Improving the Organization of Knowledge in Production by Screening Problems*

Guilherme Carmona† Krittanai Laohakunakorn‡

University of Surrey University of Surrey

July 13, 2023

Abstract

We extend Garicano's (2000) model of optimal organizations by allowing its members to screen problems, i.e. to attempt the identification of problems before trying to solve them. As for solving problems, screening is costly to learn and time consuming but has the advantage of allowing for successfully screened problems to be directed to those in the organization who can solve them. We establish several properties of optimal organizations and use them to show: (a) When screening problems is as costly as solving them, optimal organizations are hierarchies as in Garicano (2000), but (b) when the cost of learning how to screen problems is sufficiently small, optimal organizations are such that workers screen all problems that they and the managers who solve the most extraordinary problems cannot solve, those problems that they screen are directed to those managers who can solve them and those problems that they neither solve nor screen are passed to the managers who solve the most extraordinary problems. For intermediate values of the cost of learning how to screen problems, we show computationally that the optimal organization is a hybrid of the above two organizational forms.

*We wish to thank Paulo Bastos, Esteban Rossi-Hansberg, John Van Reenen, three anonymous referees and seminar participants at the 2021 RES annual conference (Belfast) and EWET 2021 (Akko) for helpful comments. Any remaining errors are, of course, ours.

†Address: University of Surrey, School of Economics, Guildford, GU2 7XH, UK; email: g.carmona@surrey.ac.uk.

‡Address: University of Surrey, School of Economics, Guildford, GU2 7XH, UK; email: k.laohakunakorn@surrey.ac.uk.

1 Introduction

Firms in certain sectors, such as the automobile industry, have a hierarchical structure: Problems that workers cannot solve are taken to managers, who in turn take those that they cannot solve to higher level managers, with the top-level managers solving only rather exceptional problems, i.e. managing by exception. In addition, the organization is typically pyramidal having more workers than lower-level managers, and so on until the top-level managers.

Because solving problems is costly, the organization of a firm might be improved by having some of its members find out who in the organization is better suited to solve a given problem before actually attempting to solve it. For instance, patients normally see a general practice doctor before being referred to the appropriate specialist; secretaries in law firms are often able to direct a client to the relevant partner by identifying e.g. whether the case falls into the realms of family or commercial law; customers of many firms first interact with an automatized system before having their questions answered by a customer service representative. As skills become more specialized (due to economies of scale as discussed in Rosen (2002)), it might make sense to learn how to identify problems to find out who can solve them.

Hierarchical organizations have been rationalized by Garicano (2000). In this paper, we propose a theory of the organization of specialization that extends his framework by allowing members of the organization to screen problems that they cannot solve and rationalizes, in a single framework, both hierarchical organizations and the above features of real-world organizations.

We show that optimal organizations are hierarchical in exactly the same sense as in Garicano (2000) when screening is as time consuming as solving problems and learning how to screen is as costly as learning how to solve problems. Indeed, in this case, screening problems is pointless since solving them costs the same but results in the problems being solved.

In contrast, in an optimal organization when learning how to screen problems is quite cheap (but even when it takes as much time to screen a problem as to solve it), workers screen all the problems that they and the managers who solve the most extraordinary problems cannot solve. The problems screened by workers are then directed to those in the organization who can solve them, and those that workers can neither solve nor identify are passed to the managers who solve the most extraordinary problems. There is no flow of information in the organization besides these ones. The organization, however, is still pyramidal in the sense that there are more workers than managers; also, with the possible exception of the managers who solve the most extraordinary problems, the more infrequent the problems that a group of managers solves, the smaller the group is. When this organizational form is optimal, the increase in output obtained by moving from the best hierarchy to the optimal organization can be substantial: We provide an example where this increase is slightly above 83%. Thus, screening problems can significantly improve the organization of knowledge in production.

The introduction of screening can also help rationalize the internal organization of real-world organizations. For instance, the optimal organization in the low screening cost case fits well with how the editorial process of Economics journals works.[1] Indeed, the lead editor looks at each paper submitted and passes it directly to the editor who is best suited to handle it.

As another example, take the data set of U.S. law offices considered in Garicano and Hubbard (2007). According to their Table 2, there were 72.7% of law offices with no associates and, in such law offices, 44.8% of partners are specialists (on average). Such law offices may be better rationalized as an optimal organization with a low screening cost rather than as a hierarchy. Indeed, since there are no associates in those firms, there is no obvious hierarchy; instead, we can think of the 55.2% of generalist partners as the first layer of workers who first look at problems brought in by their clients and pass those they cannot solve to the right one of (possibly) several layers of specialist partners specializing in different legal fields; and, as in the theoretical result, there is a higher fraction of generalist than specialized lawyers.

The optimal organizational form transitions from a hierarchy to the optimal organization with a low screening cost as the cost of learning how to screen problems decreases. When this cost is intermediate, the optimal organization is a hybrid of the two that are optimal in the extreme cases. In this hybrid organization, the lower management layer of the hierarchy is disaggregated by having the workers screen the most common problems that they cannot solve and send them to a newly formed layer that solves only those problems; in contrast, the remaining problems are passed through the organization in a

[1] We are grateful to Esteban Rossi-Hansberg for suggesting this example to us.

hierarchical way as before. This may correspond to e.g. the separation of the IT department from a generic lower management layer in an online retailer whose workers face, on top of customers' orders they can deal with, IT problems that they can't. It can also correspond to more specialized layers in organizations located in densely populated urban areas compared to rural areas since the pool of workers in urban areas may have more general knowledge and thus be cheaper to train to identify problems.[2] The use of screening in the optimal organization for low and intermediate values of the cost of learning how to screen problems can also rationalize our motivating examples, namely that general practice doctors, secretaries in law firms and automatized systems screen and direct problems to those in their organizations who can solve them.

Our results provide more broadly a theory of the organization of specialization that allows both horizontal and vertical specialization. A key determinant of how horizontal such organization should be is the cost of learning how to screen problems, which we regard as a measure of how hard it is to match problems with solutions. Horizontal specialization requires that problems are sent to the right people; this requires screening, and specialization in this dimension is limited by screening costs. Vertical specialization requires that problems are passed through multiple layers; specialization in this dimension is limited by communication costs. Our results show that, in general, the optimal organization will feature both horizontal and vertical specialization, with the organization of specialization being vertical when it is hard to match problems with solutions (i.e. when the cost of screening is high as in Garicano (2000)) and horizontal when such matching is easy. Our results thus validate the view in Becker and Murphy (1992) and Garicano (2000) that the matching of problems with solutions is a crucial limit to specialization.

Our results characterize the optimal degree of specialization in several dimensions that includes, but it is not restricted to, the degree of its horizontality. We operationalize this theory by explicitly modelling the problem of a firm organizing the knowledge and labor of its members. In particular, our results provide conclusions on the optimal way to organize the relationship between workers and managers in an organization, or more broadly, between those who need to solve problems and those who specialize in solving

[2]Tian (2021) documents the stylized fact that there is greater division of labor within firms in larger cities and suggests that larger cities provide workers with more opportunities to acquire skills that reduce the cost of training them. This is consistent with potential workers in larger cities having lower costs of learning how to screen problems than those in rural areas.

them (e.g. the general public and medical doctors). In our framework, there is an optimal degree of specialization in (i) tasks (i.e. on how people use their time, either to work or to solve problems or both), (ii) knowledge (i.e. on what people know how to solve), (iii) screening (i.e. on what people know how to screen and, thus, whether or not they can directly find who can solve the problems they face) and (iv) scope (i.e. on what the organization as a whole knows), which solves the trade-offs between specialization and communication costs already present in Garicano (2000) and the horizontality of (i.e. more direct) communication and screening costs.

We present our model in Section 3 after a brief literature review in Section 2. To obtain the above conclusions, we first establish in Section 4 general properties of optimal organizations by extending analogous results in Garicano (2000) and by obtaining new ones. This is not straightforward because of the extra generality our framework and because, while Garicano (2000) gives great intuition for his results, some details are missing.[3] Moreover, to the extent that some properties of hierarchies do not carry over to the general setting, we can distinguish between features of optimal organizations that are intrinsic to specialization and those that require the assumption of hierarchy.[4]

These results are applied in Section 5 to address the question of whether specialization should be vertical or horizontal when the cost of learning how to screen problems is high (Section 5.1), low (Section 5.2) and intermediate (Section 5.3).

In general, the optimal organization must resolve trade-offs between the cost of learning how to screen problems, the cost of learning how to solve problems and the time spent attempting to solve or screen the problem and then communicating the answer (which we refer to broadly as "communication costs"). These trade-offs are simplified when communication costs are small, a case we consider in Section 5.4 to obtain sharper characterizations of the optimal organization, namely, that all problems it faces are eventually solved but none by its workers. Such optimal organization is thus highly specialized and asymmetric as it has very knowledgeable managers alongside unskilled workers (or perhaps using automation to perform the workers' tasks).

[3]For example, Proposition 1 in Garicano (2000) states that in any optimal organization, only one layer specializes in production. This is true if for any two organizations with the same output, the one with fewer layers is preferred, which is the case in this paper but not explicitly stated in Garicano (2000).

[4]For example, the property that higher layers solve less frequent problems holds in a hierarchy but may fail when some problems are screened.

Section 5.5 contains some comments on the implications of our results for empirical work. Of course, broadening the applicability of the theory of optimal organizations can only improve its empirical fit. Hierarchical organizations, which are a particular optimal outcome of the framework of this paper, have been useful to analyze empirically manufacturing firms as shown by Caliendo, Monte, and Rossi-Hansberg (2015) and Caliendo, Mion, Opromolla, and Rossi-Hansberg (2020). However, about 18% (resp. 25%) of the firms in the data set considered in the former (resp. latter) paper are not hierarchies. Hence, the richer set of optimal organizational forms that our unified framework allows for can be helpful to empirically analyze a broader set of firms. This may be important because our model predicts that each organization's response to changes in its environment depends on its organizational form and, moreover, the response may involve changing the organizational form itself.[5]

The results described so far pertain to the case of non-cumulative knowledge. We extend our general characterization results to the cumulative knowledge case in Section 6 and show that, when the cost of learning how to screen problems is low, the optimal organization with cumulative knowledge is similar to the one with non-cumulative knowledge with one important difference: the problems that workers neither solve nor screen are now the most common ones that they do not solve. This difference may matter to distinguish between breath and depth of knowledge. To illustrate, consider a resident physician in the emergency room of an hospital.[6] In general, there will be some patients whom the resident will be able to treat and others whom she will be able to direct to the right specialist, e.g. a cardiologist. In addition, there will be some cases regarding which the resident will not know what to do and which she will pass to the attending physician. This latter feature is better described by the result in the cumulative knowledge case. Indeed, in the non-cumulative case, the problems the resident has no knowledge about are the least common that may arise, whereas the attending physician is, in general, not specialized in rare diseases but rather someone who has more advanced training than the

[5]For example, we consider the response to a 10% drop in the time needed to solve problems and show that hierarchies typically respond by increasing the knowledge of all managers, whereas optimal organizations for low screening costs will sometimes increase the knowledge of the top managers at the expense of the knowledge of lower managers. Moreover, an organization will sometimes respond by changing its organizational form, becoming more vertical as passing problems through the organization becomes cheaper relative to screening them.

[6]We are grateful to an anonymous referee for suggesting this example to us.

resident. This is consistent with cumulative knowledge, because then what may distinguish the residents from the attending physicians who receive the problems that residents neither solve nor screen is that the latter may have deeper knowledge than the former about e.g. an advanced procedure to treat relatively common diseases.

Section 7 contains some concluding remarks and the proofs of our results are in the (online) Appendix. Some omitted details are in the supplementary material to this paper.[7]

2 Literature review

There is a large literature on organizational theory; see Gibbons and Roberts (2013) for a survey. Here we focus on the literature that emerged from Garicano (2000).

The framework of Garicano (2000) and the knowledge-based hierarchies that it rationalized have been used to address many economic questions. These questions include the evolution of wage inequality (Garicano and Rossi-Hansberg (2004) and Garicano and Rossi-Hansberg (2006)), the existence and growth of firms (Fuchs, Garicano, and Rayo (2015) and Caliendo and Rossi-Hansberg (2012)), economic development (Garicano and Rossi-Hansberg (2012)) and offshoring (Antràs, Garicano, and Rossi-Hansberg (2006)). See Garicano and Van Zandt (2013) and Garicano and Rossi-Hansberg (2015) for surveys of this literature and Caicedo, Lucas, and Rossi-Hansberg (2019), Eeckhout and Kircher (2018), Geerolf (2017) and Kikuchi, Nishimura, and Stachurski (2018) for some recent developments.

Our focus is narrower than that of the above papers. Specifically, we generalize the setting of Garicano (2000) by allowing screening of problems but the goal remains the same, namely, to characterize the optimal organization. Other papers are similar in scope or goal: Chen and Suen (2019) introduce a delay cost that is incurred every time help is provided and Gumpert (2018) introduce heterogeneity in the communication costs; some extensions have been made in work whose focus is mainly empirical such as in Bloom, Garicano, Sadun, and Van Reenen (2014), who introduce heterogeneity in decision types. None of these papers allow members of the organization to screen problems.

The ability of members of the organization to screen problems can be interpreted in light of Crémer, Garicano, and Prat (2007) as the ability to communicate in a specific

[7]The supplementary material is available at https://klaohakunakorn.com/orgsm.pdf.

code. There, a code is a finite partition of the set of problems; here, in contrast, the partition is specific to each individual through the set B of problems that she can identify and consists of B^c and $\{\omega\}$ for each $\omega \in B$. That is, if someone draws a problem $\omega \in B$, she knows that she is facing problem ω but, in contrast, if $\omega \in B^c$, she does not know that she is facing problem ω, she only knows that she is facing some problem in B^c.

3 The model

3.1 Production

Production happens when an individual solves a problem with which he is faced. The set of all possible problems is an interval $\Omega \subseteq \mathbb{R}_+$ and problems are drawn according to a probability measure F on Ω. We assume that F has a continuous and strictly decreasing density f and that $0 \in \Omega$.

Individuals are characterized by a knowledge set $A \subseteq \Omega$, which consists of the problems that the individual can solve. The complement A^c of A is then the set of problems that the individual cannot solve. We extend Garicano (2000) by allowing individuals to screen problems: There is a screening set $B \subseteq \Omega$, with $A \subseteq B$, which consists of the problems that the individual can identify. The condition $A \subseteq B$ then means that a person knows how to identify all the problems that he can solve so that the elements of $B \setminus A$ are the problems that the individual can identify but not solve. If an individual draws a problem $\omega \in B \setminus A$, although no production directly results, he knows that he is facing problem ω; in contrast, if $\omega \in B^c$, the individual does not know that he is facing problem ω, he only knows that he is facing some problem in B^c — a person cannot distinguish problems ω and ω' if they both belong to B^c.

Learning to solve problems is at least as costly as learning to screen them and both are proportional to the size of the corresponding set. Thus, letting μ denote the Lebesgue measure, we let $c\mu(A)$ be the cost of learning the knowledge set A and $\xi\mu(B \setminus A)$ be the cost of learning the remaining part of the screening set B, and assume that $0 < \xi \leq c$.

3.2 Organizations

An organization consists of a set of individuals. The organization allows each of its members to ask other members for the solution of the problem he has drawn. Such communication incurs a cost: The member who is being asked incurs a cost of h units of time to solve the problem if it is in his knowledge set. If, instead, the problem is not in his knowledge set, then he incurs a cost of π units of time to attempt the identification of the problem, regardless of whether or not he succeeds. Moreover, these costs are incurred regardless of whether or not some other member has already identified it.[8] We assume that $0 < \pi \leq h$.

The members of an organization are organized in several layers, the set of those being denoted by L. Each layer $i \in L$ has knowledge and screening sets A_i and B_i respectively. In addition, it has a list l_i of the layers with whom those in layer i may communicate. The elements of l_i are ordered to indicate which one of each pair of elements of the list is contacted first by layer i. Thus, layer i is also characterized by a precedence relation, which is, formally, a linear order $\prec_i$ on l_i with the property that i is its minimal element; we may, therefore, write $l_i = \{j_0, j_1, \ldots, j_{m_i}\}$ where $i = j_0 \prec_i j_1 \prec_i \cdots \prec_i j_{m_i}$.

Members in layer i will solve the problems in A_i and pass those in $B_i \setminus A_i$ to the first layer (according to $\prec_i$) that can solve them. Problems in B_i^c are passed down the list, to layer $j_1, j_2, \ldots$, until they are either solved or abandoned. Thus, the set of problems drawn by layer i that are solved is $\cup_{l \in l_i} A_l$. In addition, out of the problems drawn by layer i, layer k solves those problems in $A_k \setminus \cup_{l \prec_i k} A_l$ and does not solve but attempts to identify problems in $A_k^c \cap (\cup_{l \prec_i k} B_l)^c$.

The final elements describing a layer $i \in L$ of an organization are its relative size $\beta_i \in (0, 1)$ and the allocation of time spent between producing (denoted by t_i^p) and helping other layers (denoted by t_i^h). The allocation of time satisfies the standard requirement that $t_i^p \geq 0$, $t_i^h \geq 0$ and $t_i^p + t_i^h \leq 1$; in addition, the total time spent helping $\beta_i t_i^h$ equals the time needed by the other layers:

$$\beta_i t_i^h = \sum_{k \in L \setminus \{i\} : i \in l_k} \beta_k t_k^p \left[hF\left(A_i \setminus \cup_{l \prec_k i} A_l\right) + \pi F\left(A_i^c \setminus \cup_{l \prec_k i} B_l\right) \right].$$

For notational convenience, let, for each $i, k \in L$,

$$\alpha_{ik} = \begin{cases} hF\left(A_i \setminus \cup_{l \prec_k i} A_l\right) + \pi F\left(A_i^c \setminus \cup_{l \prec_k i} B_l\right) & \text{if } i \neq k \text{ and } i \in l_k, \\ 0 & \text{otherwise.} \end{cases}$$

[8] For example, if someone faces a problem that he cannot identify, passes it to someone else who can identify but not solve it, who in turn passes it to a third person who can solve it, then the second (resp. third) person incurs a cost of π (resp. h) units of time.

A final requirement in the definition of an organization concerns the sets A_i and B_i that a layer $i \in L$ may have as knowledge and screening sets. We require each such set to be a finite union of disjoint intervals of the form $[a, b)$ with $a \leq b \leq \infty$. The class of such sets, which we denote by $\mathcal{I}$, is contained in the class of measurable sets and is closed under unions, intersections and set differences, making it convenient for technical reasons.[9] In addition, elements of $\mathcal{I}$ are intuitively simple, i.e. requiring layers to choose knowledge and screening sets in $\mathcal{I}$ amounts to disallowing them to learn overly complicated sets. Thus, this places some limits on the complexity of the organization.

In summary, an *organization* is $O = (L, (\beta_i, A_i, B_i, l_i, \prec_i, t_i^p, t_i^h)_{i\in L})$ such that $L \subset \mathbb{N}$ is a finite set, $\sum_{j\in L} \beta_j = 1$, and for each $i \in L$: $\beta_i > 0$, A_i and B_i belong to $\mathcal{I}$ and satisfy $A_i \subseteq B_i \subseteq \Omega$, $l_i \subseteq \{1, \ldots, L\}$ with $i \in l_i$, $\prec_i$ is a linear order on l_i such that i is its minimal element, $t_i^p \geq 0$, $t_i^h \geq 0$, $t_i^p + t_i^h \leq 1$ and $\beta_i t_i^h = \sum_{k\in L} \beta_k t_k^p \alpha_{ik}$.

The expected output of the organization O is

$$y = \sum_{i\in L} \beta_i (t_i^p F(\cup_{l\in l_i} A_l) - c\mu(A_i) - \xi\mu(B_i \setminus A_i)). \tag{1}$$

This is obtained as follows. The production of each of its members equals the time he spends in production provided that he can solve the problem he draws. Thus, the expected production of an individual in layer $i \in L$ is $t_i^p F(\cup_{l\in l_i} A_l)$. The expected output of layer i is then obtained by subtracting from its expected production the cost of learning i's knowledge set A_i and the cost of leaning how to identify but not solve the problems in $B_i \setminus A_i$. The expected output of O is then the sum of the expected output of each of the layers, weighted by their relative sizes.

3.3 Relationship with Garicano (2000)

The setting of Garicano (2000) is a particular case of ours, obtained by imposing $B_i = A_i$ for each $i \in L$ and $h = \pi$. Indeed, in this case we obtain, as in Garicano (2000), that

$$\alpha_{ik} = h\left(F\left(A_i \setminus \cup_{l\prec_k i} A_l\right) + F\left(A_i^c \setminus \cup_{l\prec_k i} A_l\right)\right) = h(1 - F\left(\cup_{l\prec_k i} A_l\right))$$

if $i \neq k$ and $i \in l_k$, and $y = \sum_{i\in L} \beta_i (t_i^p F(\cup_{l\in l_i} A_l) - c\mu(A_i))$.

To understand what the organization gains in our more flexible setting, consider first that $\pi < h$ but still $B_i = A_i$ for each $i \in L$. In this case, the amount of time that

[9] Another convenient property is that if $A \in \mathcal{I}$ is such that $F(A) = 0$, then $A = \emptyset$ since otherwise, for some $a < b$, $F(A) \geq \int_{[a,b)} f > 0$, the last inequality following since f is strictly decreasing.

layer i spends helping layer k is smaller than $h(1 - F(\cup_{l \prec_k i} A_l))$ simply because workers in layer i only attempt to identify, but not solve, problems outside their knowledge set; hence, the cost of time per problem is only π, not h. In addition, having $B_k \neq A_k$ can also reduce helping costs: Indeed, problems in $B_k \setminus A_k$ go directly to the layer that can solve it instead of being asked to each layer that precedes it in k's precedence relation. Furthermore, problems in $B_k \cap (\cup_{i \in l_k} A_i)^c$ are now known not to be solvable, hence no time is devoted to identify, let alone trying to solve, them.

3.4 Optimal organizations

We now turn to the definition of an optimal organization. Part of the definition will reflect a preference for less complex organizations.

If O is an organization and $y = 0$, then it is straightforward to characterize O: One layer is enough and we can set $A_1 = B_1 = \emptyset$. Thus, the focus is on the case where $y > 0$, which will be assumed throughout, i.e. it will be part of the definition of an optimal organization.

One requirement in the definition of an organization is that $\beta_i > 0$ for all i, which means that there are no unmanned layers. This is reasonable because unmanned layers can be removed to obtain an organization with fewer layers and the same output. This seems to be a natural way of avoiding multiplicity: Thus, given two organizations with the same output, the one with the smaller number of layers is preferred. Extending this logic to lists, we consider that, given two organizations with the same output, if one is obtained from the other by shortening some lists, then the former is preferred to the latter. In both cases, the former organization is intuitively less complex than the latter, hence, these two criteria express a preference for simplicity.[10]

The mildest way of introducing this preference for simplicity is in a lexicographic way. An organization O is *lexicographically optimal* if $y > 0$ and there is no other organization

[10]To see how such preference for simplicity avoids uninteresting multiplicity of optimal organizations, suppose that O is an optimal organization. Its output y can also be obtained by replicating O to obtain an organization $\hat{O}$ with $\hat{L} = 2L$ and, for each $i = 1, \ldots, L$, $\hat{B}_i = \hat{B}_{i+L} = B_i$, $\hat{A}_i = \hat{A}_{i+L} = A_i$, $\hat{l}_i = l_i$, $\hat{l}_{i+L} = \{j + L : j \in l_i\}$, $\hat{\prec}_i = \prec_i$, $(j+L) \hat{\prec}_{i+L} (k+L)$ if and only if $j \prec_i k$, $\hat{t}_i^h = \hat{t}_{i+L}^h = t_i^h$, $\hat{t}_i^p = \hat{t}_{i+L}^p = t_i^p$ and $\hat{\beta}_i = \hat{\beta}_{i+L} = \beta_i/2$; hence, in the absence of a preference for simplicity, $\hat{O}$ would also be optimal. In particular, note that there are more than one layer of workers in $\hat{O}$, i.e. our results do not extend to the case where such preference for simplicity is dispensed with.

$\hat{O}$ such that (i) $\hat{y} > y$, or (ii) $\hat{y} = y$ and $\hat{L} \subset L$ or (iii) $\hat{y} = y$, $L = \hat{L}$, $\hat{l}_i \subseteq l_i$ for all $i \in L$ and $\hat{l}_j \neq l_j$ for some $j \in L$.

