 = z_2 - \zeta^* \geq 1/\lambda$. ■

Proposition 2 shows that the expressions on the left-hand-side of equation (7) are increasing in φ, except for the last expression, which is constant with respect to φ. We therefore have $\partial y_i / \partial \varphi > 0$ for $i = 1, \ldots, L-1$ and $\partial y_L / \partial \varphi = 0$.

To establish the ranking of the derivatives, we use an induction argument. First, subtracting equation (7) for $i = L-2$ from that for $i = L-1$ yields:

$$e^{\lambda y_{L-1}} = e^{\lambda y_{L-2}} + \lambda y_L. \tag{8}$$

Because y_L is fixed when φ changes, and because $y_{L-1} > y_{L-2}$, this equation implies:

$$\frac{\partial y_{L-2}}{\partial \varphi} > \frac{\partial y_{L-1}}{\partial \varphi}.$$

Next, suppose it is true that $\partial y_{k-1}/\partial\varphi > \partial y_k/\partial\varphi$ for some $k \in \{4, \ldots, L-1\}$. Similar to equation (8), we have the following:

$$e^{\lambda y_{k-1}} = e^{\lambda y_{k-2}} + \lambda y_k. \tag{9}$$

Taking logs of both sides and differentiating with respect to φ gives

$$\frac{\partial y_{k-1}}{\partial \varphi} = \frac{e^{\lambda y_{k-2}}}{e^{\lambda y_{k-2}} + \lambda y_k}\frac{\partial y_{k-2}}{\partial \varphi} + \frac{\lambda y_k}{e^{\lambda y_{k-2}} + \lambda y_k}\frac{1}{\lambda y_k}\frac{\partial y_k}{\partial \varphi},$$

By Lemma 1, $\lambda y_1 \geq 1$, which implies $\lambda y_k > 1$ for all $k \geq 2$. Thus, by the induction hypothesis, we know $\partial y_{k-1}/\partial\varphi > (1/\lambda y_k)\partial y_k/\partial\varphi$. Equation (9) then implies $\partial y_{k-2}/\partial\varphi > \partial y_{k-1}/\partial\varphi$.

Finally, we deal with $\partial y_1/\partial\varphi$. The first two equations in (7) can be combined to obtain:

$$\gamma e^{\lambda y_2} = e^{\lambda y_1} + \gamma\lambda y_3.$$

We can apply the same reasoning as the above to show that $\partial y_2/\partial\varphi > \partial y_3/\partial\varphi$ implies $\partial y_1/\partial\varphi > \partial y_2/\partial\varphi$.

Part 2. Non-cumulative knowledge acquisition. The first order conditions of the objective function (6) with respect to z_i's are:

$$\begin{aligned}
\lambda(\varphi + \gamma\omega + \gamma c y_2) &= c(e^{\lambda y_1} - \gamma);\\
\lambda(\varphi + \gamma\omega + \gamma c y_{i+1}) &= \gamma c(e^{\lambda y_i} - 1), \quad i = 2, \ldots, L-1;\\
\lambda A &= \gamma c e^{\lambda y_L}.
\end{aligned} \tag{10}$$

First, we use the following the lemma to establish the ranking of knowledge acquired by workers at various layers:

Lemma 2. *When L is optimally chosen and when knowledge acquisition is non-cumulative, we have*

$$y_1 \leq y_2 \leq \ldots \leq y_L.$$

Proof. Since L is optimally chosen, we have $\overline{\Delta}^1(z_{L-1}) \geq 0$, where

$$\begin{aligned}\overline{\Delta}^1(z_{L-1}) &= \max_{\zeta \geq z_{L-1}} \left\{ A(1-e^{-\lambda\zeta}) - e^{-\lambda z_{L-1}}\left(\varphi + \gamma(\omega + c(\zeta - z_{L-1}))\right) - A(1-e^{-\lambda z_{L-1}}) \right\} \\ &= e^{-\lambda z_{L-1}} \max_{y \geq 0} \left\{ A(1-e^{-\lambda y}) - (\varphi + \gamma(\omega + cy)) \right\}.\end{aligned}$$