Lexicographically optimal organizations may fail to exist. For this reason, we also consider the following stronger preferences for simplicity. Let $\eta > 0$ and consider the case where adding a layer to the organization costs η units of output; we think of η as reflecting administrative costs which are increasing in the number of layers of an organization. In this case, define $Y = y - (L-1)\eta$. An organization O is *η-optimal* if $Y > 0$ and there is no other organization $\hat{O}$ such that (i) $\hat{Y} > Y$, or (ii) $\hat{y} = y$, $\hat{L} = L$, $\hat{l}_i \subseteq l_i$ for each $i \in L$ and $\hat{l}_j \neq l_j$ for some $j \in L$.

Except for existence, our results hold regardless of whether we focus on lexicographically optimal or η-optimal organizations. Thus, we say that an organization O is *optimal* if it is lexicographically optimal or η-optimal.

4 A theory of the organization of specialization

In this section, we show that the organization will specialize in tasks, knowledge, screening and scope. Moreover, we identify the trade-offs between the cost of learning how to screen problems, the cost of learning how to solve problems, and the time required to attempt to screen or solve problems and communicate the answer that the optimal degree of specialization resolves.

To get to the key maximization problem that determines how these trade-offs are resolved, we first show that optimal organizations satisfy the following general properties: (i) specialization in the allocation of labor and time; (ii) the union of the screening sets is an interval starting at zero and can be partitioned into the "pure knowledge sets" of problems that a given layer can solve but no one else can identify, "screened knowledge sets" of problems that one layer sends to another and "known unknowns" that a given layer screens and discards; and (iii) the frequency of each class of problems is increasing in its marginal cost to the organization. As well as being of substantive interest, these general characterization results provide enough structure to imply a numerical method for computing optimal organizations. Moreover, in the next section, we will combine these general results with parametric assumptions which will be sufficient to characterize the exact organizational form in several situations of interest.

Theorem 1 characterizes the allocation of personnel to each layer and the allocation of time for each person in each layer, as well as the list and the precedence relation of each layer. It shows that there is a layer that specializes in production whereas all the other layers specialize in helping the former layer. In light of this, the list of the former layer is L and the list of each of the latter layers has no layer other than itself. In addition, Theorem 1 describes the output of an optimal organization.

Theorem 1 *If O is an optimal organization, then there is $i \in L$ such that $t_i^p = 1$, $t_i^h = 0$, $l_i = L$, $\beta_i = \frac{1}{1+\sum_{l=1}^L \alpha_{li}}$ and $y = \frac{F(\cup_{l\in L} A_l) - (\nu_i + \sum_{l=1}^L \nu_l \alpha_{li})}{1+\sum_{l=1}^L \alpha_{li}}$, where $\nu_l = c\mu(A_l) + \xi\mu(B_l \setminus A_l)$ for each $l \in L$. Furthermore, for each $j \neq i$, $t_j^p = 0$, $t_j^h = 1$, $\alpha_{ji} > 0$, $l_j = \{j\}$ and $\beta_j = \frac{\alpha_{ji}}{1+\sum_{l=1}^L \alpha_{li}}$.*

We may assume, relabeling the layers if necessary, that layer 1 is such that $t_1^p = 1$ and that $2 \prec_1 \cdots \prec_1 L$. In this case, the precedence relation $\prec_1$ is just the standard "less than" $<$ order. Since $\alpha_{ik} = 0$ if $k \neq 1$, we simplify the notation and write, for each $i \in L \setminus \{1\}$, $\alpha_i = \alpha_{i1}$ and $\alpha_1 = 1$, the latter for notational convenience. Also for convenience, we write $\theta = \theta_1 = F(\cup_{l\in L} A_l) - \sum_{l\in L} \alpha_l \nu_l$ and $\gamma = \gamma_1 = \sum_{l\in L} \alpha_l$. Note that under this notation, $y = \frac{\theta}{\gamma}$ and $\beta_l = \frac{\alpha_l}{\gamma}$ for each $l \in L$. We refer to those in layer 1 as workers and those in layers $i \neq 1$ as managers.

Theorem 1 generalizes Proposition 1 in Garicano (2000).[11] The following results, which characterize the allocation of knowledge in an optimal organization by establishing the properties that the knowledge and screening sets $(A_1, B_1, \ldots, A_L, B_L)$ must satisfy, are also related to analogous ones in Garicano (2000) but with significant differences. In the latter setting, where $B_l = A_l$ for each $l \in L$, it turns out that the collection $\{B_1, \ldots, B_L\}$ is pairwise disjoint in any optimal organization and thus partitions the union $\cup_{l\in L} B_l$ of the screening sets. Furthermore, Garicano's (2000) key result is that $B_1 < \ldots < B_L$, i.e. these sets can be ordered according to the frequency of the problems they contain. On the other hand, in the general case, the collection $\{B_1, \ldots, B_L\}$ may not be pairwise disjoint and each B_l may not be an interval; hence $\{B_1, \ldots, B_L\}$ cannot be ordered. However, we will show that by considering the appropriate partition of $\cup_{l\in L} B_l$, the elements of such partition can be ordered and we provide a criterion with which to order them.

[11] Theorem 1 also corrects Proposition 1 in Garicano (2000). One issue with the latter has been already illustrated in Footnote 10, namely that the conclusion of Theorem 1 does not hold without a preference for simplicity. See Section A.1 in the Appendix for more details.

Intuitively, an organization may wish to treat a problem differently depending on (i) whether it is known or just screened, (ii) the layer which knows how to solve the problem and (iii) the layer which knows how to screen the problem. Thus, the appropriate partition of $\cup_{l \in L} B_l$ should contain, for each $l \in L$, its "pure knowledge set" containing problems that layer l can solve but no one else can screen, its "screened (by j) knowledge set" containing problems that layer l can solve and layer j can screen (for each $j \in L$), and the "known unknowns" which layer l can screen but no one can solve. In fact, Lemma 1 will imply that, in an optimal organization, these sets are, respectively, $A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c)$, $A_l \cap (B_j \setminus A_j)$ for each $j < l-1$ and $(B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c)$.

Specifically, Lemma 1 shows that (i) there is no duplication of effort: if layer k knows how to solve (resp. screen) a problem, there is no point in bearing the cost of having layer $l > k$ learn how to solve or screen (resp. screen but not solve) it too, (ii) the last layer does not learn how to screen problems that it cannot solve and (iii) no layer learns how to screen problems that the next layer can solve (since such problems will be passed anyway).

Lemma 1 *If O is an optimal organization, then*

1. $B_l \cap A_k = \emptyset$ *and* $(B_l \setminus A_l) \cap B_k = \emptyset$ *for each* $k, l \in L$ *such that* $k < l$,[12]

2. $B_L \setminus A_L = \emptyset$ *and*

3. $B_i \cap A_{i+1} = \emptyset$ *for each* $1 \leq i < L$.

Thus, the partition of the union of the screening sets of an optimal organization we consider is

$$\begin{aligned} \mathcal{C} &= \{A_l \cap (B_j \setminus A_j) : l, j \in L \text{ and } j < l-1\} \cup \{A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c) : l \in L\} \\ &\cup \{(B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c) : 1 \leq l < L\} \end{aligned}$$

with the usual convention that the intersection of an empty family of subsets of Ω is Ω itself.[13] Lemma 1 implies that $\mathcal{C}$ is indeed a partition of $\cup_{l \in L} B_l$.

[12] In particular, $A_l \cap A_k = \emptyset$ and $(B_l \setminus A_l) \cap (B_k \setminus A_k) = \emptyset$ for each $k, l \in L$ such that $k \neq l$.

[13] The sets in $\mathcal{C}$ can be used to describe an organization in place of the knowledge and screening sets $\{A_l, B_l\}_{l \in L}$, since the latter can be obtained from the former. Indeed, for each $l \in L$, $A_l = \big(\cup_{j<l-1} (A_l \cap (B_j \setminus A_j))\big) \cup \big(A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c)\big)$ and $B_L = A_L$, and, for each $1 \leq l < L$, $B_l \setminus A_l = \big(\cup_{j>l+1} (A_j \cap (B_l \setminus A_l))\big) \cup \big((B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c)\big)$ and $B_l = A_l \cup (B_l \setminus A_l)$.

The results below will imply that we may assume that the elements of $\mathcal{C}$ are intervals for the purposes of obtaining optimal organization. In addition, there are no gaps between the knowledge and screening sets of different layers (or equivalently, between the elements of $\mathcal{C}$) and the union of these sets contain zero. The formal statement of the "no gaps" result is as follows. Let $a_i = \min B_i$ for each $i \in L$ with the standard convention that $\min \emptyset = \infty$. Furthermore, let $b_i = \max B_i$ for each $i \in L$ with the standard convention that $\max \emptyset = -\infty$.

Theorem 2 *If O is an optimal organization, then $\cup_{i=1}^{L} B_i = [\min_{1 \le i \le L} a_i, \max_{1 \le i \le L} b_i)$ and $\min_{1 \le i \le L} a_i = 0$.*

Theorem 2 implies that $\cup_{l \in L} B_l = \cup_{C \in \mathcal{C}} C$ is an interval. We now argue that each $C \in \mathcal{C}$ can be taken to be an interval as well, and we identify the criterion according to which $\mathcal{C}$ is ordered. This is important because a crucial lesson from Garicano (2000) is that the frequency of the problems a manager knows how to solve is decreasing in the layer to which he belongs. As we will see, the correct generalization of this result is that the frequency of problems contained in each $C \in \mathcal{C}$ is increasing in the marginal cost to the organization of increasing the size of C. For each $C \in \mathcal{C}$, let the cost of learning C be c_C defined by

$$c_C = \begin{cases} \alpha_l c + \alpha_j \xi & \text{if } C = A_l \cap (B_j \setminus A_j) \text{ for some } l, j \in L \text{ with } j < l-1, \\ \alpha_l c & \text{if } C = A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c) \text{ for some } l \in L, \\ \alpha_l \xi & \text{if } C = (B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c) \text{ for some } l \in L. \end{cases}$$

It then follows that

$$\sum_{l \in L} \alpha_l \nu_l = \sum_{C \in \mathcal{C}} c_C \mu(C), \text{[14]} \tag{2}$$

i.e. c_C is the marginal cost to the organization of increasing the size of C.

To state the result formally, for each $C, C' \in \mathcal{C}$, we write $C < C'$ if $\omega < \omega'$ for each $\omega \in C$ and $\omega' \in C'$. Moreover, when O and $\hat{O}$ are two organizations with the same number of layers, i.e. $L = \hat{L}$, the sets in $\hat{\mathcal{C}}$ are in a one-to-one relationship with those in $\mathcal{C}$, namely, $\hat{A}_l \cap (\hat{B}_j \setminus \hat{A}_j)$ corresponds to $A_l \cap (B_j \setminus A_j)$ and so on, so that $\hat{C}$ corresponds to C for each $C \in \mathcal{C}$ and, thus, $\hat{\mathcal{C}} = \{\hat{C} : C \in \mathcal{C}\}$.

Theorem 3 *Let O be an optimal organization. Then:*

[14] See Lemma A.10 in the Appendix for a proof.

(a) For each $C, C' \in \mathcal{C}$, if $c_C > c_{C'}$, then $C < C'$.

(b) If $\mathcal{C} = \{C_1, \ldots, C_{|\mathcal{C}|}\}$ is such that $c_{C_1} \geq \cdots \geq c_{C_{|\mathcal{C}|}}$, then there exists an optimal organization $\hat{O}$ such that $\hat{L} = L$, $\hat{C}_1 < \cdots < \hat{C}_{|\mathcal{C}|}$, $\hat{C}$ is an interval and $F(\hat{C}) = F(C)$ for each $C \in \mathcal{C}$, $\hat{y} = y$ and $\hat{l}_i = l_i$ for each $i \in L$.

Theorem 3 says that $\mathcal{C}$ must be ordered according to the learning costs, namely that C is decreasing with c_C. If all the learning costs are different, then this gives us an ordering of $\mathcal{C}$. If, instead, $c_C = c_{C'}$ for some $C, C' \in \mathcal{C}$, then optimality does not require that $C < C'$ or that $C' < C$; hence, it is possible that some sets in $\mathcal{C}$ are unordered in some optimal organization. However, C and C' can be ordered without any loss to obtain an optimal organization $\hat{O}$ where all the sets in the partition $\hat{\mathcal{C}}$ of $\cup_{l \in \hat{L}} \hat{B}_l$ (which will be equal to $\cup_{l \in L} B_l$) are intervals and are, therefore, ordered. Moreover, any ordering $C_1 < \cdots < C_{|\mathcal{C}|}$ can be obtained provided that $c_{C_1} \geq \cdots \geq c_{C_{|\mathcal{C}|}}$.

A key result in Garicano (2000) is "management by exception": production workers solve the most common problems, while higher levels of management deal with more and more exceptional or infrequent problems. Theorem 3 implies this result when $B_l = A_l$ for all $l \in L$. Indeed, in this case, every problem that requires some time of a higher layer also requires the same time of a lower layer; hence $\alpha_j > \alpha_k$ and hence $c_{A_j} > c_{A_k}$ whenever $j < k$. However, we note that when screening is possible, Theorem 3 shows that it is not the hierarchical position of a manager that determines the frequency of the problems that he solves, but rather the marginal cost to the organization of those problems. Thus, even a top manager may deal with frequent problems if they are screened and sent directly to him.

The following general results on the order $\mathcal{C}$ are a corollary of Theorem 3. First, the problems that are solved by a layer l and were screened by a previous layer must be more frequent than the problems that are solved by that layer l but not screened by any other layer. Second, for a given layer, the problems that are screened to be solved by another layer must be more frequent than the problems that are screened and discarded. Third, in the case where learning to screen is cheaper than learning to solve problems, the problems that are solved by a given layer but not screened by any layer must be more frequent than the problems that are screened by that layer but not solved by any layer.

Corollary 1 *Let O be an optimal organization. Then:*

(a) *For each* $l \in L$ *and* $j < l-1$, $A_l \cap (B_j \setminus A_j) < A_l \cap (\cap_{k<l-1}(B_k \setminus A_k)^c)$.

(b) *For each* $j \in L$ *and* $l > j+1$, $A_l \cap (B_j \setminus A_j) < (B_j \setminus A_j) \cap (\cap_{k>j+1} A_k^c)$.

(c) *If* $\xi < c$, *then for each* $l \in L$, $A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c) < (B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c)$.

Theorem 3 is established by changing sets $C, C' \in \mathcal{C}$ to $\hat{C}, \hat{C}'$ so that $F(\hat{C}) = F(C)$ and $F(\hat{C}') = F(C')$. An alternative variation is obtained by changing sets in $\mathcal{C}$ so that, for each layer $l \in L$, ν_l remains unchanged. Using this variation, we obtain another result, Theorem A.1 in the Appendix, concerning the order of sets. As its corollary, we obtain that the knowledge set of the workers (layer 1) contains the most frequent problems. In addition, for $j < k$, the problems that are screened by layer j and sent to some layer l (resp. discarded) are more frequent than the problems that are screening by layer k and sent to the same layer l (resp. discarded). When $\pi = h$, we can establish further orders: for $j < k$, the problems that are solved by layer j but not screened by any layer are more frequent than the problems solved by layer k but not screened by any layer. Finally, still assuming $\pi = h$, the problems that are solved by a given layer are more frequent than the problems that are screened but not solved by that layer.

Corollary 2 *Let* O *be an optimal organization. Then:*

(a) $A_1 < C$ *for each* $C \in \mathcal{C} \setminus \{A_1\}$.

(b) *For each* $j, k \in L$ *such that* $j < k$, $A_l \cap (B_j \setminus A_j) < A_l \cap (B_k \setminus A_k)$ *for each* $l > k+1$.

(c) *For each* $j, k \in L$ *such that* $j < k$, $(B_j \setminus A_j) \cap (\cap_{l>j+1} A_l^c) < (B_k \setminus A_k) \cap (\cap_{l>k+1} A_l^c)$.

If $\pi = h$, *then in addition:*

(d) $A_k \cap (\cap_{j<k-1}(B_j \setminus A_j)^c) < A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c)$ *for each* $k, l \in L$ *with* $k < l$.

(e) $C < C'$ *for each* $l \in L$, $C \in \mathcal{C}(A_l)$ *and* $C' \in \mathcal{C}(B_l \setminus A_l)$, *where*

$$\mathcal{C}(A_l) = \{A_l \cap (B_j \setminus A_j) : j < l-1\} \cup \{A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c)\}$$
$$\mathcal{C}(B_l \setminus A_l) = \{A_j \cap (B_l \setminus A_l) : j > l+1\} \cup \{(B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c)\}.$$

Corollary 2 demonstrates that some features of "management by exception" survive in the general setup. First, there is "management by exception" in the sense that higher layers of management learn to screen rarer problems. However, this conclusion holds only when comparing problems that are sent to the same layer (or discarded because no one can solve them). Second, there is "management by exception" in the sense that higher layers of management learn to solve rarer problems. However, this conclusion holds only when we consider those problems that are not screened by anyone and requires that screening problems is as time consuming as solving them.

The previous results give already enough structure so that optimal organizations can be obtained by solving the problem of maximizing output by choosing the number L of layers, an ordering of $\mathcal{C}$ (i.e. to write $\mathcal{C} = \{C_1, \ldots, C_m\}$ with $C_1 < \cdots < C_m$ and $m = |\mathcal{C}|$), and the size $\mu(C)$ of each $C \in \mathcal{C}$ such that $\sum_{C \in \mathcal{C}} \mu(C) \leq \mu(\Omega)$. Indeed, letting $y(L, \psi, \mu)$ be the output of an organization with L layers, an ordering ψ of $\mathcal{C}$, and a vector of sizes $\mu = (\mu(C))_{C \in \mathcal{C}}$, an η-optimal organization resolves its trade-offs by solving:

$$\max_{L,\psi,\mu} y(L, \psi, \mu) - \eta(L-1) \text{ subject to } \sum_{C \in \mathcal{C}} \mu(C) \leq \mu(\Omega).$$ [15]

When the goal is to obtain an η-optimal organization, then the set of relevant numbers of layers is finite since the output of any organization is bounded above by 1. Thus, η-optimal organizations exist when $\mu(\Omega) < \infty$, i.e. when Ω is bounded.

Theorem 4 *If Ω is bounded, then an η-optimal organization exists.*

5 Vertical versus horizontal specialization

In this section we show that the cost ξ of learning how to screen problems, which we regard as a measure of the difficulty of matching of problems with solutions, is a key parameter shaping the organization of specialization. Indeed, our results show that the optimal organization features vertical specialization when ξ is high (Section 5.1), horizontal specialization when ξ is low (Section 5.2), and both vertical and horizontal specialization for intermediate values of ξ (Section 5.3).

When the costs h and π are small, the coordination costs that Becker and Murphy (1992) identified as limits to specialization are just the costs of learning how to solve and screen problems and the cost of adding layers. In this case, which we consider in Section 5.4, we provide a more detailed characterization of optimal organizations and

[15] A lexicographically optimal organization picks the smallest L that solves this problem with $\eta = 0$.

show that the extent of the organization's knowledge is limited "by the extent of the market" (i.e. the organization knows everything there is to know) and that the workers have no knowledge, i.e. workers and managers are fully specialized. As communication costs become very small, superstar managers emerge with unlimited span of control, independently of whether the organization is horizontal or vertical. Moreover, when there is a cost η of adding layers, the optimal organization has only two layers and vertical and horizontal specialization become indistinguishable.

The diversity of optimal organizational forms matters for the way organizations respond to shocks. In particular, the response of key observable characteristics such as wages and size of layers depend on whether the organization is horizontal or vertical. Moreover, a fall in h causes organizations to become more vertical, whereas a fall in ξ causes organizations to become more horizontal. This and other implications of our results for empirical work are discussed in Section 5.5.

5.1 Garicano (2000) revisited

As we have pointed out, Garicano's (2000) setting is obtained when $\pi = h$ and $B_l = A_l$ for each $l \in L$. The latter condition is, however, endogenous: It is up to the organization to decide on whether or not the screening set of a layer equals its knowledge set. As we will see in Section 5.2, if ξ is small relative to c, then optimal organizations will not feature $B_l = A_l$ for each $l \in L$.

Here we show that if, in addition to $\pi = h$, the cost ξ of learning how to screen problems equals the cost c of leaning how to solve them, then we have that $B_l = A_l$ for each $l \in L$. Indeed, in this case, there is no point in screening: Solving problems costs the same (in terms of time $h = \pi$ and learning costs $c = \xi$) as screening but, in contrast to screening, solving problems adds to the output of the organization.

Once we reach the conclusion that $B_l = A_l$ for each $l \in L$, we are in the framework of Garicano (2000) and we obtain his conclusions from our results. Namely, there is vertical specialization in which the most frequent problems are solved by layer 1 (the workers), the next most frequent by layer 2 (the lowest level of managers) and so on. The organization is pyramidal in the sense that layer 2 has more people than layer 3 and so on, so that higher levels of management have smaller numbers of people. If, in addition, $h < 1$ or $A_1 \neq \emptyset$, then layer 1 also has more people than layer 2.

Corollary 3 *If O is an optimal organization, $\pi = h$ and $\xi = c$, then*

(a) $B_l = A_l$ *for each* $l \in L$,

(b) $A_1 < \cdots < A_L$,

(c) $\beta_2 > \cdots > \beta_L$ *and*

(d) $\beta_1 > \beta_2$ *if* $h < 1$ *or* $A_1 \neq \emptyset$.

5.2 Optimal organizations when ξ is small

We now contrast the conclusions of Corollary 3 with those that arise when the cost of learning how to screen problems is sufficiently small. In this case, the optimal organization features horizontal rather than vertical specialization: effectively, workers solve the most common problems and send each remaining one to a manager who can solve it, without wasting the time of any manager who cannot.

We characterize optimal organizations when ξ is small under the following two assumptions. When Ω is bounded, let $\bar{\omega} \in \mathbb{R}_{++}$ be such that $\Omega = [0, \bar{\omega})$ and $f(\bar{\omega}) = \lim_{\omega \to \bar{\omega}^-} f(\omega)$. Our first assumption requires the marginal value of knowledge to be bounded away from zero.

(A1) Ω is bounded and $f(\bar{\omega}) > 0$.

Our second assumption requires the marginal value of knowledge at zero to be above the cost of learning how to solve problems so that it always pays to learn how to solve some problems.

(A2) $f(0) > c$.

Theorem 5 describes all the possible optimal organizations when ξ is small. We focus on η-optimal organizations because lexicographically optimal organizations may fail to exist; nevertheless, lexicographically optimal organizations also satisfy most of these properties when they do exist. We vary ξ and keep all the remaining parameters (Ω, f, c, h, π and η) fixed in such a way that the basic assumptions of Section 3, as well as (A1) and (A2), are satisfied.

Theorem 5 *If (A1) and (A2) hold, then there exists $\bar{\xi} > 0$ such that, if $0 < \xi < \bar{\xi}$ and O is an η-optimal organization with $L \geq 2$, then:*

1. $B_1 \setminus A_1 = (A_1 \cup A_2)^c$,

2. $B_l \setminus A_l = \emptyset$ *for each* $l > 1$,

3. $A_1 < A_l < A_2 < (B_1 \setminus A_1) \cap (\cap_{j>2} A_j^c)$ *for each* $l > 2$,

4. *there exists a bijection* $g : \{3, \ldots, L\} \to \{3, \ldots, L\}$ *such that* $A_{g(3)} < \cdots < A_{g(L)}$,

5. $\mu(A_{g(3)}) < \cdots < \mu(A_{g(L)}) < \mu(A_2)$, $F(A_{g(3)}) > \cdots > F(A_{g(L)}) > F(A_2) - \frac{\xi}{ch}$ *and* $F(A_{g(L)}) > \eta$, *and*

6. $\beta_1 > \sum_{l=2}^{L} \beta_l$ *if* $h < 1$.

In addition, for each bijection $g : \{3, \ldots, L\} \to \{3, \ldots, L\}$, *there is an optimal organization* O *such that* $A_{g(3)} < \cdots < A_{g(L)}$.

The organizations described in Theorem 5 differ considerably from those of Corollary 3 (obtained when $\xi = c$ and $\pi = h$) when $L \geq 2$. When ξ is sufficiently small, layer 1 (the workers) learns how to screen all problems it cannot solve except those that layer 2 knows how to solve. There is no point in layer 1 screening problems that layer 2 can solve since any problem that layer 1 does not screen already gets passed to layer 2. Even though layer 1 does not formally screen the problems that layer 2 can solve (i.e. $B_1 \cap A_2 = \emptyset$), effectively layer 1 knows that layer 2 can solve them since $A_2 = B_1^c$. Thus, all the problems that layer 1 faces, which, due to specialization, are those faced by the organization, are passed directly to the layer that can solve them or are abandoned if no one can solve them.[16]

Although there are several optimal organizations when $L > 3$, these are equivalent up to the relabelling of layers. Thus, we can assume that $A_3 < \ldots < A_L < A_2$. Under this normalization (or if $L \leq 3$), layer 1 knows how to solve the most frequent problems as in Corollary 3 but it is layer 3, not layer 2, that solves the next most frequent problems. In fact, layer 2 solves the least frequent problems that the organization solves.[17] Moreover,

[16] See the rightmost part of Figure 1 for a graphical illustration of the resulting flow of information when $L = 4$ and no problems are abandoned. Figure 1, in its leftmost part, also illustrates the flow of information in a hierarchy. See Section 5.3 for details on the computations behind the figure.

[17] We can interpret layer 2 as the top managers who deal only with exceptional problems that no one else in the organization can even identify.

layer 2 is the most knowledgeable of the managers in the sense of having the largest knowledge set. Problems that are screened but not solved are the least frequent of all.

Some intuition for the above properties is as follows. Layer 1 screens all problems except those that layer 2 solves because screening is sufficiently cheap. Screening such problems by layer 1 saves the time of managers in layer 2; since the marginal cost of screening goes to zero as ξ goes to zero, it is optimal for layer 1 to screen all such problems. Hence, for each $l \neq 1$, layer l does not screen any problem that it does not solve and, since all problems are sent to the correct layer, $\alpha_l = hF(A_l)$, i.e. all its helping time is devoted to solving the problems in its knowledge set.

Layer 2 is the most knowledgeable of the managers in order to minimize the learning costs of screening: Since layer 1 screens all problems it cannot solve except those solved by layer 2, its screening cost is $\xi(\bar{\omega} - \mu(A_1) - \mu(A_2))$; hence the size of layer 2's knowledge set $\mu(A_2)$ should be at least as large as $\mu(A_l)$ for each $l \geq 3$ because these layers are otherwise interchangeable. In fact, we show that this inequality must be strict for η-optimal organizations.

It then follows that $A_l < A_2$ for each $l \geq 3$. If not, we can swap a more common problem from A_2 with a less common one from A_l such that the sizes of the two sets remain $\mu(A_2)$ and $\mu(A_l)$ respectively, but the frequency of problems sent to layer 2 decreases by the same amount that the frequency of problems sent to layer l increases. Ultimately, this allows the organization to transfer people to layer l from layer 2 and decrease total learning costs of the organization, which are $\beta_1(c\mu(A_1) + \xi\mu(B_1 \setminus A_1)) + \sum_{i=2}^{L} \beta_i c\mu(A_i)$, since $\mu(A_2) > \mu(A_l)$ for each $l \geq 3$.

Problems that are screened but not solved are the rarest problems of all, i.e. $A_2 < (B_1 \setminus A_1) \cap (\cap_{j>2} A_j^c)$. This follows from Theorem 3: we have $c_{A_2} = c\alpha_2 > \xi = c_{(B_1 \setminus A_1) \cap (\cap_{j>2} A_j^c)}$ because α_2 is bounded away from zero as ξ goes to 0. Indeed, if not, then $\alpha_2 = hF(A_2)$ would converge to zero with ξ and, thus, so would $\mu(A_2)$; since $\mu(A_2) = \max_{2 \leq l \leq L} \mu(A_l)$, the output of the optimal organization would converge to the output of an organization with just one layer, contradicting the assumption that the optimal organization has at least two layers.[18]

Regarding whether the optimal organization is pyramidal, recall that $\alpha_l = hF(A_l)$ for

[18]Since there is a cost η of adding layers, if the output of the optimal organization converges to the output of the organization with one layer, eventually the optimal organization must have one layer.

each $l \neq 1$. If $h < 1$, it then follows by part 5 of Theorem 5 that $\beta_1 > \beta_3 > \cdots > \beta_L > \left(\frac{ch\eta}{ch\eta+\xi}\right)\beta_2$.[19] Thus, the organization has a *quasi-pyramidal* structure in the sense that the more infrequent the problems that a group of managers solves, the smaller the group is; the "quasi" qualification is then needed because β_2 may be bigger than β_L but not by much when ξ is small.

However, the optimal organization is not a hierarchy when ξ is small because there is no flow of information along the organization; instead the information is directed by layer 1 to the layer that can solve each problem. Therefore, the relevant comparison is between the size of the workers β_1 and that of the managers $\sum_{l=2}^{L}\beta_l$. As Theorem 5 shows, the former is bigger than the latter. Thus, in this sense, the optimal organization has a pyramidal structure.

The increase in the expected output of the optimal organization as compared with the best hierarchy can be significant: When $\Omega = [0,1)$, $f(\omega) = 1.5 - \omega$ for each $\omega \in [0,1)$, $c = 1.4$, $h = 0.5$, $\eta = 0.01$ and $\xi = \bar{\xi}/2$, where $\bar{\xi} \simeq 0$ is given by Theorem 5, the expected output net of costs of adding layers Y (expected net output, henceforth) of the optimal organization is 83.59% higher than that of the best hierarchy (i.e. the organization satisfying the properties of Corollary 3 that, at the same parameter values, maximizes expected net output).[20] Thus, screening problems can lead to a significant improvement in the organization of knowledge in production.

In the above example, there are $L = 8$ layers in the optimal organization, expected output is $y \simeq 0.60$ and expected net output is $Y \simeq 0.53$. Moreover, with $\mu = (\mu_1, \ldots, \mu_8)$ and $\mu_i = \mu(A_i)$ for each $1 \leq i \leq 8$,

$$\mu \simeq (0, 0.18, 0.12, 0.12, 0.13, 0.14, 0.15, 0.16) \text{ and}$$
$$\beta \simeq (0.5, 0.05, 0.08, 0.08, 0.08, 0.07, 0.07, 0.06).$$

In contrast, the best hierarchy has $L = 3$ layers, expected output is $y \simeq 0.31$, expected net output is $Y \simeq 0.29$, $\mu \simeq (0.05, 0.51, 0.44)$ and $\beta \simeq (0.45, 0.41, 0.14)$. The fraction of workers is similar in the two organizations (between $45\% - 50\%$) and workers know how

[19]To see this, note that $\beta_l = \frac{\alpha_l}{\gamma} = \frac{hF(A_l)}{\gamma}$ for each $l > 1$; then $F(A_L) > F(A_2) - \frac{\xi}{ch}$ implies that $\beta_2 < \beta_L + \frac{\xi}{c\gamma} = \left(1 + \frac{\xi}{chF(A_L)}\right)\beta_L < \left(1 + \frac{\xi}{ch\eta}\right)\beta_L$ since $F(A_L) > \eta$.

[20]Specifically, $\xi = 6.281367579349382e{-}10$. If we focus instead on expected output y, then the increase in output obtained by moving from the best hierarchy to the optimal organization raises to 94.34%. The python codes used in this computation are in the supplementary material to this paper.

to solve only a small fraction of problems ($0\% - 5\%$). The latter arises because the cost $c = 1.4$ of learning how to solve problems is relatively high: Indeed, recall that $y = \frac{\theta}{\gamma}$, and note that the coefficient of μ_1 in θ is c and that of μ_l, for each $l > 1$, is $c\alpha_l \leq ch = 0.7$. This makes it important to add layers to reduce the costs of knowledge. In addition, when layer 1 screens all the problems that its members and those of layer 2 cannot solve, the organization of knowledge is more efficient as manifested in a decrease in the amount of time per problem that layers $l > 1$ spend helping layer 1.

To see the two effects in detail, consider, in addition to the optimal organization O^* and the best hierarchy O and respective expected net outputs Y^* and Y, an organization $\hat{O}$ with $\hat{L} = L = 3$, $\hat{A}_1 = A_1 = [0, 0.05)$, $\hat{A}_2 = A_2 = [0.05, 0.56)$ and $\hat{A}_3 = A_3 = [0.56, 1)$ as in O but with $\hat{B}_1 \setminus \hat{A}_1 = [0, 1) \setminus (\hat{A}_1 \cup \hat{A}_2) = \hat{A}_3$, $\hat{B}_2 = \hat{A}_2$ and $\hat{B}_3 = \hat{A}_3$ as in O^*. Then, $\hat{\alpha}_2 = hF(A_2) < h(F(A_2) + F(A_3)) = h(1 - F(A_1)) = \alpha_2$ and $\hat{\alpha}_3 = hF(A_3) = h(1 - F(A_1) - F(A_2)) = \alpha_3$, showing that by adding screening reduces the helping costs of layer 2. This results in an increase of net output of $(\hat{Y} - Y)/Y = 48\%$ or 57% of the total increase $(Y^* - Y)/Y$. The remaining 43% of this increase are due to the increase in the number of layers of $\hat{O}$ to reach O^* and corresponding optimal choice for their knowledge sets.

Theorem 5 describes the optimal organizations when there are at least two layers. If there is only one layer, then the organization is trivial, with $B_1 = A_1 = [0, \mu_1^*)$ where μ_1^* solves $\max_{0 \leq \mu_1 \leq \bar{\omega}}(F(\mu_1) - c\mu_1)$ and the resulting output is $y_1 = F(\mu_1^*) - c\mu_1^*$, where $F(\omega) = F([0, \omega))$ for each $\omega \in \Omega$. When it is optimal to have only one layer, there is, strictly speaking, no role for organization. In the supplementary material to this paper we show that if communication is not too costly (i.e. $h < 1$) and organizations with one layer do not have full knowledge (i.e. $f(\bar{\omega}) < c$), then any optimal organizations has at least two layers when ξ is sufficiently small and (A1) holds. We also show that under (A1), (A2) and these two conditions, there is no lexicographically optimal organization when ξ is sufficiently small.

5.3 Optimal organizations for intermediate values of ξ

Our previous results fully characterize optimal organizations when the cost ξ of learning how to screen problems is equal to the cost c of learning how to solve them and also when ξ is sufficiently close to zero. In the former case, the optimal organization is a hierarchy featuring vertical specialization; in the latter it is the one characterized by Theorem 5, a *fully screening organization* henceforth, featuring horizontal specialization. Our goal in this section is to illustrate numerically the properties of the optimal organization for intermediate values of ξ and, in particular, how the optimal organization transforms from a hierarchy to a fully screening organization as ξ declines from c to zero.

The main conclusion is that, for intermediate values of ξ, novel organizational forms arise as optimal. For most of our simulations, the optimal organization for intermediate values of ξ is a *hybrid organization* illustrated in the middle part of Figure 1, which contrasts it with both the hierarchic and the fully screening organizations that are optimal when ξ is not intermediate (see below for details on the computations and parameter values). In such hybrid organization, some problems are screened by layer 1 and sent

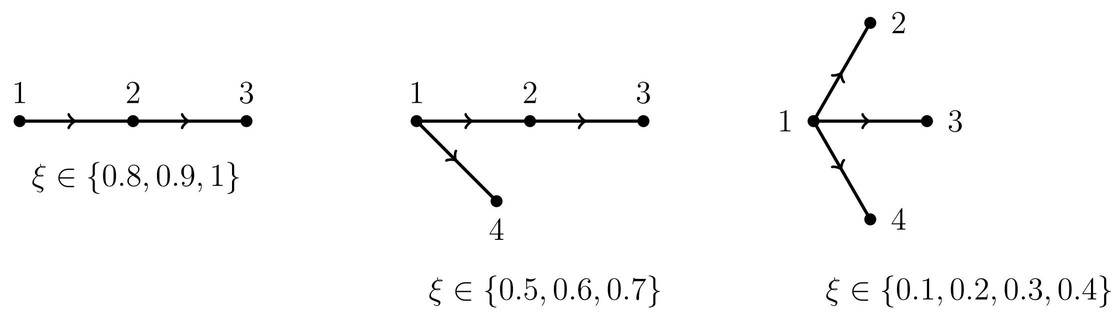

Figure 1: Flow of information in optimal organizations when $(c, h, b) = (1, 0.5, 1)$.

directly to a layer, layer 4, that solves only these problems; in contrast, the problems that layer 1 neither solves nor screens pass through the remaining layers as in a hierarchy, i.e. are sent to layer 2, which in turn send those it cannot solve to layer 3.

The above three organizational forms can, broadly speaking, match real-world organizations. The following scenario illustrates the three of them and the transition between organizational forms. Consider a firm in the services sector (e.g. an online wine shop) organized as a hierarchy with three layers. As ξ falls (which could happen over time as the organization gets better at classifying the problems it faces), it might make sense to create a new layer, layer 4, by having layer 1 learn how to distinguish between problems that are IT-specific (e.g. when the webpage is not working properly) and those that aren't; the resulting organization is then hybrid, with layer 1 sending IT-specific problems directly to IT managers (layer 4) and non-IT-specific problems to layer 2, which in turn will send those problems that fall outside their remit (e.g. issues with product design or marketing) to layer 3. As ξ falls even further, it might likewise make sense to have layer 1 learn how to distinguish the problems (e.g. accounting issues) that layer 2 can solve from those that layer 3 can, thus obtaining a fully screening organization with customer services (layer 1) dealing with initial queries and directing them to the IT, accounting or marketing departments as appropriate.

As well as rationalizing the structure of real-world organizations, our simulations generate predictions about the response of empirically relevant variables (e.g. the wage and size of each layer) to changes in the screening cost. In the remainder of this section we discuss how the optimal organization changes as a function of the cost ξ of learning how to screen problems. In particular, we will be interested in (i) the knowledge and screening sets of each layer for different values of ξ and their ordering with respect to the frequency of the problems they contain, (ii) comparing the wage and size distribution across different organizational forms, and (iii) comparing the response to changes in ξ across different organizational forms.

The conclusion that an optimal organization is a hierarchy when $\xi = c$ requires also that $\pi = h$, which we assume throughout this section. Even when the number of layers is only 4, the number of possible orderings consistent with our results is already quite large. For this reason, and because the maximum number of layers that have been used in empirical studies is 4, we assume that $\bar{L} = 4$ is the maximum number of possible layers. We focus on lexicographically optimal organizations which, given such an upper bound on the number of layers, always exist.[21] For our simulations, we consider $\Omega = [0, 1)$ and an affine density f defined by $f(\omega) = a - b\omega$ for each $\omega \in \Omega$, with $b > 0$ for f to be strictly decreasing and $a = \frac{2+b}{2}$ for f to be a density. We set $b = 1$, $h = 0.5$, and $c = 1$ (the baseline case henceforth).[22]

Figure 2 illustrates the optimal organization at different values of ξ.[23] The panels on the top row show the size of each element of C that is nonempty for some value of ξ. When $\xi \in \{0.8, 0.9, 1\}$, the optimal organization is a hierarchy with 3 layers. When $\xi \in \{0.1, 0.2, 0.3, 0.4\}$, the optimal organization is fully screening with 4 layers, the order being $A_1 < A_4 \cap (B_1 \setminus A_1) < A_3 \cap (B_1 \setminus A_1) < A_2$.[24] In each case, the ordering and the

[21]See Section A.10 in the Appendix for details on computing lexicographically optimal organizations.

[22]Other configurations of parameters are considered in the supplementary material to this paper.

[23]Variables related to layer 1 are red; layer 2, blue; layer 3, yellow; and layer 4, black.

[24]As Theorem 5 shows, the names of layers 3–L in a fully screening organization are arbitrary. So we could equally well relabel the layers to have $A_1 < A_3 \cap (B_1 \setminus A_1) < A_4 \cap (B_1 \setminus A_1) < A_2$ but the

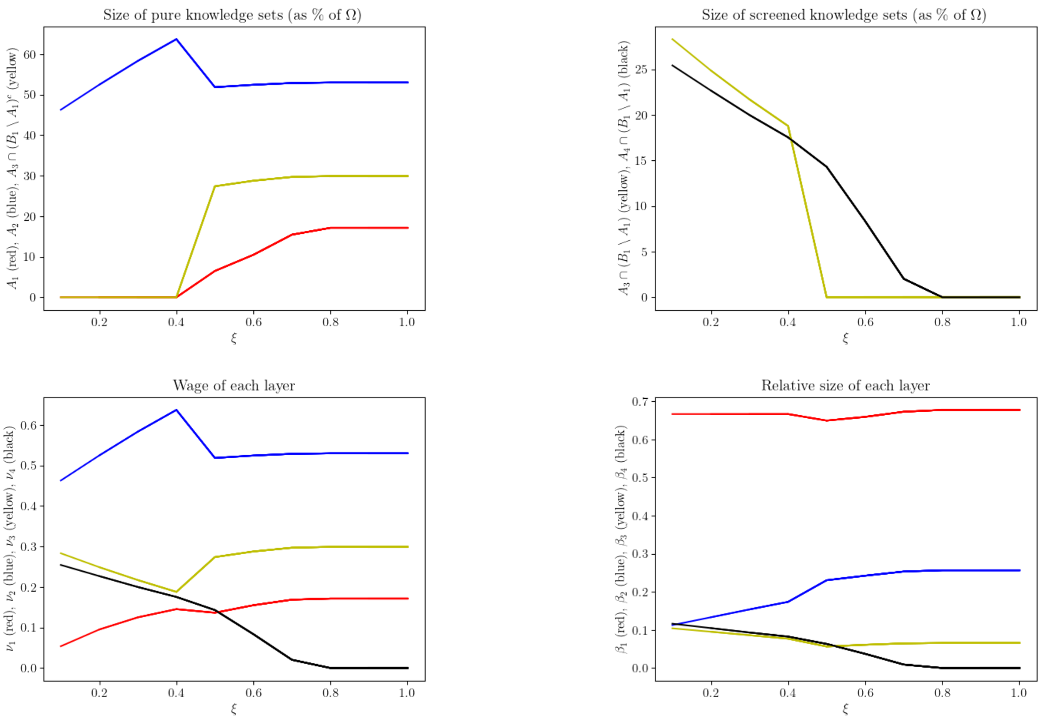

Figure 2: Optimal organization for different values of ξ when $(c, h, b) = (1, 0.5, 1)$.

sizes of the knowledge and screening sets, as well as the relative sizes and learning costs of each layer, are as in the theoretical results presented in Sections 5.1 and 5.2.

Comparing hierarchies to fully screening organizations: Regarding learning costs $\nu_i = c\mu(A_i) + \xi\mu(B_i \setminus A_i)$ as a measure of wages (as in e.g. Caliendo, Monte, and Rossi-Hansberg (2015)), there is greater wage inequality between workers and managers in fully screening organizations, and this inequality is increasing as screening becomes cheaper. Inequality among managers is greatest when the organization is fully screening but ξ is relatively high (in that case, the organization uses the knowledge of the top managers as much as possible to save on screening costs). The number of workers (and hence managers) in the organization is roughly constant as the cost of screening changes; however, for fully screening organizations, managers are more equally distributed across the different layers than in a hierarchy, and this distribution becomes even more equal as the cost of screening falls.

For intermediate values of the screening cost, i.e. when $\xi \in \{0.5, 0.6, 0.7\}$, the optimal organization takes the hybrid form. It has 4 layers; only layer 1 screens problems that it does not solve, all such problems are solved by layer 4, and layer 4 only solves such

ordering we adopted here is more convenient for the comparison of the different organizational forms in this section.

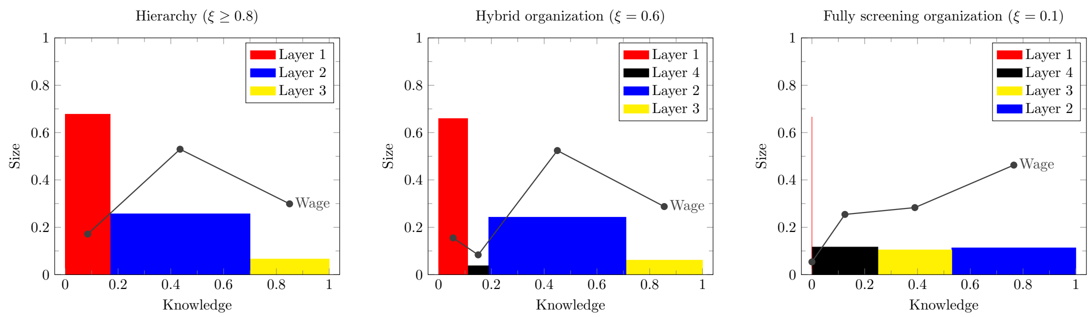

Figure 3: Knowledge, size and wage of layers when $(c, h, b) = (1, 0.5, 1)$.

problems: $B_1 \setminus A_1 = A_4 \cap (B_1 \setminus A_1) = A_4$. Layers 2 and 3 solve the problems that layer 1 does not screen. The order is $A_1 < A_4 \cap (B_1 \setminus A_1) < A_2 < A_3 \cap (B_1 \setminus A_1)^c$ and, thus, layer 1 solves the most common problems, followed by layers 4, 2 and 3 — layer 3 is, in this sense, the one with the top managers of the organization.

Figure 3 shows, for the three types of organization, the knowledge sets of each layer on the horizontal axis and the corresponding size and wage of the layer on the vertical axis. It can clearly be seen that for the hybrid organization, the expansion of $A_4 \cap (B_1 \setminus A_1)$ is made mostly at the expense of A_1: The least common problems that were solved by the workers in layer 1 when $\xi = 0.8$ (i.e. when the optimal organization was still a hierarchy), are now being solved by the managers in layer 4. When screening becomes sufficiently cheap, i.e. when $\xi = 0.4$, the organization becomes fully screening with $A_1 < A_4 \cap (B_1 \setminus A_1) < A_3 \cap (B_1 \setminus A_1) < A_2$. In this case, layer 1 screens some problems that were previously solved by layer 2 and sends them directly to layer 3; layer 3 solves only these problems, with layer 2 taking over the rarest problems.

Recall that the distribution of problems F has a strictly decreasing density; thus in Figure 3 the layers appear on the horizontal axis in decreasing order with respect to the frequency of the problems they solve. It can then be seen that only hierarchies are pyramidal in the sense of layers solving less frequent problems having fewer members. However, regarding layer 3 as the top managers in the hybrid organization (as they are the ones who solve the least frequent problems) and layers 2 and 4 as those of middle managers, the organization is then pyramidal in the sense of $\beta_3 < \beta_2 + \beta_4 < \beta_1$ but not in the sense of $\beta_3 < \beta_2 < \beta_4 < \beta_1$. When the optimal organization is fully screening, it is quasi-pyramidal as defined in Section 5.2 (i.e. $\beta_3 < \beta_4 < \beta_1$ reflecting $A_1 < A_4 \cap (B_1 \setminus A_1) < A_3 \cap (B_1 \setminus A_1)$) and is pyramidal in the sense that there are more workers than managers (i.e. $\sum_{i=2}^{4} \beta_i < \beta_1$).[25] In addition, only in fully screening organizations do wages reflect the ordering of the layers with respect to the rarity of the problems they solve. In particular, when the hierarchy is optimal, $A_1 < A_2 < A_3$ but $\nu_1 < \nu_3 < \nu_2$.[26]

The response of the wage and size of each layer to a fall in the cost of screening also differs by organizational form. As screening becomes cheaper over the range where the optimal organization is hybrid, the wage of layer 4 increases while the wages of the other layers decrease. In contrast, when the optimal organization is fully screening, a fall in the cost of screening increases the wages of layers 3 and 4 and decreases those of layers 1 and 2. Cheaper screening means that the organization screens more problems; thus the wages of those layers that solve the screened problems go up. In addition, as screening becomes cheaper, the wage of the layer that does the screening falls.

As for personnel changes in the hybrid organization as the cost of screening falls, there is an increase in the number of people in the new layer 4 at the expense of all the other layers, but mostly of layers 1 and 2. In contrast, for fully screening organizations, a fall in the cost of screening increases the number of managers in layers 3 and 4 at the expense of the number of managers in layer 2. Cheaper screening means that the organization screens more problems; thus, the layers that solve these problems increase in size.

In summary, as the cost of screening falls, we see larger organizations (i.e. an increase in the number of layers), direct flows of information (in addition or instead of sequential flows), quasi-pyramidal organizations where the layer of top managers may be larger than the previous one, and less knowledgeable workers.

In the supplementary material to this paper, we consider other combinations of the parameters in comparison to the baseline case and observe qualitatively similar patterns. One interesting case that provides some additional insight is when the density is steeper than in the baseline case ($b = 1.9$). Here, the transformation of a hybrid organization into a fully screening one has some intermediate steps, illustrated in Figure 4. In this case, the optimal organization is hierarchical when $\xi \in \{0.6, \ldots, 1\}$, hybrid when $\xi \in \{0.4, 0.5\}$

[25] The emergence of novel optimal organizational forms other than hierarchies implies adjustments to standard notions such as that of pyramidal structure, which is specific to each organizational form.

[26] We have checked all parameter values such that, when $\xi = c$, the optimal organization has 3 layers in the range $c \in \{0.1, 0.2, \ldots, 2\}$, $h \in \{0.1, 0.2, \ldots, 1\}$ and $b \in \{0.1, 1, 1.9\}$ such that $b > 2(c-1)$ (the latter implying that the best organization with 1 layer has strictly positive output and improves the efficiency of our code). In none of them did we obtain $\nu_1 < \nu_2 < \nu_3$.

but only fully screening when $\xi \leq 0.02$.