Since $y_L = z_L - z_{L-1}$ is the solution to the above maximization problem, $\overline{\Delta}^1(z_{L-1}) \geq 0$ implies

$$\lambda A \geq \lambda\left(\varphi + \gamma(\omega + c y_L)\right) + \lambda A e^{-\lambda y_L}.$$

Using the first-order conditions (10) with respect to y_L and y_{L-1}, this inequality implies $y_L \geq y_{L-1}$. An induction argument using the first-order conditions (10) for $i = 2, \ldots, L-1$ shows that $y_k \geq y_{k-1}$ implies $y_{k-1} \geq y_{k-2}$. Finally, comparing the first-order conditions (10) for y_2 and y_1 and using the fact that $\gamma < 1$ shows that $y_2 > y_1$. ■

From the first-order condition (10) for y_L, we have $\partial y_L/\partial\varphi = 0$. Proceeding iteratively with the first-order conditions (10) for $y_{L-1}, y_{L-2}, \ldots, y_1$ shows that $\partial y_i/\partial\varphi > 0$ for all $i < L$.

To establish the ranking of the derivatives, we use an induction argument. First, subtracting the first-order condition for y_{L-2} from that for y_{L-1} yields:

$$e^{\lambda y_{L-1}} = e^{\lambda y_{L-2}} + \lambda(y_L - y_{L-1}). \tag{11}$$

Because $y_L - y_{L-1}$ decreases with φ and $y_L > y_{L-1}$, this equation implies:

$$\frac{\partial y_{L-2}}{\partial\varphi} > \frac{\partial y_{L-1}}{\partial\varphi}.$$

Next, suppose it is true that $\partial y_{k-1}/\partial\varphi > \partial y_k/\partial\varphi$ for some $k \in \{4, \ldots, L-1\}$. Similar to equation (11), we have the following equation:

$$e^{\lambda y_{k-1}} = e^{\lambda y_{k-2}} + \lambda(y_k - y_{k-1}). \tag{12}$$

Taking logs of both sides and differentiating with respect to φ gives

$$\frac{\partial y_{k-1}}{\partial\varphi} = \frac{e^{\lambda y_{k-2}}}{e^{\lambda y_{k-2}} + \lambda(y_k - y_{k-1})}\frac{\partial y_{k-2}}{\partial\varphi} + \frac{1}{e^{\lambda y_{k-2}} + \lambda(y_k - y_{k-1})}\frac{\partial(y_k - y_{k-1})}{\partial\varphi},$$

By the induction hypothesis, we know $\partial(y_k - y_{k-1})/\partial\varphi < 0$. Furthermore, Lemma 2 shows that $y_k - y_{k-1} \geq 0$. Equation (12) then implies $\partial y_{k-2}/\partial\varphi > \partial y_{k-1}/\partial\varphi$.

Finally, we deal with $\partial y_1/\partial\varphi$. The first two equations in (10) can be combined to obtain:

$$\gamma e^{\lambda y_2} = e^{\lambda y_1} + \gamma\lambda(y_3 - y_2).$$

We can apply the same reasoning as the above to show that $\partial y_2/\partial\varphi > \partial y_3/\partial\varphi$ implies $\partial y_1/\partial\varphi > \partial y_2/\partial\varphi$.

References

Acemoglu, Daron, and Andrew F. Newman, "The Labor Market and Corporate Structure," *European Economic Review* 46 (2002), 1733–1756.

Aoki, Masahiko, "Information, Incentives and Bargaining in the Japanese Economy: A Microtheory of the Japanese Economy," Cambridge University Press (1989).

Aoki, Masahiko, "Toward an Economic Model of the Japanese Firm," *Journal of Economic Literature* 28 (1990), 1–27.

Beckmann, Martin J., "Management Production Function and the Theory of the Firm," *Journal of Economic Theory* 14 (1977), 1–18.

Beggs, Alan W., "Queues and hierarchies," *Review of Economic Studies* 68 (2001), 297–322.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen, "The Distinct Effects of Information Technology and Communication Technology on Firm Organization," *Management Science* 60 (2014), 2859–2885.

Bloom, Nicholas, Raffaella Sadun, and J. Van Reenen, "Does Product Market Competition Lead Firms to Decentralize?" *American Economic Review* 100 (2010), 434–438.