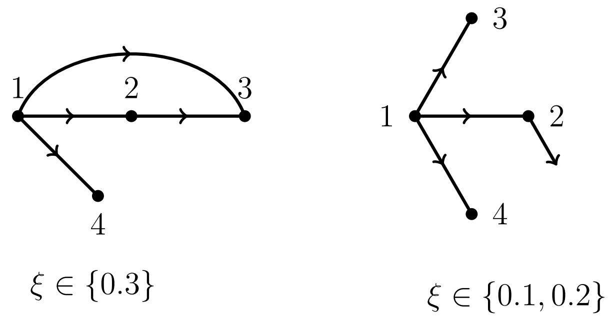

Figure 4: Flow of information in optimal organizations when $(c, h, b) = (1, 0.5, 1.9)$.

When $\xi = 0.3$, layer 1 screens and sends directly to layer 3 some problems that were solved by layer 2. However, screening is not sufficiently cheap for layer 1 to screen all problems to be solved by layer 3; hence, both $A_3 \cap (B_1 \setminus A_1)$ and $A_3 \cap (B_1 \setminus A_1)^c$ are nonempty. The order is $A_1 < A_4 \cap (B_1 \setminus A_1) < A_3 \cap (B_1 \setminus A_1) < A_2 < A_3 \cap (B_1 \setminus A_1)^c$.

When $\xi \in \{0.1, 0.2\}$, screening is now sufficiently cheap for layer 1 to screen all problems solved by layers 3 and 4, as in a fully screening organization, but $(B_2 \setminus A_2) \cap A_4^c$ is nonempty. This is because when $b = 1.9$, some problems are so rare that they are not worth solving when it is sufficiently cheap to screen them. The reason that it is layer 2 screening these problems instead of layer 1 is the usual trade-off between layer 1's learning costs and layer 2's helping time. The order is $A_1 < A_4 \cap (B_1 \setminus A_1) < A_3 \cap (B_1 \setminus A_1) < A_2 < (B_2 \setminus A_2) \cap A_4^c$. For a real-world example, consider an organization that uses a chatbot to deal with customers' queries. Customers with well-defined problems are sent to the appropriate department; those who select "all other queries" are passed on to a general customer service agent. If the problem is very rare, nothing can be done – but crucially, it is the customer service agent, not the chatbot, who is able to decide which customers to send away.

5.4 The case of small helping costs h

When h is close to zero, communicating solutions to problems is very cheap and one of the trade-offs determining the optimal degree of specialization is effectively shut down. This allows us to derive additional characteristics of optimal organizations. The case of small h may also be relevant to rationalize changes to firms' organization and the wage distribution as argued in Garicano and Rossi-Hansberg (2015) since recent decades have witnessed an improvement in communication technology.

Theorem 6 below establishes that, when h is small, optimal organizations have full knowledge and its workers have no knowledge. In other words, it shows that the fraction of problems that are abandoned (i.e. not solved) by an optimal organization is zero and that there is full specialization in the sense that managers (i.e. layers other than layer 1) specialize not only in helping layer 1 but also in learning how to solve or screen problems. This latter aspect may be interpreted as automation or as quite mechanical work by the workers. In addition, for small enough $\eta > 0$, η-optimal organizations are fully characterized by the organization O_2^* with two layers, full knowledge and full specialization in knowledge acquisition: $L = 2$, $A_1 = B_1 = \emptyset$ and $A_2 = B_2 = \Omega$. In general, when h is low, managers have a very high span of control[27] and their wages, measured by learning costs, relative to those of workers are also very high — managers are superstars.

Analogously to Theorem 5, we vary h and keep all the remaining parameters (Ω, f, c, ξ, π and η) fixed in such a way that the basic assumptions of Section 3, as well as (A1), are satisfied. Recall that y_1 is the highest output obtained by an organization with only one layer.

Theorem 6 *If (A1) and $\eta < 1 - y_1$ hold, then there exists $\bar{h} > 0$ such that, if $0 < h < \bar{h}$ and O is an optimal organization, then:*

1. *$F(\cup_{l \in L} A_l) = 1$,*

2. *$B_1 = \emptyset$, and*

3. *$O = O_2^*$ if O is η-optimal.*

Part 3 of Theorem 6 does not extend to the case of lexicographically optimal organizations. This follows from Theorem 7 below which focuses on the simpler case where optimal organizations are hierarchies (i.e. $\xi = c$ and $\pi = h$). It roughly shows that, when h is sufficiently small, lexicographically optimal organizations have two layers when $\frac{c}{f(\bar{\omega})}$ is small, three layers when $\frac{c}{f(\bar{\omega})}$ is intermediate and four or more layers when $\frac{c}{f(\bar{\omega})}$ is high.[28] This may explain why many firms in construction have two layers (workers

[27]By Theorem 1, for each $l \in L \setminus \{1\}$, $\frac{\beta_l}{\beta_1} = \alpha_l \leq h$.

[28]This paper is agnostic about which one of the two notions of optimality is better, the punch line here being that they are equivalent for the general properties of optimal organizations. However, as the contrasting conclusions of Theorems 6 and 7 show, the choice matters for some specific questions.

and manager/owner) whereas, for some very specific medical problems, patients need to consult with many doctors before a solution is found since $\frac{c}{f(\bar{\omega})}$ is likely to be considerably higher in the latter case.

Theorem 7 uses the following notation. For each $L \in \mathbb{N}$, let y_L be the expected output of the best organization with L layers. Moreover, let O_2^* be as above and O_3^* be the hierarchy with three layers defined by $A_1 = \emptyset$, $A_2 = [0, \mu_2^*)$, and $A_3 = [\mu_2^*, \bar{\omega})$, where μ_2^* is the unique solution to $\mu_2 + \frac{F(\mu_2)}{f(\mu_2)} = \frac{y_3}{c} + \bar{\omega}$.

Theorem 7 *If (A1), $\xi = c$ and $\pi = h$ hold, then there exists $\bar{h} > 0$ such that the following holds for each $0 < h < \bar{h}$:*

1. *A lexicographically optimal organization exists.*

2. *O_2^* is a lexicographically optimal organization if and only if*

$$\frac{c}{f(\bar{\omega})} \leq \sup_{L \in \mathbb{N}} y_L. \tag{3}$$

   In addition, under (3), O_2^* is the unique optimal organization and (3) holds if and only if $\frac{c}{f(\bar{\omega})} \leq y_2$.

3. *Suppose that f is differentiable and*

$$\frac{\partial \left(\frac{F(x+y)-F(x)}{f(x+y)} \right)}{\partial x} > -1 \tag{4}$$

   for each $x, y \in \Omega$ such that $x + y \leq \bar{\omega}$. Then, O_3^* is a lexicographically optimal organization if and only if

$$\frac{c(1 - F(\mu_2^*))}{f(\bar{\omega})} \leq \sup_{L \in \mathbb{N}} y_L < \frac{c}{f(\bar{\omega})} \tag{5}$$

   holds. In addition, under (5), O_3^* is the unique optimal organization and (5) holds if and only if $\frac{c(1-F(\mu_2^*))}{f(\bar{\omega})} \leq y_3 < \frac{c}{f(\bar{\omega})}$.

The first part of Theorem 7 establishes the existence of lexicographically optimal organizations by effectively showing that the optimal number of layers is finite. As we show in its proof, the condition needed for this, on top of (A1), is that $f(0) > \min\{ch, c\}$, which follows from (A1) when h is sufficiently small. Requiring the set Ω of possible problems to be bounded is really only needed for $f(\bar{\omega}) > 0$ to be possible, which is thus the main condition.[29]

Theorem 7 partitions the parameter space according to whether the ratio $\frac{c}{f(\bar{\omega})}$ of the marginal cost to the marginal benefit of knowledge accumulation is low, medium or high. This is easily seen in the limit as $h \to 0$ since, due to $\sup_L y_L \to 1$ as $h \to 0$, a sufficient condition for (3) to hold when h is sufficiently small is $\frac{c}{f(\bar{\omega})} < 1$ and for (5) is $1 < \frac{c}{f(\bar{\omega})} < \frac{1}{1-F(\bar{\mu}_2)}$, where $\bar{\mu}_2$ is the unique solution to $\mu_2 + \frac{F(\mu_2)}{f(\mu_2)} = \frac{1}{c} + \bar{\omega}$. Accordingly, the optimal number of layers will be two, three or four or above if the ratio of the marginal cost to the marginal benefit of knowledge accumulation is low, medium or high.[30]

5.5 Implications for empirical research

As already mentioned in the introduction, the emergence of optimal organizational forms other than hierarchies broadens the applicability of the theory of knowledge-based organizations. In particular, our model can help rationalize common organizational structures (such as specialized HR or IT departments in organizations that are otherwise hierarchical) and delivers testable predictions regarding the response of different organizations to shocks such as an improvement in communication technology.

As well as rationalizing particular organizations, our results may provide an explanation for the diversity of real-world organizational forms by tracing it to differences in the parameters of the model (i.e. Ω, f, c, ξ, h and π). This raises the following empirical question: Are there observable characteristics of firms that explain the differences in their organizational forms?

Answering the above question requires a way of empirically identifying the organizational form of each organization. Our results also suggest that the empirical definition of layers may need to be significantly different from the usual one, used in e.g. Caliendo, Monte, and Rossi-Hansberg (2015), which classifies occupations according to whether they are lower, middle, or upper management. For example, in the hybrid organization with 4 layers that is optimal for intermediate values of ξ in Section 5.3, workers (layer 1) pass

[29] Garicano (2000, Section III) presents an example where the number of layers is infinite, i.e. of no existence of optimal hierarchies. In it, $\Omega = \mathbb{R}_+$ and thus (A1) fails.

[30] We note that condition (4) that is needed in Theorem 7 is not too demanding. As discussed in Section A.15 in the Appendix, each one of the following two conditions alone is sufficient (but not necessary) for it: (i) $f(0) < 2f(\bar{\omega})$, and (ii) f is concave.

some problems directly to layer 4, which solves only these problems. Other problems, which the workers cannot identify, are passed on to layer 2, and some of these are passed on to layer 3. Layers 2 and 4 both constitute lower forms of management in the sense that their members communicate directly with workers. However, these managers will differ in their knowledge sets and therefore wages. Thus, it may not be appropriate to place them into the same layer.

The above suggests that a more firm-specific approach to the definition of layers might be needed so that subtle distinctions between layers, such as between layers 2 and 4 in the above example, are captured. The ideal scenario would be to have information about the way each firm operates, namely on who does what, who knows what, who communicates with whom and even on the nature of knowledge its members have, i.e. whether it is general or specialized knowledge.[31] The importance of the latter for placing members of the organization in a specific layer can be seen by noting that members of layer 1 are the least knowledgeable in hierarchies but they are the ones with the highest general knowledge in fully screening organizations.

The consideration of novel organizational forms may shed new light on important empirical questions such as the impact of shocks on firms' activity. In Section 5.3, we have already seen how a fall in the screening cost ξ may cause the firm to change its organizational structure. Moreover, different organizational forms respond differently to such shock. We now consider the impact of a fall in helping cost h (for example, corresponding to an improvement in communication technology), comparing the responses of different organizations and contrasting the predictions with those obtained from a fall in the screening cost ξ.

Figure 5 illustrates the changes in the optimal organization in the baseline case where $b = 1$, $h = 0.5$, and $c = 1$ in response to a 10% decline in the helping cost h (its new value is then 0.45).[32] The main patterns we observe are that hierarchies and hybrid organizations increase the knowledge of all managers at the expense of the knowledge of workers, but with small changes in personnel; on the other hand, fully screening organizations increase the number of workers at the expense of managers, but with small changes in knowledge (and only the knowledge of top managers increase as unscreened knowledge becomes

[31] For example, by looking at email and meeting metadata as in Impink, Prat, and Sadun (2022) or, at a small scale, by using field experiments as in Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013).

[32] See the supplementary material for results for other parameter configurations.

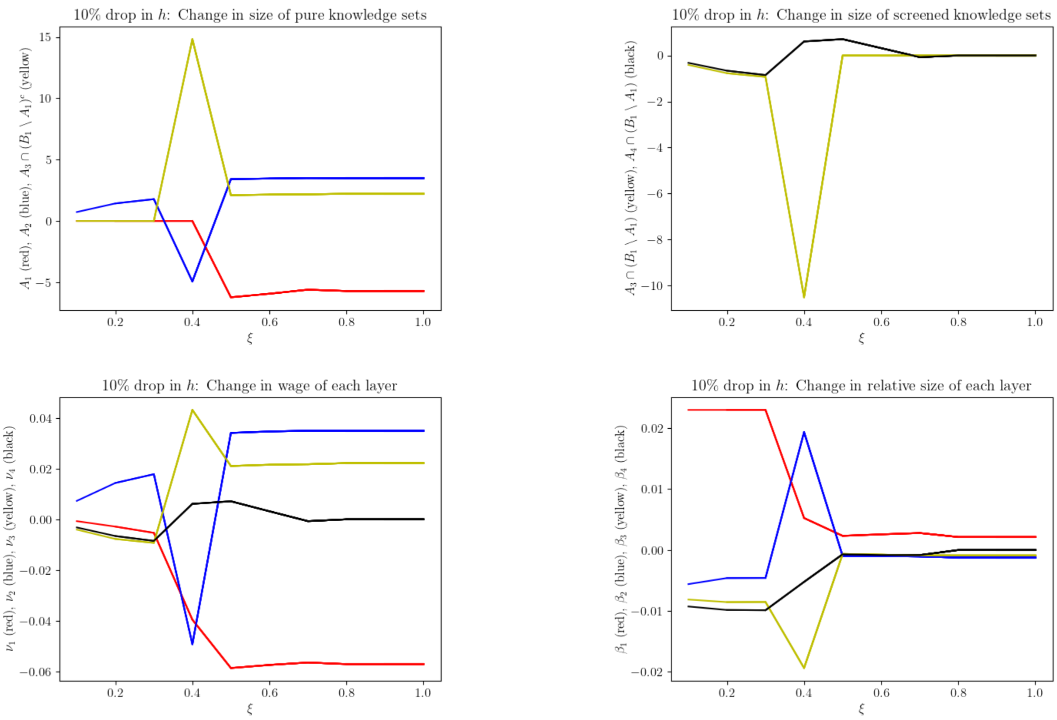

Figure 5: Effects of a 10% drop in h when $(c, h, b) = (1, 0.5, 1)$.

cheaper in relation to screened knowledge).

As Caliendo, Monte, and Rossi-Hansberg (2015) and Caliendo, Mion, Opromolla, and Rossi-Hansberg (2020) have shown, the effect of shocks on the activity of hierarchical firms depends on whether or not they change the numbers of their layers. More generally, shocks can lead firms to adjust their organizational form in ways that differ from changes to the number of layers. As we now illustrate, the effects of a fall in h also depend on whether the firm changes its organizational structure as a response to the shock. Hence, the estimation of the impact of shocks on firms' activity should condition on this too.

Reorganization in response to a fall in h occurs at $\xi = 0.4$ where the optimal organization changes from a fully screening organization (when $h = 0.5$) to a hybrid form (when $h = 0.45$). The fall in h results in the least common problems that were previously solved by layer 2 being passed on to layer 3. The order is $A_4 \cap (B_1 \setminus A_1) < A_3 \cap (B_1 \setminus A_1) < A_2 < A_3 \cap (B_1 \setminus A_1)^c$, with $A_3 \cap (B_1 \setminus A_1)^c$ becoming nonempty when $h = 0.45$. Thus, layer 3 now solves some relatively common problems, which are screened by layer 1, and also the least common ones, which are passed to it via layer 2. Thus, improvements in communication technology can result in more vertical organizations as the knowledge of higher managers can substitute for screening – note that the wage of layer 2, which is the highest in the organization, can go down in this case as more problems are instead passed on to layer 3.

We now contrast the predictions of a fall in h with those of a fall in ξ. Both improvements lead to a fall in the wage of the workers, but they have different implications regarding the wages of the managers. When no reorganization takes place, a fall in ξ increases the sizes of the screened knowledge sets (and hence wages of the corresponding layers) at the expense of the pure knowledge sets, whereas a fall in h increases the sizes of the pure knowledge sets at the expense of the screened ones. In general, a fall in ξ causes the organization to become more horizontal, whereas a fall in h causes the organization to become more vertical.[33]

Our results also provide novel testable implications for the theory of knowledge-based organizations by focusing on special cases and, thus, on specific sub-samples of firms corresponding to each of these cases. For instance, Theorem 5 predicts that workers of firms with small ξ have considerable general knowledge; moreover, Theorem 6 predicts that workers of firms with small h have little specialized knowledge.

6 Cumulative Knowledge

In this section we consider the case where knowledge is cumulative. The cumulative knowledge case is important because it often arises when knowledge is obtained through learning by doing; to borrow Garicano's (2000) example, "a *chef de cuisine* has usually previously been employed in all lower-rank jobs in the kitchen."

Assuming that knowledge is cumulative means that, in order to learn how to solve (respectively, screen) a "hard" problem, an individual needs to learn how to solve (resp. screen) all the problems that are "easier" than it. The notions of hard and easy problems are specific to the organization as it is up to the organization to decide which problems everyone in it knows how to solve (resp. screen) — these are then the easy problems —, which ones only one layer knows how to solve (resp. screen) — these are then the hard problems — and all the problems in between. Formally, this amounts to adding a linear order $\prec$ on L to the description of an organization with the following properties:

[33] Several papers, e.g. Akerman, Gaarder, and Mogstad (2015) and Bastos, Monteiro, and Straume (2018), have studied the effects of improvements in technology interpreted as a fall in h. However, such an improvement may also be interpreted as a fall in ξ. By clarifying the organizational responses to these changes, our model can potentially shed light on the relative importance of these channels.

$\prec$ is a linear order on L such that $A_i \subseteq A_j$ and $B_i \subseteq B_j$ whenever $i \prec j$. We can use $\prec$ to order L and write $L = \{i_1, \ldots, i_L\}$, where $i_1 \prec \cdots \prec i_L$, $A_{i_1} \subseteq \cdots \subseteq A_{i_L}$ and $B_{i_1} \subseteq \cdots \subseteq B_{i_L}$. We refer to $\prec$ as the *cumulative knowledge order.* In summary, an *organization with cumulative knowledge* is $O = (L, \prec, (\beta_i, A_i, B_i, l_i, \prec_i, t_i^p, t_i^h)_{i \in L})$ such that $(L, (\beta_i, A_i, B_i, l_i, \prec_i, t_i^p, t_i^h)_{i \in L})$ is an organization and $\prec$ is a cumulative knowledge order.

Theorem 1 on specialization extends to the case of cumulative knowledge with exactly the same statement.[34] As in Section 4, we may assume that layer 1 is such that $t_1^p = 1$ and that $2 \prec_1 \cdots \prec_1 L$; $\prec_1$ is then just the standard "less than" $<$ order. Also as in Section 4, we write $\alpha_1 = 1$ and $\alpha_l = \alpha_{l1}$ for each $l \in L \setminus \{1\}$.

In the non-cumulative knowledge case, the knowledge sets of different layers are disjoint and this, together with related results for the screening sets, allowed us to obtain a partition of the union of screening sets which we use to derive our results in Section 4. In contrast, knowledge sets are not disjoint in the cumulative knowledge case by definition, which implies that a new partition of the union of the screening sets is needed to derive the results of this section. The partition of the union of the screening sets of an optimal organization with cumulative knowledge we consider is

$$\begin{aligned}\mathcal{C} \; &= \{\cap_{j<l}(B_{i_j} \setminus A_{i_j})^c \cap (A_{i_l} \setminus A_{i_{l-1}}) : l \in L\} \\ &\cup \{\cap_{j<l}(B_{i_j} \setminus A_{i_j})^c \cap (B_{i_l} \setminus A_{i_l}) \cap (A_{i_k} \setminus A_{i_{k-1}}) : l, k \in L \text{ and } k > l\} \\ &\cup \{\cap_{j<l}(B_{i_j} \setminus A_{i_j})^c \cap (B_{i_l} \setminus A_{i_l}) \cap (B_{i_L} \setminus A_{i_L}) : l \in L\}\end{aligned}$$

with the usual convention that the intersection of an empty family of subsets of Ω is Ω itself and with $A_{i_0} = \emptyset$.

The above partition $\mathcal{C}$ of $\cup_{l \in L} B_l$ allows us, in the cumulative knowledge case, to show that there are no gaps and to order the elements of $\mathcal{C}$ according to their learning costs. Moreover, η-optimal organizations exist.

We characterize the cumulative knowledge order in optimal organizations and show that it equals the precedence relation of the layer of workers which, given our normalization, equals the "less than" order, i.e. $i \prec j$ if and only if $i < j$. This means that workers are those who know the least, followed by those in layer 2 who are the first to solve or screen problems that the workers cannot deal with, and so on until the top layer L which consists of those who know the most in the organization.

[34] Detailed statements and proofs for the results of this section are in the supplementary material.

More detailed characterizations of optimal organizations are possible in extreme cases. When screening is as costly as solving problems, we obtain the same conclusions as in the cumulative knowledge case of Garicano (2000). In contrast, Theorem 8 below characterizes η-optimal organizations with cumulative knowledge when the cost ξ of learning how to screen problems is small.

Theorem 8 *If (A1) and (A2) hold, then there exists* $\bar{\xi} > 0$ *such that, if* $0 < \xi < \bar{\xi}$ *and* O *is an* η*-optimal organization with cumulative knowledge with* $L \geq 2$*, then:*

1. $B_1 \setminus A_1 = A_2^c$,

2. $B_l = \Omega$ *for each* $l > 1$,

3. $A_1 < A_2 \setminus A_1 < A_l \setminus A_{l-1} < A_L^c$ *for each* $l > 2$,

4. $\beta_2 > \cdots > \beta_L$, *and*

5. $\beta_1 > \sum_{l=2}^{L} \beta_l$ *if* $h < 1$.

When ξ is sufficiently small, the optimal organization in the cumulative knowledge case is analogous to the non-cumulative case. The main difference is that, without cumulative knowledge, layer 2 knows how to solve the least frequent problems whereas, with cumulative knowledge, it solves the most frequent problems among the managers. This happens because, with cumulative knowledge, the knowledge set of each layer $l > 2$ must contain that of layer 2. This implies that the learning costs are lower for layer 2 than for any of the following layers. Hence, since $y = \frac{F(A_L) - \sum_{l \in L} \alpha_l \nu_l}{\sum_{l \in L} \alpha_l}$, it is optimal to increase the helping costs α_2 of layer 2 by the same amount that, say, α_3 is reduced, which can be obtained by moving $A_2 \setminus A_1$ to be immediately after A_1.

7 Concluding remarks

In this paper we propose a model of an organization that optimally combines the time and knowledge of its members to solve problems that arise in production. Our innovation is to allow members of the organization to screen problems before attempting to solve them. Screening has the potential to improve the organization of knowledge since problems can be sent directly to those who can solve them, saving the time of managers and hence allowing the organization to have more workers and produce more output.

It might be that the organization finds it optimal not to screen problems. This happens when the learning and time costs of screening are equal to those of solving problems and, thus, in this case the optimal organization is a hierarchy exactly as in Garicano (2000). However, novel organizational forms – where some problems are screened and others passed through the organization in a hierarchical way – emerge as optimal when learning how to screen problems is cheaper than learning how to solve them. When screening is very cheap, every problem faced by the organization is sent directly to the person who can solve it.

Our work delivers a broader theory of the organization of specialization, whereby the optimal degree of specialization resolves trade-offs between the cost of learning how to screen problems, the cost of learning how to solve problems and the time spent attempting to solve or screen problems and then communicating the answer. As some of these costs become more or less important, the optimal organization changes in several dimensions, namely on its degree of specialization in tasks, knowledge, screening and scope.

The richness of organizational forms we obtain makes it easier to rationalize real-world organizations and understand their changes in response to shocks. However, a detailed empirical test of the theory developed here is not without challenges as it requires a systematic identification of the organizational form of each firm considered. We plan to address this question in future research.

Another important question concerns the impact of screening in a market setting e.g. on the distribution of wages and firm sizes. In Carmona and Laohakunakorn (2023), we extended Greinecker and Kah's (2021) matching model to include many-to-one matching and occupational choice; in a future paper, this will allow us to embed the framework of this paper in a market setting in an analogous way to what Garicano and Rossi-Hansberg (2004) and Garicano and Rossi-Hansberg (2006) did for the framework of Garicano (2000).

References

Akerman, A., I. Gaarder, and M. Mogstad (2015): "The Skill Complementarity of Broadband Internet," *Quarterly Journal of Economics*, 130, 1781–1824.

Antràs, P., L. Garicano, and E. Rossi-Hansberg (2006): "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 121, 31–77.

Bastos, P., N. P. Monteiro, and O. R. Straume (2018): "Foreign Acquisition and Internal Organization," *Journal of International Economics*, 114, 143–163.

Becker, G., and K. Murphy (1992): "The Division of Labor, Coordination Costs, and Knowledge," *Quarterly Journal of Economics*, 107, 1137–1160.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does Management Matters? Evidence from India," *Quarterly Journal of Economics*, 128, 1–51.

Bloom, N., L. Garicano, R. Sadun, and J. Van Reenen (2014): "The Distinct Effect of Information Technology and Communication Technology on Firm Organization," *Management Science*, 60, 2859–2885.

Caicedo, S., R. Lucas, and E. Rossi-Hansberg (2019): "Learning, Career Paths, and the Distribution of Wages," *American Economic Journal: Macroeconomics*, 11, 49–88.

Caliendo, L., G. Mion, L. Opromolla, and E. Rossi-Hansberg (2020): "Productivity and Organization in Portuguese Firms," *Journal of Political Economy*, 128, 4211–4257.

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015): "The Anatomy of French Production Hierarchies," *Journal of Political Economy*, 123, 809–852.

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

Carmona, G., and K. Laohakunakorn (2023): "Stable Matching in Large Markets with Occupational Choice," University of Surrey.

Chen, C., and W. Suen (2019): "The Comparative Statics of Optimal Hierarchies," *American Economic Journal: Microeconomics*, 11, 1–25.

Crémer, J., L. Garicano, and A. Prat (2007): "Language and the Theory of the Firm," *Quarterly Journal of Economics*, 122, 373–407.

Eeckhout, J., and P. Kircher (2018): "Assortative Matching with Large Firms," *Econometrica*, 86, 85–132.

Fuchs, W., L. Garicano, and L. Rayo (2015): "Optimal Contracting and the Organization of Knowledge," *Review of Economic Studies*, 82, 632–658.

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

Garicano, L., and T. Hubbard (2007): "Managerial Leverage is Limited by the Extent of the Market: Hierarchies, Specialization, and the Utilization of Lawyers' Human Capital," *Journal of Law and Economics*, 50, 1–43.

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

——— (2006): "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 121, 1383–1435.

——— (2012): "Organizing Growth," *Journal of Economic Theory*, 147, 623–656.

——— (2015): "Knowledge-Based Hierarchies: Using Organizations to Understand the Economy," *Annual Review of Economics*, 7, 1–30.

Garicano, L., and T. Van Zandt (2013): "Hierarchies and the Division of Labor," in *The Handbook of Organizational Economics*, ed. by R. Gibbons, and J. Roberts. Princeton University Press, Princeton.

Geerolf, F. (2017): "A Theory of Pareto Distributions," UCLA.

Gibbons, R., and J. Roberts (2013): *The Handbook of Organizational Economics*. Princeton University Press, Princeton.

Greinecker, M., and C. Kah (2021): "Pairwise Stable Matching in Large Economies," *Econometrica*, 89, 2929–2974.

Gumpert, A. (2018): "The Organization of Knowledge in Multinational Firms," *Journal of the European Economic Association*, 16, 1929–1976.

Impink, S., A. Prat, and R. Sadun (2022): "Communication within Firms: Evidence from CEO Turnovers," NBER Working Paper Series, No. 29042.

Kikuchi, T., K. Nishimura, and J. Stachurski (2018): "Span of Control, Transaction Costs, and the Structure of Production Chains," *Theoretical Economics*, 13, 729–760.

Luenberger, D., and Y. Ye (2008): *Linear and Nonlinear Programming*. Springer, New York.

Rosen, S. (2002): "Markets and Diversity," *American Economic Review*, 92, 1–15.

Rudin, W. (1976): *Principles of Mathematical Analysis*. McGraw-Hill, New York.

Tian, L. (2021): "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil," CEPR Press Discussion Paper No. 16590.

A Online Appendix: Proofs

This section contains the proofs of the results stated in the main body of the paper and some lemmas that are used in those proofs. We begin by proving Theorems 1–3, which generalizes and corrects some statements in Garicano (2000). Section A.1 discusses these issues and in particular, compares our approach with his.

A.1 Technical difficulties in Garicano (2000)

This paper corrects some statements in Garicano (2000) and provides detailed proofs for them. In this section, we briefly describe what is missing in those statements and proofs in Garicano (2000), and indicate how we have corrected them.

The difficulties concern mainly Proposition 1 in Garicano (2000), from which the other propositions in that paper follow. One issue is that $t_i^h + t_i^p = 1$ for each $i \in L$ is simply assumed in its proof, but not demonstrated. The standard argument of increasing t_i^p whenever $t_i^h + t_i^p < 1$ does not work because, in general, an increase in t_i^p requires an increase in t_j^h in the layers j that help layer i. We show that $t_i^h + t_i^p = 1$ for each $i \in L$ roughly by increasing both t_i^h and t_i^p proportionally, by reducing the size β_i of layer i and by increasing the size of all other layers to increase the output of the organization.

A more difficult issue concerns the system of equations (A3) and the problem (A4) in Garicano (2000). In (A3) there are, as claimed, L unknowns and L equations, but to be able to solve it uniquely, one needs the resulting matrix to be invertible. Nothing guarantees that. Indeed, the matrix in question is

$$\begin{pmatrix} 1 & \alpha_{12} & \cdots & \alpha_{1L} \\ \alpha_{21} & 1 & \cdots & \alpha_{2L} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{L1} & \alpha_{L2} & \cdots & 1 \end{pmatrix};$$

hence, if $L = 2$, the matrix fails to be invertible if $1 - \alpha_{12}\alpha_{21} = 0$; if $L = 3$, the matrix fails to be invertible if $1 - \alpha_{23}\alpha_{32} - \alpha_{12}(\alpha_{21} - \alpha_{23}\alpha_{31}) + \alpha_{13}(\alpha_{21}\alpha_{32} - \alpha_{31}) = 0$. The values of L and α_{ik} are endogenous, hence one cannot simply assume that they are such that the above matrix is invertible. But if the above matrix is not invertible, then problem (A4) is not defined.

But even when problem (A4) is defined, the proposed solution is not clear. The goal in (A4) is to determine β_i for each $i \in L$ given the set L of layers and, for each layer $i \in L$, A_i, l_i and $\prec_i$, and given the solution to (A3) expressed as $\beta_i t_i^p = \rho_i' \beta = \sum_{j \in L} \rho_{i,j} \beta_j$ for each $i \in L$, where $(\rho_{i,j})_{i,j \in L}$ is the solution to (A3). The claim is that there is $i \in L$ such that $\beta_i = \rho_i' \beta$ and $\rho_j' \beta = 0$ for all $j \neq i$ but there is no proof of this claim.

More generally, the claim of Proposition 1 in Garicano (2000) is that, given $(L, (l_i, \prec_i, A_i)_{i \in L})$, output can be increased unless there is $i \in L$ with $t_i^p = 1$ and $t_j^p = 0$ for all $j \in L \setminus \{i\}$. A trivial example shows that this is not the case: Let $L = 2$, $A_1 = A_2 = A$ and $l_i = \{i\}$ for each $i = 1, 2$. Because $2 \notin l_1$ and $1 \notin l_2$, $t_1^h = t_2^h = 0$. Then it is optimal to set $t_1^p = t_2^p = 1$ and, given $\beta_1 + \beta_2 = 1$, output is $y = F(A) - c\mu(A)$. But it is not the case that there is $i \in L$ with $t_i^p = 1$ and $t_j^p = 0$ for all $j \in L \setminus \{i\}$.

We solve the difficulties surrounding Garicano (2000, Proposition 1) as follows. First, we apply the fundamental theorem of linear programming to a linear programming problem, different from the one in Garicano (2000), to obtain a solution to the maximization of output such that there is $i \in L$ with $\beta_i t_i^p = 1$ and $\beta_j t_j^p = 0$ for all $j \in L \setminus \{i\}$. This is clearly possible in the above example by setting $\beta_1 = 1$ and $\beta_2 = 0$ (or $\beta_1 = 0$ and $\beta_2 = 1$).

As in the example, solutions to linear programming problems are often not unique. We then solve this problem by using the optimality condition in Section 3.4 which, when comparing two organizations with the same output, ranks the one with the smallest L above the other. In the context of the example, this implies that the organization with $L = 2$, $A_1 = A_2 = A$ and $l_i = \{i\}$ for each $i = 1, 2$ cannot be optimal since the organization with just one layer (say layer 1) obtains the same output with a smaller number of layers. More generally, the optimality criterion that we use allows us to obtain that there is a layer $i \in L$ such that $t_i^p = 1$ and $t_j^p = 0$ for each $j \neq i$ in any optimal organization.

A final issue concerns the proofs of Proposition 2 and 3 in Garicano (2000) which are correct provided that, as we do, the class of sets in which each A_i lies is $\mathcal{I}$. The reason is that these results are established, both in Garicano (2000) and in this paper, by transferring an interval $[\omega, \omega + \varepsilon)$ for some $\omega \in \Omega$ and $\varepsilon > 0$ from e.g. A_i to A_j and this is only possible when A_i is the (finite) union of disjoint intervals of the form $[a, b)$, i.e. it belongs to $\mathcal{I}$.

A.2 Allocation of labor

We establish that, in an optimal organization, $t_i^p + t_i^h = 1$ for each $i \in L$. This requires the following two lemmas. Lemma A.1 states that, in an optimal organization, the list of each layer that does not devote time to production consists of itself only. This is because such layer requires no help and our notion of optimality favours organizations with smaller lists.

Lemma A.1 *If O is an optimal organization, then $l_i = \{i\}$ for each $i \in L$ with $t_i^p = 0$.*

Proof. Suppose not; then there is $i \in L$ such that $t_i^p = 0$ and $\{i\} \subset l_i$. Consider $\hat{O}$ equal to O except that $\hat{l}_i = \{i\}$.

Since $\hat{t}_i^p = t_i^p = 0$, we have that

$$\hat{\beta}_j \hat{t}_j^h = \beta_j t_j^h = \sum_{k \in L \setminus \{i\}} \beta_k t_k^p \alpha_{jk} = \sum_{k \in L \setminus \{i\}} \hat{\beta}_k \hat{t}_k^p \hat{\alpha}_{jk} = \sum_{k \in L} \hat{\beta}_k \hat{t}_k^p \hat{\alpha}_{jk}$$

for each $j \in L$ and, hence, $\hat{O}$ is an organization. Using again $\hat{t}_i^p = t_i^p = 0$, it follows that $\hat{y} = \sum_{j \in L \setminus \{i\}} \beta_j (t_j^p F(\cup_{l \in l_j} A_l) - c\mu(A_j) - \xi\mu(B_j \setminus A_j)) - \beta_i (c\mu(A_i) + \xi\mu(B_i \setminus A_i)) = y$. This, together with $\hat{L} = L$, $\hat{l}_j = l_j$ for all $j \in L \setminus \{i\}$ and $\hat{l}_i \subset l_i$, contradicts the optimality of O. ■

Lemma A.2 states that members of each layer use strictly positive amount of time; if not, such layers could be removed and its member be moved to other layers, thus obtaining a higher output with a smaller number of layers.

Lemma A.2 *If O is an optimal organization, then $t_i^p + t_i^h > 0$ for each $i \in L$.*

Proof. Suppose not; then let $i \in L$ be such that $t_i^p + t_i^h = 0$. Since $y > 0$, then $\{i\} \neq L$; hence $\lambda := 1/(\sum_{i \in L \setminus \{i\}} \beta_i) > 1$. Consider $\hat{O}$ equal to O except that $\hat{L} = L \setminus \{i\}$ and, for each $j \in \hat{L}$, $\hat{\beta}_j = \lambda \beta_j$, $\hat{l}_j = l_j \setminus \{i\}$ and $\hat{\prec}_j = \prec_j |_{\hat{l}_j}$.

Let $k \in \hat{L}$. Since $t_i^h = 0$, we have that $0 = \beta_i t_i^h = \sum_{l \in L} \beta_l t_l^p \alpha_{il}$. In particular, it follows that $t_k^p = 0$ or $\alpha_{ik} = 0$. If $\alpha_{ik} > 0$, then $i \in l_k$ and $t_k^p = 0$; Lemma A.1 then implies that $l_k = \{k\}$, a contradiction to $i \in l_k$ and $i \neq k$. Thus, $\alpha_{ik} = 0$, implying that either $i \notin l_k$ or $F(A_i \setminus \cup_{l \prec_k i} A_l) = 0 = F(A_i^c \setminus \cup_{l \prec_k i} B_l)$. In the latter case, since both $A_i \setminus \cup_{l \prec_k i} A_l \in \mathcal{I}$ and $A_i^c \setminus \cup_{l \prec_k i} B_l \in \mathcal{I}$, we have that $A_i \subseteq \cup_{l \prec_k i} A_l$ and $A_i^c \subseteq \cup_{l \prec_k i} B_l$ by footnote 9.

We now show that $\hat{\alpha}_{jk} = \alpha_{jk}$ for each $j, k \in L \setminus \{i\}$ such that $j \neq k$ and $j \in l_k$. This is clear if $i \notin l_k$ or $i \prec_k j$ does not hold, hence assume that $i \in l_k$ and $i \prec_k j$. Let

$\mathcal{L}_{jk} = \{l \in l_k : l \prec_k j\}$ and $\hat{\mathcal{L}}_{jk} = \mathcal{L}_{jk} \setminus \{i\}$. The former implies that $\cup_{l\in\mathcal{L}_{jk}} A_l = \cup_{l\in\hat{\mathcal{L}}_{jk}} A_l$ and, thus, $A_j \setminus \cup_{l\in\mathcal{L}_{jk}} A_l = A_j \setminus \cup_{l\in\hat{\mathcal{L}}_{jk}} A_l$. In addition, we have that $\Omega = A_i \cup A_i^c \subseteq \cup_{l\prec_k i} B_l$ and, hence, $A_j^c \setminus \cup_{l\in\hat{\mathcal{L}}_{jk}} B_l = \emptyset = A_j^c \setminus \cup_{l\in\mathcal{L}_{jk}} B_l$. Thus, $\hat{\alpha}_{jk} = \alpha_{jk}$ as claimed.

We have that $\sum_{j\in\hat{L}} \hat{\beta}_j = 1$. Since $\hat{\alpha}_{jk} = \alpha_{jk}$ for each $j, k \in \hat{L}$ such that $j \neq k$ and $j \in l_k$, then, as $\hat{t}_i^p = t_i^p = 0$, $\hat{\beta}_j \hat{t}_j^h = \lambda\beta_j t_j^h = \sum_{k\in L\setminus\{i\}} \lambda\beta_k t_k^p \alpha_{jk} = \sum_{k\in L\setminus\{i\}} \hat{\beta}_k \hat{t}_k^p \hat{\alpha}_{jk}$ for each $j \in \hat{L}$. Hence, $\hat{O}$ is an organization.

Since $A_i \subseteq \cup_{l\prec_k i} A_l$ if $i \in l_k$, it follows that $\cup_{l\in\hat{l}_k} A_l = \cup_{l\in l_k} A_l$ for each $k \in L \setminus \{i\}$. Thus, since $\lambda > 1$ and $y > 0$, $\hat{y} = \lambda \sum_{k\in L\setminus\{i\}} \beta_k(t_k^p F(\cup_{l\in l_k} A_k) - c\mu(A_k) - \xi\mu(B_k \setminus A_k)) > \sum_{k\in L\setminus\{i\}} \beta_k(t_k^p F(\cup_{l\in l_k} A_k) - c\mu(A_k) - \xi\mu(B_k \setminus A_k)) - \beta_i(c\mu(A_i) + \xi\mu(B_i \setminus A_i)) = y$. This shows that $\hat{y} > y > 0$ and, together with $\hat{L} \subset L$, contradicts the optimality of O. ■

We can now state and prove our main result of this section.

Lemma A.3 *If O is an optimal organization, then $t_i^p + t_i^h = 1$ for each $i \in L$.*

Proof. Suppose not; then let $i \in L$ be such that $t_i^p + t_i^h < 1$. We have that $y > 0$ since O is optimal and $t_i^p + t_i^h > 0$ by Lemma A.2. Set $\lambda = \frac{1}{t_i^p + t_i^h} > 1$ and $\gamma = \frac{1}{1-\beta_i+\frac{\beta_i}{\lambda}} > 1$. Consider $\hat{O}$ equal to O except that $\hat{\beta}_i = \frac{\gamma}{\lambda}\beta_i$, $\hat{t}_i^p = \lambda t_i^p$, $\hat{t}_i^h = \lambda t_i^h$ and, for each $j \neq i$, $\hat{\beta}_j = \gamma\beta_j$.

We have that $\hat{\beta}_j \hat{t}_j^p = \gamma\beta_j t_j^p$ and $\hat{\beta}_j \hat{t}_j^h = \gamma\beta_j t_j^h$ for all $j \in L$. Since $\hat{\alpha}_{jk} = \alpha_{jk}$ for each $k, j \in L$ (as $\hat{L} = L$, $\hat{A}_j = A_j$, $\hat{B}_j = B_j$, $\hat{l}_j = l_j$ and $\hat{\prec}_j = \prec_j$ for all $j \in L$), it follows that $\hat{\beta}_j \hat{t}_j^h = \sum_{k\in L} \hat{\alpha}_{jk} \hat{\beta}_k \hat{t}_k^p$ for each $j \in L$. Moreover, $\sum_{j\in L} \hat{\beta}_j = \gamma \sum_{j\neq i} \beta_j + \frac{\gamma}{\lambda}\beta_i = \gamma\left(1 - \beta_i + \frac{\beta_i}{\lambda}\right) = 1$. Thus, $\hat{O}$ satisfies all requirements of an organization.

We have that $-\frac{\gamma}{\lambda}\beta_i\nu_i \geq -\gamma\beta_i\nu_i$ since $\lambda > 1$. Since $y > 0$ and $\gamma > 1$, it follows that $\hat{y} = \gamma \sum_{j\in L} \beta_j t_j^p F(\cup_{l\in l_j} A_l) - \gamma \sum_{j\neq i} \beta_j\nu_j - \frac{\gamma}{\lambda}\beta_i\nu_i \geq \gamma \sum_{j\in L} \beta_j t_j^p F(\cup_{l\in l_j} A_l) - \gamma \sum_{j\in L} \beta_j\nu_j = \gamma y > y$. This shows that $\hat{y} > y > 0$ and, together with $L = \hat{L}$, contradicts the optimality of O. ■

A.3 Specialization

In this section we establish Theorem 1. For convenience, for each $i \in L$, let

$$\nu_i = c\mu(A_i) + \xi\mu(B_i \setminus A_i).$$

The problem that $\beta = (\beta_1, \ldots, \beta_L)$, $t^p = (t_1^p, \ldots, t_L^p)$ and $t^h = (t_1^h, \ldots, t_L^h)$ solve is:

$$\max_{\beta, t^p, t^h} \sum_{i=1}^{L} \beta_i \big(t_i^p F(\cup_{l \in l_i} A_l) - \nu_i \big) \tag{6}$$

$$\text{subject to } \sum_{i=1}^{L} \beta_i = 1, \text{ and, for each } i \in L, \tag{7}$$

$$\beta_i > 0, \tag{8}$$

$$t_i^p \geq 0, \tag{9}$$

$$t_i^h \geq 0, \tag{10}$$

$$t_i^p + t_i^h = 1, \tag{11}$$

$$\beta_i t_i^h = \sum_{j=1}^{L} \alpha_{ij} \beta_j t_j^p. \tag{12}$$

Note that (11) uses the conclusion of Lemma A.3.

We consider an equivalent but easier problem defined as follows. Instead of having everyone in a layer with the same allocation of labor, we now specialize people as follows. For each $i \in L$, let δ_i be the fraction of people that are in layer i and are a *worker* (i.e. spend all their time working) and let $\delta = (\delta_1, \ldots, \delta_L)$. There is a need of $\sum_{j=1}^{L} \delta_j \alpha_{ij}$ of helpers or *managers* (i.e. people who spend all their time helping other layers) in layer i and, hence, a fraction of $\delta_i + \sum_{j=1}^{L} \delta_j \alpha_{ij}$ people in layer i. Putting $\beta_i = \delta_i + \sum_{j=1}^{L} \delta_j \alpha_{ij}$, $\beta_i t_i^p = \delta_i$ and $\beta_i t_i^h = \sum_{j=1}^{L} \delta_j \alpha_{ij}$ in (6)–(12), we obtain the following problem:

$$\max_{\delta} \sum_{i=1}^{L} \big(F(\cup_{l \in l_i} A_l) - \nu_i \big) \delta_i - \sum_{i=1}^{L} \sum_{j=1}^{L} \nu_i \alpha_{ij} \delta_j \tag{13}$$

$$\text{subject to } \sum_{i=1}^{L} \Big(\delta_i + \sum_{j=1}^{L} \delta_j \alpha_{ij} \Big) = 1, \text{ and, for each } i \in L, \tag{14}$$

$$\delta_i + \sum_{j=1}^{L} \delta_j \alpha_{ij} > 0, \tag{15}$$

$$\delta_i \geq 0. \tag{16}$$

Lemma A.4 states the equivalence between the optimization problems (6)–(12) and (13)–(16); its proof is a simple algebraic argument.

Lemma A.4 *If* (β, t^p, t^h) *solves problem* (6)–(12) *and* $\delta_i = \beta_i t_i^p$ *for each* $i \in L$*, then* δ *solves problem* (13)–(16)*. Conversely, if* δ *solves problem* (13)–(16) *and, for each* $i \in L$*,* $\beta_i = \delta_i + \sum_{j=1}^{L} \delta_j \alpha_{ij}$*,* $t_i^p = \frac{\delta_i}{\delta_i + \sum_{j=1}^{L} \delta_j \alpha_{ij}}$ *and* $t_i^h = \frac{\sum_{j=1}^{L} \delta_j \alpha_{ij}}{\delta_i + \sum_{j=1}^{L} \delta_j \alpha_{ij}}$*, then* (β, t^p, t^h) *solves problem* (6)–(12)*.*

Note that we may write (13) as $\sum_{i=1}^{L} \theta_i \delta_i$, where, for each $i \in L$,

$$\theta_i = F(\cup_{l \in l_i} A_l) - \left(\nu_i + \sum_{j=1}^{L} \nu_j \alpha_{ji} \right).^{35} \tag{17}$$

The coefficient θ_i reveals the impact of increasing the fraction of workers of layer i in the objective function to be an increase in the production of layer i ($F(\cup_{l \in l_i} A_l)$), an increase in the learning costs of layer i (ν_i) and an increase in the learning costs of the layers that help layer i ($\sum_{j=1}^{L} \nu_j \alpha_{ji}$).

Similarly, we may write (14) as $\sum_{i=1}^{L} \gamma_i \delta_i = 1$, where, for each $i \in L$,

$$\gamma_i = 1 + \sum_{j=1}^{L} \alpha_{ji}.^{36} \tag{18}$$

The coefficient γ_i describes the impact of increasing the fraction of workers of layer i in the constraint as being that increase (represented by 1) plus the increase in the fraction of helpers in the layers that help layer i ($\sum_{j=1}^{L} \alpha_{ji}$).