Bolton, Patrick, and Mathias Dewatripont, "The Firm as a Communication Network," *Quarterly Journal of Economics* 109 (1994), 809–839.

Caliendo, Lorenzo, and Esteban Rossi-Hansberg, "The Impact of Trade on Organization and Productivity," *Quarterly Journal of Economics* 127 (2012), 1393–1467.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg, "The Anatomy of French Production Hierarchies," *Journal of Political Economy* 123 (2015), 809–852.

Caliendo, Lorenzo, Giordano Mion, Luca David Opromolla, and Esteban Rossi-Hansberg, "Productivity and Organization in Portuguese Firms," National Bureau of Economic Research, working paper no. 21811 (2015).

Calvo, Guillermo, and Stanislaw Wellisz, "Supervision, Loss of Control, and the Optimum Size of the Firm," *Journal of Political Economy* 86 (1978), 943–952.

Calvo, Guillermo, and Stanislaw Wellisz, "Hierarchy, Ability, and Income Distribution," *Journal of Political Economy* 87 (1979), 991–1010.

Chen, Cheng, "Management Quality and Firm Organization in Industry Equilibirum," *American Economic Journal: Microeconomics*, Forthcoming.

Friedrich, Benjamin, "Trade Shocks, Firm Hierarchies and Wage Inequality," Yale University, working paper (2015).

Galbraith, Jay R., "Organization Design," Reading, Mass.: Addison-Wesley (1977).

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy* 108 (2000), 874–904.

Garicano, Luis, and Esteban Rossi-Hansberg, "Inequality and the Organization of Knowledge," *American Economic Review* 94 (2004), 197–202.

Garicano, Luis, and Esteban Rossi-Hansberg, "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics* 121 (2006), 1383–1435.

Garicano, Luis, and Esteban Rossi-Hansberg, "Organizing Growth," *Journal of Economic Theory* 147 (2012), 623-656.

Garicano, Luis, and Timothy Van Zandt, "Hierarchies and the Division of Labor," in *The Handbook of Organizational Economics*, edited by Robert Gibbons and John Roberts, Princeton: Princeton University Press (2013).

Keren, Michael, and David Levhari, "The Optimal Span of Control in a Pure Hierarchy," *Management Science* 25 (1979), 1162-1172.

Meagher, Kieron J., "Generalizing Incentives and Loss of Control in an Optimal Hierarchy: The Role of Information Technology," *Economics Letters* 78 (2003), 273–280.

Milgrom, Paul, and Chris Shannon, "Monotone Comparative Statics," *Econometrica* 62 (1994), 157–180.

Mookherjee Dilip, "Incentives in Hierarchies," in *The Handbook of Organizational Economics*, edited by Robert Gibbons and John Roberts, Princeton: Princeton University Press (2013).

Patacconi, Andrea, "Coordination and Delay in Hierarchies," *RAND Journal of Economics* 40 (2009), 190–208.

Qian, Yingyi, "Incentives and Loss of Control in an Optimal Hierarchy," *Review of Economic Studies* 61 (1994), 527–544.

Radner, Roy, "Hierarchy: The Economics of Managing," *Journal of Economic Literature* 30 (1992), 1382–1415.

Radner, Roy, "The Organization of Decentralized Information Processing," *Econometrica* 61 (1993), 1109–1146.

Radner, Roy, and Timothy Van Zandt, "Information Processing in Firms and Returns to Scale," *Annales d'Economie et de Statistique* 25 (1992), 265–298.

Rajan, Raghuram G., and Julie M. Wulf, "The Flattening Firm: Evidence on the Changing Nature of Firm Hierarchies from Panel Data," *Review of Economics and Statistics* 88 (2006), 759–773.

Rosen, Sherwin, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics* 13 (1982), 311–323.

Tag, Joacim, "Production Hierarchies in Sweden," *Economics Letters* 121 (2013), 210–213.

Williamson, Oliver, "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy* 75 (1967), 123–138.

Whitney, Daniel, "Manufacturing by Design," *Harvard Business Review* 66 (1988), 83–91.

Wu, Yanhui, "Organizational Structure and Product Choice in Knowledge-Intensive Firms," *Management Science* 61 (2015), 1830–1848.