Furthermore, as Lemma A.5 shows, we may drop (15) to characterize the solutions of (13)–(16) and, thus, focus on the following linear programming problem:

$$\max_{\delta} \sum_{i=1}^{L} \theta_i \delta_i \tag{19}$$

$$\text{subject to } \sum_{i=1}^{L} \gamma_i \delta_i = 1, \text{ and} \tag{20}$$

$$\delta_i \geq 0 \text{ for each } i \in L. \tag{21}$$

Lemma A.5 *If* (β, t^p, t^h) *solves problem* (6)–(12)*, then* $\delta = (\beta_1 t_1^p, \cdots, \beta_L t_L^p)$ *solves problem* (19)–(21)*.*

Proof. Let (β, t^p, t^h) be a solution to (6)–(12) and $\delta = (\beta_1 t_1^p, \cdots, \beta_L t_L^p)$. By Lemma A.4, δ solves (13)–(16). We now claim that δ solves (19)–(21). To see this, let δ^* be

[35] Indeed,

$$\sum_{i=1}^{L} \left(F(\cup_{l \in l_i} A_l) - \nu_i \right) \delta_i - \sum_{i=1}^{L} \sum_{j=1}^{L} \nu_i \alpha_{ij} \delta_j = \sum_{j=1}^{L} \left(F(\cup_{l \in l_j} A_l) - \nu_j \right) \delta_j - \sum_{j=1}^{L} \delta_j \sum_{i=1}^{L} \nu_i \alpha_{ij}$$

$$= \sum_{j=1}^{L} \left[F(\cup_{l \in l_j} A_l) - \left(\nu_j + \sum_{i=1}^{L} \nu_i \alpha_{ij} \right) \right] \delta_j = \sum_{i=1}^{L} \left[F(\cup_{l \in l_i} A_l) - \left(\nu_i + \sum_{j=1}^{L} \nu_j \alpha_{ji} \right) \right] \delta_i.$$

[36] Indeed, $\sum_{i=1}^{L} \left(\delta_i + \sum_{j=1}^{L} \alpha_{ij} \delta_j \right) = \sum_{j=1}^{L} \delta_j + \sum_{j=1}^{L} \delta_j \sum_{i=1}^{L} \alpha_{ij} = \sum_{j=1}^{L} \left(1 + \sum_{i=1}^{L} \alpha_{ij} \right) \delta_j = \sum_{i=1}^{L} \left(1 + \sum_{j=1}^{L} \alpha_{ji} \right) \delta_i$.

a solution to (19)–(21). Since δ satisfies (20)–(21), it suffices to show that $\sum_{i=1}^{L}\theta_i\delta_i \geq \sum_{i=1}^{L}\theta_i\delta_i^*$. Let $\{\lambda_n\}_{n=1}^{\infty}$ be a sequence such that $\lambda_n \to 1$ and $\lambda_n \in (0,1)$ for each $n \in \mathbb{N}$, and define $\delta_n = \lambda_n\delta^* + (1-\lambda_n)\delta$ for each n. Then, for each $n \in \mathbb{N}$, δ_n satisfies (14)–(16) and, thus, $\sum_{i=1}^{L}\theta_i\delta_i \geq \sum_{i=1}^{L}\theta_i\delta_{n,i}$. Since $\delta_n \to \delta^*$, it follows that $\sum_{i=1}^{L}\theta_i\delta_i \geq \sum_{i=1}^{L}\theta_i\delta_i^*$ as claimed. ■

Problem (19)–(21) has a linear objective function and a linear constraint besides non-negativity constraints. Lemma A.6 states that it has a corner solution.

Lemma A.6 *There exists a solution $\hat{\delta}$ of the problem* (19)–(21) *such that there is $i \in L$ with $\hat{\delta}_i > 0$ and $\hat{\delta}_j = 0$ for all $j \neq i$. Furthermore, $\hat{\delta}_i = \frac{1}{\gamma_i}$, $\sum_{l\in L}\theta_l\hat{\delta}_l = \frac{\theta_i}{\gamma_i}$ and, for each $l \neq i$, $\hat{\delta}_l + \sum_{j=1}^{L}\hat{\delta}_j\alpha_{lj} > 0$ holds if and only if $\alpha_{li} > 0$.*

Proof. By the fundamental theorem of linear programming (e.g. Luenberger and Ye (2008, p. 20-21)), there exists an optimal basic feasible solution $\hat{\delta}$ of the problem (19)–(21). Since there is only one constraint other than the non-negativity constraints, there exists $i \in L$ such that $\hat{\delta}_j = 0$ for all $j \neq i$. The constraint $\sum_{l=1}^{L}\gamma_l\hat{\delta}_l = 1$ then implies that $\hat{\delta}_i = \frac{1}{\gamma_i}$. We also have that $\sum_{l\in L}\theta_l\hat{\delta}_l = \theta_i\hat{\delta}_i = \frac{\theta_i}{\gamma_i}$. In addition, we have that $\hat{\delta}_l + \sum_{j=1}^{L}\hat{\delta}_j\alpha_{lj} \geq 0$ for each $l \in L$ (recall that $\alpha_{lj} \geq 0$ for all $l, j \in L$). Moreover $\hat{\delta}_l + \sum_{j=1}^{L}\hat{\delta}_j\alpha_{lj} > 0$ holds if and only if $\alpha_{li} > 0$ for each $l \neq i$. ■

Given an organization O, let M be the set of $i \in L$ such that $\delta_i = \frac{1}{\gamma_i}$ and $\delta_j = 0$ for all $j \in L \setminus \{i\}$ is a solution to (19)–(21). It follows by Lemma A.6 that $M \neq \emptyset$.

Corollary A.1 *If O is an optimal organization and $i \in M$, then $y = \frac{\theta_i}{\gamma_i}$.*

Proof. Indeed, y is (6) at (β, t^p, t^h) and (β, t^p, t^h) satisfies (7)–(12) by Lemma A.3. Thus, by Lemmas A.4, A.5 and A.6 respectively, y is also (13) at $\delta = (\beta_1 t_1^p, \ldots, \beta_L t_L^p)$, (19) at δ and (19) at $\hat{\delta}$, the latter being given in Lemma A.6. ■

Lemmas A.7 and A.8 explore the above conclusion that there is a corner solution i.e. $i \in L$ such that $\delta_i > 0$ while $\delta_j = 0$ for each $j \in L \setminus \{i\}$. Lemma A.7 then shows that every layer is in the list of layer i and that the expected amount of time that each of the other layers spend helping layer i is strictly positive. The reason is that, otherwise, the organization could use such corner solution and remove some layer $j \neq i$ to obtain the same output with less layers.

Lemma A.7 *If O is an optimal organization and $i \in M$, then $l_i = L$ and $\alpha_{ji} > 0$ for all $j \in L \setminus \{i\}$.*

Proof. Suppose not. Let $i \in M$ and $\hat{L} = \{j \in l_i : \alpha_{ji} > 0\} \cup \{i\}$; then $\emptyset \subset \hat{L} \subset L$. Define $\hat{O}$ as follows: Layers are $\hat{L}$; for each $j \in \hat{L}$, set: $\hat{A}_j = A_j$ and $\hat{B}_j = B_j$; for each $j \in \hat{L} \setminus \{i\}$, set $\hat{l}_j = \{j\}$, $\hat{\beta}_j = \frac{\alpha_{ji}}{\gamma_i}$, $\hat{t}^p_j = 0$, and $\hat{t}^h_j = 1$; finally, set $\hat{l}_i = \hat{L}$, $\hat{\prec}_i = \prec_i |_{\hat{l}_i}$, $\hat{\beta}_i = \frac{1}{\gamma_i}$, $\hat{t}^p_i = 1$ and $\hat{t}^h_i = 0$.

Arguing as in Lemma A.2, we obtain that $A_j \setminus \cup_{l \prec_i j} A_l = A_j \setminus \cup_{l \in \hat{l}_i : l \prec_i j} A_l$, $A^c_j \setminus \cup_{l \in \hat{l}_i : l \prec_i j} B_l = \emptyset = A^c_j \setminus \cup_{l \prec_i j} B_l$ and $\cup_{l \in \hat{l}_i} A_l = \cup_{l \in l_i} A_l$ for each $j \in \hat{L} \setminus \{i\}$. Hence, $\hat{\alpha}_{ji} = \alpha_{ji}$ for each $j \in \hat{L} \setminus \{i\}$. Thus, as $\alpha_{ji} = 0$ for each $j \in L \setminus \hat{L}$, $\theta_i = F(\cup_{l \in l_i} A_l) - \nu_i - \sum_{j \in L} \alpha_{ji} \nu_j = F(\cup_{l \in \hat{l}_i} A_l) - \nu_i - \sum_{j \in \hat{L}} \hat{\alpha}_{ji} \nu_j = \hat{\theta}_i$. Furthermore, $\gamma_i = 1 + \sum_{j \in L} \alpha_{ji} = 1 + \sum_{j \in \hat{L}} \hat{\alpha}_{ji} = \hat{\gamma}_i$. It then follows that $\hat{O}$ is an organization. Moreover, $\hat{y} = \frac{\theta_i}{\gamma_i} = y > 0$ (the last equality by Corollary A.1) which, together with $\hat{L} \subset L$, contradicts the optimality of O. ■

Lemma A.8 complements Lemma A.7 by showing that the list of every layer other than layer i contains only itself. Indeed, otherwise, the organization could use the corner solution with $\delta_i > 0$ and $\delta_j = 0$ for each $j \neq i$ and shorten the list of some layer $j \neq i$.

Lemma A.8 *If O is an optimal organization and $i \in M$, then $l_j = \{j\}$ for all $j \in L \setminus \{i\}$.*

Proof. Suppose not. Let $i \in M$ and $k \neq i$ such that $\{k\} \subset l_k$. By Lemma A.7 and Corollary A.1, $l_i = L$, $\alpha_{ji} > 0$ for all $j \in L \setminus \{i\}$ and $y = \frac{\theta_i}{\gamma_i}$. Consider $\hat{O}$ equal to O except (possibly) that $\hat{l}_j = \{j\}$, $\hat{\beta}_j = \frac{\alpha_{ji}}{\gamma_i}$, $\hat{t}^p_j = 0$ and $\hat{t}^h_j = 1$ for each $j \neq i$, and $\hat{\beta}_i = \frac{1}{\gamma_i}$, $\hat{t}^p_i = 1$ and $\hat{t}^h_i = 0$.

Since $\hat{l}_i = l_i$, $\hat{\prec}_i = \prec_i$, $\hat{A}_j = A_j$ and $\hat{B}_j = B_j$ for each $j \in L$, it follows that $\hat{\alpha}_{ij} = \alpha_{ij}$ for all $j \in L$ and, hence, $\hat{\gamma}_i = \gamma_i$ and $\hat{\theta}_i = \theta_i$. This, together with $\alpha_{ji} > 0$ for all $j \neq i$, implies that $\hat{O}$ is an organization and that $\hat{y} = \frac{\theta_i}{\gamma_i}$. Finally, $\hat{y} = \frac{\theta_i}{\gamma_i} = y$, $\hat{L} = L$ and $\hat{l}_j \subseteq l_j$ for all $j \in L$ and $\hat{l}_k \subset l_k$ contradicts the optimality of O. ■

We turn now to the proof of Theorem 1.

Proof of Theorem 1. Let O be an optimal organization and $i \in M$. Then $l_i = L$, $\alpha_{ji} > 0$ and $l_j = \{j\}$ for each $j \in L \setminus \{i\}$ by Lemmas A.7 and A.8. It remains to show that $(\beta, t^p, t^h) = (\hat{\beta}, \hat{t}^p, \hat{t}^h)$ where, for each $j \neq i$, $\hat{\beta}_j = \frac{\alpha_{ji}}{\gamma_i}$, $\hat{t}^p_j = 0$ and $\hat{t}^h_j = 1$, and $\hat{\beta}_i = \frac{1}{\gamma_i}$, $\hat{t}^p_i = 1$ and $\hat{t}^h_i = 0$.

Suppose not; then $(\beta, t^p, t^h) \neq (\hat{\beta}, \hat{t}^p, \hat{t}^h)$. Letting $\delta = (\beta_1 t^p_1, \ldots, \beta_L t^p_L)$ and $\hat{\delta} = (\hat{\beta}_1 \hat{t}^p_1, \ldots, \hat{\beta}_L \hat{t}^p_L)$, it follows that $\delta \neq \hat{\delta}$. Thus, there exists $j \neq i$ such that $\delta_j > 0$; in particular, it follows that $|L| \geq 2$. Since the convex combination of two solutions of a linear programming problem is also a solution, there exists a solution $\tilde{\delta}$ to (19)–(21)

such that $\tilde{\delta}_i > 0$ and $\tilde{\delta}_j > 0$. Put $z_i = 1$, $z_j = -\frac{\gamma_i}{\gamma_j}$ and $z_l = 0$ for all $l \notin \{i, j\}$; then, for all ε in a neighborhood of zero (in $\mathbb{R}$), $\tilde{\delta} - \varepsilon z$ satisfies (20)–(21). Optimality of $\tilde{\delta}$ then implies that $\frac{\theta_i}{\gamma_i} = \frac{\theta_j}{\gamma_j}$. We have that $y = \theta_i/\gamma_i$ by Lemmas A.5 and A.6. Moreover, $l_j = \{j\}$ implies that $\cup_{l \in l_j} A_l = A_j$ and that $\alpha_{lj} = 0$ for each $l \in L$; hence, $\theta_j = F(\cup_{l \in l_j} A_l) - (\nu_j + \sum_{l=1}^{L} \nu_l \alpha_{lj}) = F(A_j) - \nu_j$ and $\gamma_j = 1 + \sum_{l=1}^{L} \nu_l \alpha_{lj} = 1$. Thus, $y = \frac{\theta_i}{\gamma_i} = \frac{\theta_j}{\gamma_j} = F(A_j) - \nu_j$. Therefore, the organization $\hat{O}$ with just layer j, i.e. $\hat{L} = \{j\}$, and $\hat{B}_j = B_j$, $\hat{A}_j = A_j$, $\hat{\beta}_j = 1$, $\hat{t}_j^p = 1$ and $\hat{t}_j^h = 0$ and obtains as much output as O. Since $|L| > 1$, this contradicts the optimality of O. ■

A.4 No overlap

We prove Lemma 1 in this section.

Lemma A.9 *If O is an optimal organization, then $A_l \cap A_k = \emptyset$ and $(B_l \setminus A_l) \cap B_k = \emptyset$ for each $k, l \in L$ such that $k < l$.*

Proof. Suppose not; then there exists $k, l \in L$ with $k < l$ such that $A_l \cap A_k \neq \emptyset$ or $(B_l \setminus A_l) \cap B_k \neq \emptyset$. Define an organization $\hat{O}$ to be equal to O except that $\hat{A}_l = A_l \setminus A_k$ and $\hat{B}_l = (B_l \setminus B_k) \cup \hat{A}_l$. We have that $\hat{A}_j \setminus \cup_{i<j} \hat{A}_i = A_j \setminus \cup_{i<j} A_i$ and $\hat{A}_j^c \setminus \cup_{i<j} \hat{B}_i = A_j^c \setminus \cup_{i<j} B_i$ for each $j \in L$. Indeed, this is clear for all $j < l$. When $j = l$, we have that $\hat{A}_l \setminus \cup_{i<l} \hat{A}_i = (A_l \cap A_k^c) \cap (\cap_{i<l} A_i^c) = A_l \setminus \cup_{i<l} A_i$ and $\hat{A}_l^c \setminus \cup_{i<l} \hat{B}_i = (A_l^c \cup A_k) \cap (\cap_{i<l} B_i^c) = A_l^c \cap (\cap_{i<l} B_i^c) = A_l^c \setminus \cup_{i<l} B_i$ since $A_k \cap (\cap_{i<l} B_i^c) \subseteq A_k \cap B_k^c = \emptyset$. Finally, if $j > l$, $\cup_{i<j} \hat{A}_i = \cup_{i<j} A_i$ and $\cup_{i<j} \hat{B}_i = \cup_{i<j} B_i$ and the result follows. We then have that $\hat{\alpha}_j = \alpha_j$ for each $j \in L$. Hence, $\hat{O}$ is an organization.

As $A_k \subseteq B_k$, we have that $\hat{B}_l \setminus \hat{A}_l = (B_l \setminus B_k) \setminus (A_l \setminus A_k) = (B_l \cap B_k^c \cap A_l^c) \cup (B_l \cap B_k^c \cap A_k) = B_l \cap B_k^c \cap A_l^c = (B_l \setminus A_l) \cap B_k^c$. Moreover, $(\hat{B}_l \setminus \hat{A}_l) \cap B_k = (B_l \setminus A_l) \cap B_k^c \cap B_k = \emptyset$, $\hat{A}_l \cap A_k = A_l \cap A_k^c \cap A_k = \emptyset$ and

$$
\begin{aligned}
\hat{y} - y &= \beta_l \left[c(\mu(A_l) - \mu(A_l \setminus A_k)) + \xi(\mu(B_l \setminus A_l) - \mu((B_l \setminus A_l) \cap B_k^c)) \right] \\
&= \beta_l \left[c\mu(A_k \cap A_l) + \xi\mu((B_l \setminus A_l) \cap B_k) \right].
\end{aligned}
$$

Since A_k, A_l, $B_l \setminus A_l$ and B_k belong to $\mathcal{I}$, it follows from $A_k \cap A_l \neq \emptyset$ or $(B_l \setminus A_l) \cap B_k \neq \emptyset$ that $\mu(A_k \cap A_l) > 0$ or $\mu((B_l \setminus A_l) \cap B_k) > 0$. In either case, $\hat{y} > y$. Since $\hat{L} = L$, this contradicts the optimality of O. ■

We can now prove Lemma 1.

Proof of Lemma 1. (Part 1) Let O be an optimal organization and $k, l \in L$ be such that $k < l$. Then $(B_l \setminus A_l) \cap B_k = \emptyset$ and $A_l \cap A_k = \emptyset$. Thus, $B_l \cap A_k = B_l \cap A_k \cap A_l^c = (B_l \setminus A_l) \cap A_k \subseteq (B_l \setminus A_l) \cap B_k = \emptyset$.

Next, let $k, l \in L$ be such that $k \neq l$. Then either $k < l$ or $l < k$; for concreteness, let $k < l$. Then $A_l \cap A_k = \emptyset$ and $(B_l \setminus A_l) \cap (B_k \setminus A_k) \subseteq (B_l \setminus A_l) \cap B_k = \emptyset$ from the previous paragraph. This completes the proof.

(Part 2) Suppose not; then $B_L \setminus A_L \neq \emptyset$. Define an organization $\hat{O}$ to be equal to O except that $\hat{B}_L = A_L$. We have, clearly, that $\hat{A}_j \setminus \cup_{i<j} \hat{A}_i = A_j \setminus \cup_{i<j} A_i$ and $\hat{A}_j^c \setminus \cup_{i<j} \hat{B}_i = A_j^c \setminus \cup_{i<j} B_i$ for each $j \in L$. Hence, $\hat{\alpha}_j = \alpha_j$ for all $j \in L$. It then follows that $\hat{O}$ is an organization. We have that $\hat{\nu}_L < \nu_L$ which, together with $\hat{\alpha}_j = \alpha_j$ for all $j \in L$, implies that $\hat{y} > y$. Since $\hat{L} = L$, this contradicts the optimality of O.

(Part 3) Suppose not; then there is $i \in \{1, \ldots, L-1\}$ such that $B_i \cap A_{i+1} \neq \emptyset$. Hence, $(B_i \setminus A_i) \cap A_{i+1} \neq \emptyset$ since $A_i \cap A_{i+1} = \emptyset$ by part 1. Define an organization $\hat{O}$ to be equal to O except that $\hat{B}_i = B_i \setminus A_{i+1}$. We clearly have that $\hat{\alpha}_j = \alpha_j$ for all $j \leq i$. Since $\cup_{l<j} \hat{B}_l = \cup_{l<j} B_l$ for each $j > i$, it follows that $\hat{\alpha}_j = \alpha_j$ for all $j > i$ as well. Hence, $\hat{O}$ is an organization.

We have that $\hat{B}_i \setminus \hat{A}_i = (B_i \setminus A_i) \setminus A_{i+1}$. Hence, $\mu(\hat{B}_i \setminus \hat{A}_i) = \mu(B_i \setminus A_i) - \mu((B_i \setminus A_i) \cap A_{i+1}) < \mu(B_i \setminus A_i)$ since $(B_i \setminus A_i) \cap A_{i+1} \neq \emptyset$ and $(B_i \setminus A_i) \cap A_{i+1} \in \mathcal{I}$. Thus, $\hat{\nu}_i < \nu_i$ which, together with $\hat{\alpha}_j = \alpha_j$ for all $j \in L$, implies that $\hat{y} > y$. Since $\hat{L} = L$, this contradicts the optimality of O. ■

We conclude this section with the following lemma which establishes the decomposition of an optimal organization's costs using the sets in the partition $\mathcal{C}$.

Lemma A.10 *If O is an optimal organization, then $\sum_{l \in L} \alpha_l \nu_l = \sum_{C \in \mathcal{C}} c_C \mu(C)$.*

Proof. Indeed,

$$
\begin{aligned}
&\sum_{l\in L}\alpha_l\nu_l=\sum_{l\in L}\alpha_l(c\mu(A_l)+\xi\mu(B_l\setminus A_l))\\
&=\sum_{l\in L}\alpha_l\Big[\sum_{j<l-1}c\mu(A_l\cap(B_j\setminus A_j))+c\mu(A_l\cap(\cap_{j<l-1}(B_j\setminus A_j)^c))\\
&+\sum_{j>l+1}\xi\mu(A_j\cap(B_l\setminus A_l))+\xi\mu((B_l\setminus A_l)\cap(\cap_{j>l+1}A_j^c))\Big]\\
&=\sum_{l\in L}\sum_{j<l-1}\alpha_l c\mu(A_l\cap(B_j\setminus A_j))+\sum_{l\in L}\alpha_l c\mu(A_l\cap(\cap_{j<l-1}(B_j\setminus A_j)^c))\\
&+\sum_{l\in L}\sum_{j<l-1}\alpha_j\xi\mu(A_l\cap(B_j\setminus A_j))+\sum_{l\in L}\alpha_l\xi\mu((B_l\setminus A_l)\cap(\cap_{j>l+1}A_j^c))\\
&=\sum_{C\in\mathcal{C}}c_C\mu(C).
\end{aligned}
$$

■

A.5 No gaps

In this section we establish Theorem 2. We start by stating two technical lemmas. Lemma A.11 provides two useful formulas for the expected amount of time α_i that layer i spends helping layer 1.

Lemma A.11 *If O satisfies $A_l \cap A_k = \emptyset$ for each $k, l \in L$ such that $k \neq l$, then, for each $i > 1$, $\alpha_i = hF(A_i) + \pi(1 - F(A_i \cup (\cup_{j<i}B_j))) = (h-\pi)F(A_i) + \pi(1 - F(\cup_{j<i}B_j \setminus A_i))$.*

Proof. Let $i \in L \setminus \{1\}$. We have that $A_i \subseteq (\cup_{j<i}A_i)^c = \cap_{j<i}A_j^c$ and, hence, $F(A_i \setminus \cup_{j<i}A_j) = F(A_i)$. Furthermore, $(A_i^c \setminus (\cup_{j<i}B_j))^c = A_i \cup (\cup_{j<i}B_j) = A_i \cup (\cup_{j<i}B_j \setminus A_i)$. Hence, $F(A_i^c \setminus (\cup_{j<i}B_j)) = 1 - F(A_i \cup (\cup_{j<i}B_j)) = 1 - F(A_i) - F(\cup_{j<i}B_j \setminus A_i)$ since A_i and $\cup_{j<i}B_j \setminus A_i$ are disjoint. Since $\alpha_i = hF(A_i \setminus \cup_{j<i}A_j) + \pi F(A_i^c \setminus \cup_{j<i}B_j)$, the result follows. ■

Lemma A.12 is a consequence of the assumption that the density f of F is strictly decreasing.

Lemma A.12 *Let $a, \hat{a}, b, \hat{b} \in \Omega$ be such that $a < b$, $\hat{a} < \hat{b}$, $\hat{a} < a$ and $b - a = \hat{b} - \hat{a}$. Then:*

(a) $F([\hat{a}, \hat{b})) > F([a, b))$.

(b) There exists a unique $b' \in (\hat{a}, \hat{b})$ such that $F([\hat{a}, b')) = F([a, b))$.

Proof. To prove part (a), let $\varphi : [a,b] \to [\hat{a},\hat{b}]$ be defined by $\varphi(x) = x + \hat{a} - a$ for each $x \in [a,b]$. Then $\varphi'(x) = 1$ for each $x \in [a,b]$, $\varphi(a) = \hat{a}$ and $\varphi(b) = \hat{b}$. Moreover, for each $x \in [a,b]$, $f \circ \varphi(x) = f(x + \hat{a} - a) > f(x)$ since f is strictly decreasing. Hence, by the change of variable theorem, e.g. Rudin (1976, Theorem 6.19, p. 132), $F([\hat{a},\hat{b})) = \int_{\hat{a}}^{\hat{b}} f\mathrm{d}\mu = \int_a^b f \circ \varphi \mathrm{d}\mu > \int_a^b f\mathrm{d}\mu = F([a,b))$.

As for (b), let $g(\omega) = F([\hat{a},\omega))$ for each $\omega \geq \hat{a}$. Since $g(\hat{a}) = 0 < F([a,b))$, $g(\hat{b}) = F([\hat{a},\hat{b})) > F([a,b))$ by part (a) and g is continuous, it follows by the intermediate value theorem that such b' exists. Since g is strictly increasing, such b' is unique. ■

Recall that $a_i = \min B_i$ for each $i \in L$ with the standard convention that $\min \emptyset = \infty$. Lemma A.13 establishes part of Theorem 2.

Lemma A.13 *If O is an optimal organization, then* $\min_{1\leq i\leq L} a_i = 0$.

Proof. Suppose not. For each $C \in \mathcal{C}$, let $a_C = \min C$; then $\min_{1\leq i\leq L} a_i = \min_{C\in\mathcal{C}} a_C$. Letting $C \in \mathcal{C}$ be such that $a_C = \min_{1\leq i\leq L} a_i$, then $a_C > 0$. Thus, $[0, a_C) \subseteq (\cup_{l=1}^L B_l)^c$.

Let $\varepsilon > 0$ be such that $[a_C, a_C + \varepsilon) \subseteq C$ and let $0 < \varepsilon' < \varepsilon$ be such that $F([0,\varepsilon')) = F([a_C, a_C + \varepsilon))$; the existence of ε' follows by Lemma A.12. Define an organization $\hat{O}$ equal to O except that $\hat{C} = [0,\varepsilon') \cup (C \setminus [a_C, a_C + \varepsilon))$ and $\{\hat{A}_l, \hat{B}_l\}_{l=1}^L$ are defined from $\{\hat{D} : D \in \mathcal{C}\}$ via the formulas in Footnote 13. Note that $\hat{C} \cap \hat{D} = \emptyset$ whenever $D \in \mathcal{C}$ is such that $C \neq D$ because $[0,\varepsilon) \subseteq (\cup_{l=1}^L B_l)^c$.

We have that $F(\hat{D}) = F(D)$ for each $D \in \mathcal{C}$. Thus, $F(\hat{A}_j) = \sum_{D\in\mathcal{C}(A_j)} F(\hat{D}) = \sum_{D\in\mathcal{C}(A_j)} F(D) = F(A_j)$ for each $j \in L$. In addition, for each $j \in L$, let $\mathcal{C}_j = (\cup_{l\leq j}\mathcal{C}(A_l)) \cup (\cup_{l<j}\mathcal{C}(B_l \setminus A_l))$. Then, $F(A_j \cup (\cup_{l<j} B_l)) = F((\cup_{l\leq j} A_l) \cup (\cup_{l<j}(B_l \setminus A_l))) = F(\cup_{D\in\mathcal{C}_j} D) = \sum_{D\in\mathcal{C}_j} F(D) = \sum_{D\in\mathcal{C}_j} F(\hat{D}) = F(\hat{A}_j \cup (\cup_{l<j} \hat{B}_l))$. It then follows from Lemma A.11 that $\hat{\alpha}_j = \alpha_j$ for all $j \in L$. Thus, $\hat{O}$ is an organization.

We have that $\mu(\hat{C}) < \mu(C)$ and $\mu(\hat{D}) = \mu(D)$ for each $D \in \mathcal{C}\setminus\{C\}$. Moreover, by part 1 of Lemma 1, $F(\cup_{l\in L}\hat{A}_l) = \sum_{l\in L} F(\hat{A}_l) = \sum_{l\in L} F(A_l) = F(\cup_{l\in L} A_l)$ since $F(\hat{A}_l) = F(A_l)$ for each $l \in L$. It then follows that $\hat{y} > y$. This, together with $\hat{L} = L$, contradicts the optimality of O. ■

We now prove Theorem 2.

Proof of Theorem 2. Suppose not. Since $\cup_{l\in L} B_l = \cup_{C\in\mathcal{C}} C$ and $C \in \mathcal{I}$ for each $C \in \mathcal{C}$, we write $C = \cup_{r=1}^{m_C} [a_{Cr}, b_{Cr})$ where $[a_{Cr}, b_{Cr}) \cap [a_{Cr'}, b_{Cr'}) = \emptyset$ whenever $r \neq r'$. We then order the set $\{a_{Cr}, b_{Cr} : C \in \mathcal{C}, 1 \leq r \leq m_C\}$ and write it as $\{a_1, b_1, \ldots, a_m, b_m\}$

with $a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_m < b_m$, so that $\cup_{l \in L} B_l = \cup_{r=1}^{m} [a_r, b_r)$. It follows by Lemma A.13 that there is $i \in \{2, \ldots, m\}$ such that $a_i > b_{i-1}$.

We have that $[b_{i-1}, a_i) \subseteq (\cup_{l=1}^{L} B_l)^c$. Let $C \in \mathcal{C}$ and $1 \leq r' \leq m_C$ be such that $a_{Cr'} = a_i$. By Lemma A.12, let $\varepsilon > 0$ be such that $F([b_{i-1}, b_{Cr'} - \varepsilon)) = F([a_{Cr'}, b_{Cr'}))$ and $\mu([b_{i-1}, b_{Cr'} - \varepsilon)) < \mu([a_{Cr'}, b_{Cr'}))$. Define an organization $\hat{O}$ equal to O except that $\hat{C} = (C \setminus [a_{Cr'}, b_{Cr'})) \cup [b_{i-1}, b_{Cr'} - \varepsilon)$ and $\{\hat{A}_l, \hat{B}_l\}_{l=1}^{L}$ are defined from $\{\hat{D} : D \in \mathcal{C}\}$ via the formulas in Footnote 13. Note that $\hat{C} \cap \hat{D} = \emptyset$ whenever $D \in \mathcal{C}$ is such that $C \neq D$ because $[b_{i-1}, a_i) \subseteq (\cup_{l=1}^{L} B_l)^c$.

We clearly have that $\hat{\mathcal{C}} = \{\hat{D} : D \in \mathcal{C}\}$ and that $\hat{D} = \cup_{r=1}^{m_D} [\hat{a}_{Dr}, \hat{b}_{Dr})$ for each $D \in \mathcal{C}$ (with $\hat{a}_{Cr'} = b_{i-1}$, $\hat{b}_{Cr'} = b_{Cr'} - \varepsilon$, and $\hat{a}_{Dr} = a_{Dr}$ and $\hat{b}_{Dr} = b_{Dr}$ whenever $(D, r) \neq (C, r')$). An argument completely analogous to the one in the proof of Lemma A.13 shows that $\hat{O}$ is an organization and $\hat{y} > y$. But this, together with $\hat{L} = L$, contradicts the optimality of O. ■

A.6 Order of sets

In this section we prove Theorem 3 by establishing Lemmas A.14 and A.15 below.

Lemma A.14 *If O is an optimal organization, then $C < C'$ for all $C, C' \in \mathcal{C}$ with $c_C > c_{C'}$.*

Proof. Order $\mathcal{C}$ and write $\mathcal{C} = \{C_1, \ldots, C_n\}$ such that $c_{C_l} \geq c_{C_{l+1}}$ for each $l = 1, \ldots, n$. If the conclusion does not hold, then there exists $i, j \in \{1, \ldots, n\}$ such that $c_{C_i} > c_{C_j}$ but $C_i < C_j$ does not hold. Since $C_i, C_j \in \mathcal{I}$, $C_i = \cup_{r=1}^{m} E_r$ and $C_j = \cup_{r=1}^{m'} E'_r$ where $\{E_r : r = 1, \ldots, m\}$ is a collection of pairwise disjoint intervals and so is $\{E'_r : r = 1, \ldots, m'\}$. Since $C_i \cap C_j = \emptyset$, $E_r \cap E'_s = \emptyset$ for each $1 \leq r \leq m$ and $1 \leq s \leq m'$. Since $C_i < C_j$ does not hold, there is $1 \leq r \leq m$ and $1 \leq s \leq m'$ such that $E'_s < E_r$.

Let $E_r = [a, b)$ and $E'_s = [a', b')$; then $a' < b' \leq a < b$. Let $\Omega' = E_r \cup E'_s$ and let $\hat{a} \in \Omega$ be such that $F(\Omega' \cap [0, \hat{a})) = F(E_r)$; the existence and uniqueness of $\hat{a}$ follows by an argument analogous to that of Lemma A.12. In addition, $\mu(\Omega' \cap [0, \hat{a})) < \mu(E_r)$ by Lemma A.12. Let $\hat{O}$ be equal to O except that $\hat{E}_r = \Omega' \cap [0, \hat{a})$ and $\hat{E}'_s = \Omega' \setminus [0, \hat{a})$. Thus, $F(\hat{D}) = F(D)$ for each $D \in \mathcal{C}$ and, hence, $\hat{\alpha}_l = \alpha_l$ for each $l \in L$. Thus, $\hat{O}$ is an organization.

Furthermore, $\mu(\hat{D}) = \mu(D)$ for each $D \in \mathcal{C} \setminus \{C_i, C_j\}$, $\mu(\hat{C}_i) < \mu(C_i)$ and $\mu(\hat{C}_j) + \mu(\hat{C}_i) = \mu(C_j) + \mu(C_i)$. We have that $F(\cup_{l \in L} A_l) = F(\cup_{l \in L} \hat{A}_l)$ since $F(D) = F(\hat{D})$ for each $D \in \mathcal{C}$. By Lemma A.10, $y = \frac{F(\cup_{l \in L} A_l) - \sum_{D \in \mathcal{C}} \mu(D) c_D}{\gamma}$ and $\hat{y} = \frac{F(\cup_{l \in L} A_l) - \sum_{D \in \mathcal{C}} \mu(\hat{D}) c_D}{\gamma}$. Consequently, letting $\rho = \mu(C_i) - \mu(\hat{C}_i)$, we have that $\hat{y} - y = \frac{(c_{C_i} - c_{C_j})\rho}{\gamma} > 0$. This, together with $\hat{L} = L$, contradicts the optimality of O. ■

Lemma A.15 *If O is an optimal organization and $\mathcal{C} = \{C_1, \ldots, C_{|\mathcal{C}|}\}$ is such that $c_{C_1} \geq \cdots \geq c_{C_{|\mathcal{C}|}}$, then there is an optimal organization $\hat{O}$ such that $\hat{y} = y$, $\hat{L} = L$, $\hat{l}_i = l_i$ for each $i \in L$, $\hat{\mathcal{C}} = \{\hat{C} : C \in \mathcal{C}\}$, $\hat{C}_1 < \cdots < \hat{C}_{|\mathcal{C}|}$ and $F(\hat{C}) = F(C)$ for each $C \in \mathcal{C}$.*

Proof. Define $\{c_1, \ldots, c_n\} = \{c_C : C \in \mathcal{C}\}$ with $c_1 > \ldots > c_n$ and $\mathcal{C}_i = \{C \in \mathcal{C} : c_C = c_i\}$ for each $1 \leq i \leq n$. For each $1 \leq i \leq n$, there is $k_i \in \{1, \ldots, |\mathcal{C}|\}$ and $r_i \in \{0, \ldots, |\mathcal{C}| - 1\}$ such that $\mathcal{C}_i = \{C_{k_i}, \ldots, C_{k_i + r_i}\}$. Let I be the set of $i \in \{1, \ldots, n\}$ such that $C_{k_i} < \cdots < C_{k_i + r_i}$ does not hold; we may assume that $I \neq \emptyset$ since, otherwise, just set $\hat{O} = O$. We will define an organization $\tilde{O}$ such that $\tilde{y} = y$, $\tilde{L} = L$, $\tilde{l}_i = l_i$ for each $i \in L$, $\tilde{\mathcal{C}} = \{\tilde{C} : C \in \mathcal{C}\}$, $\tilde{C}_{k_i} < \cdots < \tilde{C}_{k_i + r_i}$ for each $i \notin \tilde{I}$, $F(\tilde{C}) = F(C)$ for each $C \in \mathcal{C}$ and $|\tilde{I}| = |I| - 1$. By repeating this argument at most $|I|$ times, we obtain the desired $\hat{O}$.

By Lemma A.14, $\cup_{C \in \mathcal{C}_i} C < \cup_{C \in \mathcal{C}_{i+1}} C$ for each $i = 1, \ldots, n-1$. By Theorem 2, $\cup_{l \in L} B_l = \cup_{i=1}^{n} \cup_{C \in \mathcal{C}_i} C$ is an interval. Thus, it follows that $\cup_{C \in \mathcal{C}_i} C$ is an interval for each $i = 1, \ldots, n$.

Let $i \in I$ and $\cup_{C \in \mathcal{C}_i} C = [a, b)$. Then obtain $\{\tilde{C} : C \in \mathcal{C}_i\}$ such that $\tilde{C}$ is an interval and $F(\tilde{C}) = F(C)$ for each $C \in \mathcal{C}_i$ as follows. Write $\mathcal{C}_i = \{C_{k_i}, \ldots, C_{k_i + r_i}\} = \{C_{i_1}, \ldots, C_{i_{r_i+1}}\}$. Let b_1 be such that $F([a, b_1)) = F(C_{i_1})$ and set $\tilde{C}_{i_1} = [a_1, b_1)$ with $a_1 = a$; assuming that $\tilde{C}_{i_1}, \ldots, \tilde{C}_{i_{j-1}}$ are such that, for each $1 \leq l \leq j-1$, $F(\tilde{C}_{i_l}) = F(C_{i_l})$ and $\tilde{C}_{i_l} = [a_l, b_l)$ with $a = a_1 < b_1 = a_2 < b_2 = \cdots = a_{j-1} < b_{j-1}$, let $a_j = b_{j-1}$ and b_j such that $F([a_j, b_j)) = F(C_{i_j})$. The existence and uniqueness of b_j follows by an argument analogous to that of Lemma A.12.

Let $\tilde{O}$ be equal to O except that, for each $C \in \mathcal{C}_i$, C is replaced with $\tilde{C}$. We have that $\tilde{L} = L$ with $\tilde{l}_i = l_i$ for each $i \in L$, $\tilde{\mathcal{C}} = \{\tilde{C} : C \in \mathcal{C}\}$, $\tilde{C}_{k_i} < \cdots < \tilde{C}_{k_i + r_i}$, $F(\tilde{C}) = F(C)$ for each $C \in \mathcal{C}$ and $|\tilde{I}| = |I| - 1$ by construction. Also, $\cup_{C \in \mathcal{C}_i} C = \cup_{C \in \mathcal{C}_i} \tilde{C}$ so that $\sum_{C \in \mathcal{C}_i} \mu(C) = \sum_{C \in \mathcal{C}_i} \mu(\tilde{C})$. It then follows by Lemma A.10 that $\tilde{y} - y = c_i \left(\sum_{C \in \mathcal{C}_i} \mu(C) - \sum_{C \in \mathcal{C}_i} \mu(\tilde{C}) \right) = 0$. This completes the proof. ■

A.7 Statement and proof of Theorem A.1

Whenever the knowledge sets are pairwise disjoint, thus, in any optimal organization, $\alpha_l = hF(A_l) + \pi(1 - F(A_l \cup (\cup_{j<l} B_j)))$ for each $l > 1$ (see Lemma A.11). Since A_l and B_l can be written as unions of sets in the partition $\mathcal{C}$, we can also write the output of an organization in terms of the sets of the partition $\mathcal{C}$. In fact, let, for each $1 \leq l < L$,

$$\mathcal{C}(B_l \setminus A_l) = \{A_j \cap (B_l \setminus A_l) : j > l+1\} \cup \{(B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c)\}$$

and, for each $l \in L$,

$$\mathcal{C}(A_l) = \{A_l \cap (B_j \setminus A_j) : j < l-1\} \cup \{A_l \cap (\cap_{j<l-1} (B_j \setminus A_j)^c)\} \text{ and}$$
$$\mathcal{C}_l = (\cup_{j \leq l} \mathcal{C}(A_j)) \cup (\cup_{j<l} \mathcal{C}(B_j \setminus A_j)).$$

Then, noting that $\alpha_1 = 1$ and $\mathcal{C}(A_1) = \{A_1\}$,

$$\theta = F(A_1) - \nu_1 + \sum_{l=2}^{L} \left(\sum_{C \in \mathcal{C}(A_l)} (1 - h\nu_l) F(C) - \pi\nu_l (1 - \sum_{C \in \mathcal{C}_l} F(C)) \right),$$

$$\gamma = 1 + \sum_{l=2}^{L} \left(h \sum_{C \in \mathcal{C}(A_l)} F(C) + \pi(1 - \sum_{C \in \mathcal{C}_l} F(C)) \right)$$

and, as before, $y = \frac{\theta}{\gamma}$.

In the following result, we consider changes to the knowledge and screening sets of an organization such that ν_l remains unchanged for each $l \in L$ and, for some $C, C' \in \mathcal{C}$ and $\varepsilon > 0$, $F(C)$ increases by $\varepsilon > 0$, $F(C')$ decreases by ε and $F(D)$ remains unchanged for each $D \in \mathcal{C} \setminus \{C, C'\}$. To state it formally, we treat the vector $(\nu_l)_{l \in L}$ as a constant in the above formula for output and replace the vector $(F(C))_{C \in \mathcal{C}}$ with an arbitrary vector $(\phi_C)_{C \in \mathcal{C}}$. The resulting output is denoted by $y((\phi_C)_{C \in \mathcal{C}})$, i.e.

$$y((\phi_C)_{C \in \mathcal{C}}) = \frac{\phi_{A_1} - \nu_1 + \sum_{l=2}^{L} \left(\sum_{C \in \mathcal{C}(A_l)} (1 - h\nu_l)\phi_C - \pi\nu_l (1 - \sum_{C \in \mathcal{C}_l} \phi_C) \right)}{1 + \sum_{l=2}^{L} \left(h \sum_{C \in \mathcal{C}(A_l)} \phi_C + \pi(1 - \sum_{C \in \mathcal{C}_l} \phi_C) \right)}.$$

We then let $y_{F(C),F(C')}(\varepsilon)$ be the output resulting from increasing $F(C)$ and decreasing $F(C')$ by $\varepsilon > 0$, i.e.

$$y_{F(C),F(C')}(\varepsilon) = y((\phi_C))_{C \in \mathcal{C}})$$

with $\phi_C = F(C) + \varepsilon$, $\phi_{C'} = F(C') - \varepsilon$ and $\phi_D = F(D)$ for each $D \in \mathcal{C} \setminus \{C, C'\}$.

Theorem A.1 *If O is an optimal organization and $C, C' \in \mathcal{C}$ are such that there is $0 < \bar{\varepsilon} \leq \min\{F(C), F(C')\}$ such that $y_{F(C),F(C')}(\varepsilon) > y$ for each $0 < \varepsilon < \bar{\varepsilon}$, then $C < C'$.*

Proof. Let O be an optimal organization and $C, C' \in \mathcal{C}$ and $\bar{\varepsilon} > 0$ be as in the statement. Suppose that $C < C'$ does not hold. Then there exist $a, a' \in \Omega$ and $\varepsilon > 0$ such that $a' + \varepsilon < a$, $[a', a' + \varepsilon) \subseteq C'$ and $[a, a + \varepsilon) \subseteq C$. Let $\varepsilon' = F([a', a' + \varepsilon)) - F([a, a + \varepsilon))$; it follows by Lemma A.12 that $\varepsilon' > 0$. Since $\varepsilon' \to 0$ as $\varepsilon \to 0$, reduce ε if necessary so that $\varepsilon' < \bar{\varepsilon}$.

Consider an organization $\hat{O}$ equal to O except that $\hat{C} = (C \setminus [a, a + \varepsilon)) \cup [a', a' + \varepsilon)$ and $\hat{C}' = (C' \setminus [a', a' + \varepsilon)) \cup [a, a + \varepsilon)$. Reducing ε further if necessary, we may assume that $\hat{\alpha}_l > 0$ for each $l \in L$ and, therefore, that $\hat{O}$ is indeed an organization. Then $\mu(\hat{D}) = \mu(D)$ for each $D \in \mathcal{C}$, $F(\hat{C}) = F(C) + \varepsilon'$, $F(\hat{C}') = F(C') - \varepsilon'$ and $F(\hat{D}) = F(D)$ for each $D \in \mathcal{C} \setminus \{C, C'\}$. It then follows that $\hat{y} = y_{F(C), F(C')}(\varepsilon') > y$. But this together with $\hat{L} = L$ contradicts the optimality of O. This contradiction then shows that $C < C'$. ■

A.8 Proof of Corollary 1

For part (a), note that $c_{A_l \cap (B_j \setminus A_j)} = \alpha_l c + \alpha_j \xi > \alpha_l c = c_{A_l \cap (\cap_{k<l-1}(B_k \setminus A_k)^c)}$ since $\alpha_j > 0$ by Lemma A.7. Theorem 3 then implies $A_l \cap (B_j \setminus A_j) < A_l \cap (\cap_{k<l-1}(B_k \setminus A_k)^c)$.

For part (b), note that $c_{A_l \cap (B_j \setminus A_j)} = \alpha_l c + \alpha_j \xi > \alpha_j \xi = c_{(B_j \setminus A_j) \cap (\cap_{k>j+1} A_k^c)}$ since $\alpha_l > 0$ by Lemma A.7. Theorem 3 then implies $A_l \cap (B_j \setminus A_j) < (B_j \setminus A_j) \cap (\cap_{k>j+1} A_k^c)$.

For part (c), $\xi < c$ implies $c_{A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c)} = \alpha_l c > \alpha_l \xi = c_{(B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c)}$ since $\alpha_l > 0$ by Lemma A.7. Theorem 3 then implies $A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c) < (B_l \setminus A_l) \cap (\cap_{j>l+1} A_j^c)$.

A.9 Proof of Corollary 2

We will use the following lemma in the proof of Corollaries 2 and 3. It states that every layer other than layer 1 has a nonempty screening set. This is because a layer with an empty screening set is of no use to layer 1 and could be removed to obtain an organization with a higher output and a smaller number of layers.

Lemma A.16 *If O is an optimal organization, then $\mu(B_j) > 0$ for all $j \in L \setminus \{1\}$.*

Proof. Let O be an optimal organization and let $J = \{j \in L \setminus \{1\} : \mu(B_j) = 0\}$. Assume that $J \neq \emptyset$ and define an organization $\hat{O}$ as follows: $\hat{L} = L \setminus J$; for each $k \in \hat{L}$,

$\hat{A}_k = A_k$ and $\hat{B}_k = B_k$; for each $k \in \hat{L} \setminus \{1\}$, $\hat{l}_k = l_k$, $\hat{\beta}_k = \frac{\hat{\alpha}_k}{\hat{\gamma}}$, $\hat{t}_k^p = 0$ and $\hat{t}_k^h = 1$; and $\hat{l}_1 = \hat{L}$, $\hat{\prec}_1 = <$, $\hat{\beta}_1 = \frac{1}{\hat{\gamma}}$, $\hat{t}_1^p = 1$ and $\hat{t}_1^h = 0$.

Let $k \in \hat{L} \setminus \{1\}$, $\mathcal{L}_k = \{1, \ldots, k-1\}$ and $\hat{\mathcal{L}}_k = \{l \in \hat{L} : l < k\}$. For each $j \in J$, $\mu(B_j) = 0$ and $B_j \in \mathcal{I}$ imply that $B_j = \emptyset$ and $A_j = \emptyset$, the latter because $A_j \subseteq B_j$. Hence, $\cup_{l \in \hat{\mathcal{L}}_k} A_l = \cup_{l \in \mathcal{L}_k} A_l$ and $\cup_{l \in \hat{\mathcal{L}}_k} B_l = \cup_{l \in \mathcal{L}_k} B_l$. Thus, $\hat{\alpha}_k = \alpha_k$ for each $k \in \hat{L}$. We have that $\alpha_j > 0$ for each $j \in L$ (otherwise, $\beta_j = 0$ for some $j \in L$) and, hence, $\beta_j > 0$ for all $j \in \hat{L}$. Thus, $\hat{O}$ is indeed an organization. Furthermore, $\hat{\gamma} = \sum_{k \in \hat{L}} \alpha_k < \sum_{k \in L} \alpha_k = \gamma$. Since $\mu(B_j) = 0$ for each $j \in J$, it follows that $\nu_j = 0$ for each $j \in J$ and, thus, $\hat{\theta} = F(\cup_{k \in \hat{L}} A_k) - \sum_{k \in \hat{L}} \alpha_k \nu_k = F(\cup_{k \in L} A_k) - \sum_{k \in L} \alpha_k \nu_k = \theta$. Therefore, $\hat{y} = \frac{\theta}{\hat{\gamma}} > \frac{\theta}{\gamma} = y$. This, together with $\hat{L} < L$, contradicts the optimality of O. ■

Proof of Corollary 2. For part (a), let $C \in \mathcal{C} \setminus \{A_1\}$. We will show that $y_{F(A_1),F(C)}(\varepsilon) > y$ for each $0 < \varepsilon < \min\{F(A_1), F(C)\}$.

Analogously to $y = \frac{\theta}{\gamma}$, write $y_{F(A_1),F(C)}(\varepsilon) = \frac{\theta_{F(A_1),F(C)}(\varepsilon)}{\gamma_{F(A_1),F(C)}(\varepsilon)}$. Since $A_1 \in \mathcal{C}_l \setminus \mathcal{C}(A_l)$, it follows that

$$\theta_{F(A_1),F(C)}(\varepsilon) - \theta = \varepsilon \left[1 - \sum_{l=2}^{L} \left((1 - h\nu_l) 1_{\mathcal{C}(A_l)}(C) + \pi \nu_l (1 - 1_{\mathcal{C}_l}(C)) \right) \right] \text{ and}$$

$$\gamma_{F(A_1),F(C)}(\varepsilon) - \gamma = \varepsilon \sum_{l=2}^{L} \left[-h 1_{\mathcal{C}(A_l)}(C) + \pi(-1 + 1_{\mathcal{C}_l}(C)) \right].$$

We have that $\gamma_{F(A_1),F(C)}(\varepsilon) - \gamma \leq 0$. In addition, note that if $C \in \mathcal{C}(A_l)$ for some $l > 1$, then $C \subseteq A_l$ and, hence, $C \notin \mathcal{C}(A_j)$ for each $j \neq l$. Thus, $\theta_{F(A_1),F(C)}(\varepsilon) - \theta \geq \varepsilon[1 - \max_{l \geq 2}(1 - h\nu_l)] = \varepsilon h \min_{l \geq 2} \nu_l > 0$, where the last inequality follows from Lemma A.16. It then follows that $y_{F(A_1),F(C)}(\varepsilon) > y$. Thus, part (a) follows from Theorem A.1.

For part (b), let $C = A_l \cap (B_j \setminus A_j)$ and $C' = A_l \cap (B_k \setminus A_k)$. Then $C, C' \in \mathcal{C}(A_l)$, $C \in \mathcal{C}_i$ for each $i \geq j+1$ and $C' \in \mathcal{C}_i$ for each $i \geq k+1$. We may assume that C and C' are nonempty since otherwise there is nothing to prove. Then, for each $0 < \varepsilon < \min\{F(C), F(C')\}$,

$$y_{F(C),F(C')}(\varepsilon) = \frac{\theta + \varepsilon\pi \sum_{i=j+1}^{k} \nu_i}{\gamma - \varepsilon(k-j)\pi} > y.$$

Thus, part (b) follows from Theorem A.1.

For part (c), let $C = (B_j \setminus A_j) \cap (\cap_{l > j+1} A_l^c)$ and $C = (B_k \setminus A_k) \cap (\cap_{l > k+1} A_l^c)$. Then $C, C' \notin \mathcal{C}(A_l)$ for each $l \in L$, $C \in \mathcal{C}_i$ for each $i \geq j+1$ and $C' \in \mathcal{C}_i$ for each $i \geq k+1$. We may assume that C and C' are nonempty since otherwise there is nothing to prove.

Then, for each $0 < \varepsilon < \min\{F(C), F(C')\}$,

$$y_{F(C),F(C')}(\varepsilon) = \frac{\theta + \varepsilon\pi \sum_{i=j+1}^{k} \nu_i}{\gamma - \varepsilon(k-j)\pi} > y.$$

Thus, part (c) follows from Theorem A.1.

For part (d), let $C = A_k \cap (\cap_{j<k-1}(B_j \setminus A_j)^c)$ and $C' = A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c)$. Then $C \in \mathcal{C}(A_k) \cap (\cap_{i\geq k}\mathcal{C}_i)$, $C' \in \mathcal{C}(A_l) \cap (\cap_{i\geq l}\mathcal{C}_i)$. We may assume that C and C' are nonempty since otherwise there is nothing to prove. Then, using $\pi = h$, for each $0 < \varepsilon < \min\{F(C), F(C')\}$,

$$y_{F(C),F(C')}(\varepsilon) = \frac{\theta + \varepsilon h \sum_{j=k+1}^{l} \nu_j}{\gamma - \varepsilon(l-k)h} > y.$$

Thus, part (d) follows from Theorem A.1.

For part (e), we may assume that $l > 1$ since the conclusion follows from part (a) of Corollary 2 when $l = 1$. We may also assume that C and C' are nonempty since otherwise there is nothing to prove.

First assume that $C = A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c)$ and $C' = (B_l \setminus A_l) \cap (\cap_{j>l+1}A_j^c)$. It then follows that $C \in \mathcal{C}(A_l) \cap (\cap_{i\geq l}\mathcal{C}(A_i))$ and $C' \in (\cap_{i\in L}\mathcal{C}(A_l)^c) \cap (\cap_{i\geq l+1}\mathcal{C}(A_i))$. Then, for each $0 < \varepsilon < \min\{F(C), F(C')\}$,

$$y_{F(C),F(C')}(\varepsilon) = \frac{\theta + \varepsilon(1 + (\pi - h)\nu_l)}{\gamma - \varepsilon(\pi - h)} = \frac{\theta + \varepsilon}{\gamma} > y.$$

If instead $C' = A_j \cap (B_l \setminus A_l)$ for some $j > l+1$, then $C' \in \mathcal{C}(A_j) \cap (\cap_{i\geq l+1}\mathcal{C}(A_i))$ and

$$y_{F(C),F(C')}(\varepsilon) = \frac{\theta + \varepsilon(\pi - h)\nu_l + \varepsilon h \nu_j}{\gamma - \varepsilon(\pi - h) - \varepsilon h} = \frac{\theta + \varepsilon h \nu_j}{\gamma - \varepsilon h} > 0.$$

In each case, $C < C'$ follows from Theorem A.1.

Part (e) then follows by noting that in the remaining case where $C = A_l \cap (B_j \setminus A_j)$ for some $j < l-1$, then $C < A_l \cap (\cap_{j<l-1}(B_j \setminus A_j)^c) < C'$ by part (a) of Corollary 1. ■

A.10 Existence and computation of optimal organizations

In this section we outline our proof of the existence of η-optimal organizations and provide a method to compute them, which we use in our simulations.

It follows from our previous results that any optimal organization is such that, for some $L \in \mathbb{N}$, $\beta_1 = \frac{1}{\gamma}$, $(t_1^p, t_1^h) = (1, 0)$, $l_1 = \{1, \ldots, L\}$, $\beta_i = \frac{\alpha_i}{\gamma}$, $(t_i^p, t_i^h) = (0, 1)$ and $l_i = \{i\}$. Furthermore, $\cup_{C\in\mathcal{C}}C$ is an interval that contains 0 and we may assume that each $C \in \mathcal{C}$ is an interval as well. Thus, all that is left to determine is the number L of layers, an ordering of $\mathcal{C}$, i.e. to write $\mathcal{C} = \{C_1, \ldots, C_m\}$ with $C_1 < \cdots < C_m$ and $m = |\mathcal{C}|$, and the size $\mu(C)$ of each $C \in \mathcal{C}$. Letting $\mu_j = \mu(C_j)$ for each $j = 1, \ldots, m$, we then have $C_1 = [0, \mu_1)$, $C_2 = [\mu_1, \mu_1 + \mu_2)$ and so on, so that, for each $j = 1, \ldots, m$, $C_j = \Big[\sum_{i=1}^{j-1} \mu_i, \sum_{i=1}^{j} \mu_i\Big)$. We then obtain $\{A_1, B_1, \ldots, A_L, B_L\}$ via the formulas in Footnote 13.

Note, however, that fixing the number $L \in \mathbb{N}$ of layers, an ordering ψ of $\mathcal{C}$ (formally, ψ is a bijection from $\mathcal{C}$ onto $\{1, \ldots, m\}$) and $\mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}_+^m$ such that $\sum_{j=1}^m \mu_j \leq \mu(\Omega)$ may fail to define an organization because the requirement that $\beta_l > 0$ for each $l \in L$ may fail. To allow for this case, we say that O is a *quasi-organization* if it satisfies $\beta_l \geq 0$ for each $l \in L$ and all the conditions of the definition of an organization except possibly the one requiring $\beta_l > 0$ for each $l \in L$. For each (L, ψ, μ), let $y_{L,\psi}(\mu_1, \ldots, \mu_m)$ be the output of the resulting quasi-organization (computed using (1)) and

$$y_{L,\psi} = \max_{(\mu_1,\ldots,\mu_m)\in\mathbb{R}_+^m} y_{L,\psi}(\mu_1, \ldots, \mu_m) \tag{22}$$

$$\text{subject to } \sum_{j=1}^{m} \mu_j \leq \mu(\Omega). \tag{23}$$

Then an optimal organization O^* is obtained by letting L^* be such that

$$\max_{\psi} y_{L^*,\psi} - \eta(L^* - 1) = \max_{L} \big(\max_{\psi} y_{L,\psi} - \eta(L-1) \big), \tag{24}$$

ψ^* be such that

$$y_{L^*,\psi^*} = \max_{\psi} y_{L^*,\psi} \tag{25}$$

and $(\mu_1^*, \ldots, \mu_m^*) \in \mathbb{R}_+^m$ be such that

$$\sum_{j=1}^{m} \mu_j^* \leq \mu(\Omega) \text{ and } y_{L^*,\psi^*}(\mu_1^*, \ldots, \mu_m^*) = y_{L^*,\psi^*}. \tag{26}$$

It turns out that O^* is actually an organization which is η-optimal since it maximizes $Y = y - (L-1)\eta$. Problem (22)–(23) has a solution provided that Ω is bounded and problems (24) and (25) have choice sets that are effectively finite.[37] Thus, η-optimal organizations exist.

The proof is more complicated than what has been described above for the following reason. The above argument produces an organization O^* which is optimal *only* within

[37] Note that the output of any organization is bounded above by 1, hence $y_{L,\psi} - (L-1)\eta \leq 1 - (L-1)\eta \leq 0$ if $L \geq \frac{1+\eta}{\eta}$. Thus, the set of relevant numbers of layers is $\{1, \ldots, L'\}$ with $L' = \max\{L \in \mathbb{N} : L < (1+\eta)/\eta\}$.

the class of organizations that satisfy our previous results. If O^* fails to be optimal, then there is an organization O such that $y-(L-1)\eta > y^*-(L^*-1)\eta$ or $y=y^*$, $L=L^*$, $l_i \subseteq l_i^*$ for each $i \in L$ and $l_j \neq l_j^*$ for some $j \in L$. The organization O cannot satisfy our previous results because O is optimal within those organizations that satisfy them. We complete the proof by showing that there is an organization $\hat{O}$ that satisfy our results and is weakly better than O in the sense that it has at least the same output and at the most the same number of layers than O. The proof of this last step requires strengthening the argument of our previous results and it is left for the supplementary material of this paper.

The above method to compute η-optimal organizations can be easily modified to compute lexicographically optimal organizations. Indeed, all it takes is to replace condition (24) that determines the optimal number of layers with the requirement that L^* be the smallest $\hat{L}$ such that

$$\max_{\psi} y_{\hat{L},\psi} = \max_{L} \max_{\psi} y_{L,\psi}. \tag{27}$$

A.11 Proof of Corollary 3

The proof of (a) is by induction and, thus, we start by showing that $B_1 = A_1$. Suppose not; then $B_1 \setminus A_1 \neq \emptyset$. Define $\tilde{O}$ equal to O except that $\tilde{A}_1 = B_1$. It may happen that $\tilde{O}$ fails to be an organization since the requirement $\tilde{\beta}_i > 0$ may fail for some $i \in L$, which happens if and only if $\tilde{\alpha}_i = 0$.

Let $i \in L$ be such that $\tilde{\alpha}_i = 0$. Then $\tilde{A}_i \subseteq \cup_{l<i}\tilde{A}_l \subseteq \cup_{l<i}\tilde{B}_l$ and $\tilde{A}_i^c \subseteq \cup_{l<i}\tilde{B}_l$. Since $\tilde{B}_i \setminus \tilde{A}_i \subseteq \tilde{A}_i^c$, we have that $\tilde{B}_i = \tilde{A}_i \cup (\tilde{B}_i \setminus \tilde{A}_i) \subseteq \cup_{l<i}\tilde{B}_l$. Let $\hat{O}$ be such that $\hat{L} = L \setminus \{i \in L : \tilde{\alpha}_i = 0\}$ and $\hat{l}_1 = \hat{L}$ but otherwise equal to O. It then follows easily by induction that, for each $j \in L$, $\cup_{l<j}\tilde{A}_l = \cup_{l\in\hat{L}:l<j}\tilde{A}_l$ and $\cup_{l<j}\tilde{B}_l = \cup_{l\in\hat{L}:l<j}\tilde{B}_l$. It then follows that $\hat{\alpha}_j = \tilde{\alpha}_j$ for each $i \in \hat{L}$. Hence, $\hat{O}$ is an organization and that, for each $j \in \hat{L}$, $\hat{\alpha}_j = hF(\hat{A}_j \setminus \cup_{l<j}\hat{A}_l) + \pi F(\hat{A}_j^c \setminus \cup_{l<j}\hat{B}_l)$.

Then $\hat{\nu}_1 = c\mu(B_1) = c\mu(A_1) + c\mu(B_1 \setminus A_1) = c\mu(A_1) + \xi\mu(B_1 \setminus A_1) = \nu_1$. Furthermore, we clearly have that, for each $j > 1$, $\nu_j = \hat{\nu}_j$, $\cup_{l<j}B_l = \cup_{l<j}\hat{B}_l$, $\cup_{l<j}A_l \subseteq \cup_{l<j}\hat{A}_l$ and $\cup_{l\in L}A_l \subseteq \cup_{l\in L}\hat{A}_l$. In particular, this implies that $\hat{\alpha}_j \leq \alpha_j$ for each $j > 1$.

Suppose first that $\cup_{l\in L}A_l \subset \cup_{l\in L}\hat{A}_l$. Since $(\cup_{l\in L}\hat{A}_l) \setminus (\cup_{l\in L}A_l) \in \mathcal{I}$, it follows that $F(\cup_{l\in L}\hat{A}_l) > F(\cup_{l\in L}A_l)$. This, together with $\hat{\alpha}_j \leq \alpha_j$ for each $j > 1$, implies that $\hat{\theta} > \theta$ and $\hat{\gamma} \leq \gamma$; hence, $\hat{y} > y$. Since $\hat{L} \leq L$, this contradicts the optimality of O.

Thus, assume that $\cup_{l\in L}A_l = \cup_{l\in L}\hat{A}_l$. Hence, $B_1 \setminus A_1 \subseteq \cup_{l\in L}A_l$ and, in fact, $B_1 \setminus A_1 \subseteq \cup_{l>1}A_l$ since $(B_1 \setminus A_1) \cap A_1 = \emptyset$. Thus, there is $j > 1$ and $a, b \in \Omega$ such that $a < b$ and $[a, b) \subseteq A_j \cap (B_1 \setminus A_1)$. Hence, $A_j \setminus (\cup_{l<j}\hat{A}_l) \subseteq A_j \setminus [a, b)$ whereas $A_j \setminus (\cup_{l<j}A_l) = A_j$ by part 1 of Lemma 1. It then follows that $F(A_j \setminus (\cup_{l<j}\hat{A}_l)) < F(A_j \setminus (\cup_{l<j}A_l))$. This, together with $A_j^c \setminus (\cup_{l<j}\hat{B}_l) = A_j^c \setminus (\cup_{l<j}B_l)$, implies that $\hat{\alpha}_j < \alpha_j$. This, together with $\cup_{l\in L}A_l = \cup_{l\in L}\hat{A}_l$ and $\hat{\alpha}_l \leq \alpha_l$ for each $l > 1$, implies that $\hat{\theta} \geq \theta$ and $\hat{\gamma} < \gamma$; hence, $\hat{y} > y$. Since $\hat{L} \leq L$, this contradicts the optimality of O. This contradiction shows that $B_1 = A_1$.

Let $1 < i < L$ and assume that $B_j = A_j$ for all $j < i$. Suppose that $B_i \neq A_i$; then $B_i \setminus A_i \neq \emptyset$. Define an organization $\hat{O}$ equal to O except that $\hat{A}_i = B_i$ and $\hat{L} = L \setminus \{j \in L : \hat{\alpha}_j = 0\}$. Then, for each $j \in L$, $\nu_j = \hat{\nu}_j$, $\cup_{l<j}B_l = \cup_{l<j}\hat{B}_l$, $\cup_{l<j}A_l \subseteq \cup_{l<j}\hat{A}_l$ and $\cup_{l\in L}A_l \subseteq \cup_{l\in L}\hat{A}_l$ as above. In particular, this implies that $\hat{\alpha}_j \leq \alpha_j$ for each $j \neq i$. Furthermore, since $B_j = A_j$ for all $j < i$ and $h = \pi$, $\alpha_i = h(F(A_i \setminus \cup_{l<i}A_l) + F(A_i^c \setminus \cup_{l<i}A_l)) = h(1 - F(\cup_{l<i}A_l)) = h(F(\hat{A}_i \setminus \cup_{l<i}A_l) + F(\hat{A}_i^c \setminus \cup_{l<i}A_l)) = \hat{\alpha}_i$. Hence, $\hat{\alpha}_j \leq \alpha_j$ for each $j \in L$.

If $\cup_{l\in L}A_l \subset \cup_{l\in L}\hat{A}_l$, then $\hat{y} > y$ as above. But this, together with $\hat{L} \leq L$, contradicts the optimality of O. Thus, we may assume that $\cup_{l\in L}A_l = \cup_{l\in L}\hat{A}_l$. Hence, $B_i \setminus A_i \subseteq \cup_{l\in L}A_l$ and, in fact, $B_i \setminus A_i \subseteq \cup_{l>i}A_l$ since $(B_i \setminus A_i) \cap A_i = \emptyset$ obviously and $(B_i \setminus A_i) \cap A_j = \emptyset$ for each $j < i$ by part 1 of Lemma 1. Thus, as above, there is $j > i$ such that $\hat{\alpha}_j < \alpha_j$ and $\hat{y} > y$. Since $\hat{L} \leq L$, this contradicts the optimality of O. This contradiction shows that $B_i = A_i$.

The above shows that $B_l = A_l$ for each $l \in \{1, \ldots, L-1\}$ and part 2 of Lemma 1 shows that $B_L = A_L$. Thus, (a) follows.

By (a), we have that $B_l \setminus A_l = \emptyset$ and, hence, $A_l \in \mathcal{C}$ for each $l \in L$. For each $1 < i < L$, we have that $\alpha_i = h(1 - \sum_{l=1}^{i-1} F(A_l)) > h(1 - \sum_{l=1}^{i-1} F(A_l) - F(A_i)) = \alpha_{i+1}$ by Lemma A.16. Hence, Theorem 3 implies that $A_i < A_{i+1}$. It follows by Corollary 2 that $A_1 < A_2$. Hence, $A_1 < A_2 < \ldots < A_L$ and, thus, (b) holds. Moreover, $\beta_i = \frac{\alpha_i}{\gamma} > \frac{\alpha_{i+1}}{\gamma} = \beta_{i+1}$, proving (c).

Since $\alpha_2 = h(1 - F(A_1))$, we have that $\alpha_2 < 1$ if $h < 1$. If $A_1 \neq \emptyset$, then $A_1 < A_2$ together with Theorem 3 imply that $c = c_{A_1} > c_{A_2} = \alpha_2 c$ and, hence, $\alpha_2 < 1$. Thus, in both cases, $\beta_1 = \frac{1}{\gamma} > \frac{\alpha_2}{\gamma} = \beta_2$, proving (d).

A.12 Proof of Theorem 5

We establish Theorem 5 in a series of lemmas. Throughout this section, we assume that Assumptions (A1) and (A2) hold. Let $\eta > 0$; we will consider η-optimal as well as lexicographically optimal organizations but in the latter case η has no interpretation other than a strictly positive constant.

We next define a function $y^*_{\bar{L}}$ which we latter use to describe the output of any optimal organization. Let $\bar{L} = \max\{l \in \mathbb{N} : l < \frac{1+\eta}{\eta}\}$ and define $X_{\bar{L}} = \{(\mu_1, \ldots, \mu_{\bar{L}}) \in \mathbb{R}^{\bar{L}}_+ : \sum_{l=1}^{\bar{L}} \mu_l \leq \bar{\omega}\}$. Let, for each $(\mu_1, \ldots, \mu_{\bar{L}}) \in X_{\bar{L}}$,

$$y_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}}) = \frac{F(\sum_{i=1}^{\bar{L}} \mu_i) - c\mu_1 - c\sum_{i=2}^{\bar{L}} \alpha_i(\mu_1, \ldots, \mu_{\bar{L}})\mu_i - \xi(\bar{\omega} - \mu_1 - \mu_2)}{1 + h(\sum_{i=1}^{\bar{L}} F(\mu_i) - F(\mu_1))},$$

$$\alpha_2(\mu_1, \ldots, \mu_{\bar{L}}) = h(F(\sum_{i=1}^{\bar{L}} \mu_i) - F(\mu_1 + \sum_{i=3}^{\bar{L}} \mu_i)), \text{ and}$$

$$\alpha_i(\mu_1, \ldots, \mu_{\bar{L}}) = h(F(\mu_1 + \sum_{l=3}^{i} \mu_i) - F(\mu_1 + \sum_{l=3}^{i-1} \mu_l)) \text{ for each } i = 3, \ldots, \bar{L}.$$

This is the output of an organization that satisfies the conclusions of the theorem; if, instead, $(B_1 \setminus A_1) \cap (\cap_{j>2} A_j^c) < A_2$, then the resulting output is as follows (see Lemma A.27 below for both claims). Let, for each $(\mu_1, \ldots, \mu_{\bar{L}}) \in X_{\bar{L}}$,

$$\tilde{y}_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}}) = \frac{\phi(\mu_1, \ldots, \mu_{\bar{L}}) - c\mu_1 - c\sum_{i=2}^{\bar{L}} \tilde{\alpha}_i(\mu_1, \ldots, \mu_{\bar{L}})\mu_i - \xi(\bar{\omega} - \mu_1 - \mu_2)}{1 + h(1 - F(\bar{\omega} - \mu_2) + F(\mu_1 + \sum_{i=3}^{\bar{L}} \mu_i) - F(\mu_1)))},$$

$$\phi(\mu_1, \ldots, \mu_{\bar{L}}) = F(\mu_1 + \sum_{i=3}^{\bar{L}} \mu_i) + 1 - F(\bar{\omega} - \mu_2),$$

$$\tilde{\alpha}_2(\mu_1, \ldots, \mu_{\bar{L}}) = h(1 - F(\bar{\omega} - \mu_2)), \text{ and}$$

$$\tilde{\alpha}_i(\mu_1, \ldots, \mu_{\bar{L}}) = \alpha_i(\mu_1, \ldots, \mu_{\bar{L}}) \text{ for each } i = 3, \ldots, \bar{L}.$$

As we will show below (see Lemma A.27), $\mu_i = \mu(A_i)$ for each $i \in \{1, \ldots, \bar{L}\}$ in both cases.

Let $y^*_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}}) = \max\{y_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}}), \tilde{y}_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}})\}$ and recall that

$$y_1 = \max_{0 \leq \mu_1 \leq \bar{\omega}} (F(\mu_1) - c\mu_1).$$

Lemma A.17 states that the output of an organization where μ_i is small for each $i > 1$ cannot be much bigger than the output of best organization with just one layer.

Lemma A.17 *If $(\mu_1, \mu_2, \ldots, \mu_{\bar{L}}) \in X_{\bar{L}}$ and $\max_{2 \leq l \leq \bar{L}} \mu_l < \frac{\eta^2}{f(0)}$, then $y^*_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}}) < y_1 + \eta$.*

Proof. For convenience, let $\delta = \frac{\eta^2}{f(0)}$. We have that if $\max_{2\leq l\leq \bar{L}} \mu_l < \delta$ then $y_{\bar{L}}(\mu_1,\ldots,\mu_{\bar{L}}) \leq F(\mu_1+\delta(\bar{L}-1)) - c\mu_1$ and also $\tilde{y}_{\bar{L}}(\mu_1,\ldots,\mu_{\bar{L}}) \leq F(\mu_1+\delta(\bar{L}-1)) - c\mu_1$. The former is immediate and so is the latter when $\mu_1+\delta(\bar{L}-1) \geq \bar{\omega}$; when $\mu_1+\delta(\bar{L}-1) < \bar{\omega}$, then $\mu_1 + \sum_{i=3}^{\bar{L}} \mu_i < \mu_1+\delta(\bar{L}-2) < \bar{\omega}-\delta < \bar{\omega}-\mu_2$, hence $\phi(\mu_1,\mu_2,\ldots,\mu_{\bar{L}}) < F(\sum_{i=1}^{\bar{L}} \mu_i)$ by Lemma A.12. Thus, $\phi(\mu_1,\mu_2,\ldots,\mu_{\bar{L}}) \leq F(\mu_1 + \delta(\bar{L}-1))$ and $\tilde{y}_{\bar{L}}(\mu_1,\ldots,\mu_{\bar{L}}) \leq F(\mu_1+\delta(\bar{L}-1)) - c\mu_1$.

Next note that $F(\mu_1+\delta(\bar{L}-1)) = F(\mu_1) + F(\mu_1+\delta(\bar{L}-1)) - F(\mu_1) < F(\mu_1) + f(\mu_1)\delta(\bar{L}-1)$. This is clear when $\mu_1+\delta(\bar{L}-1) \leq \bar{\omega}$ since f is strictly decreasing. When $\mu_1+\delta(\bar{L}-1) > \bar{\omega}$, then $F(\mu_1+\delta(\bar{L}-1)) - F(\mu_1) = F(\bar{\omega}) - F(\mu_1) \leq f(\mu_1)(\bar{\omega}-\mu_1) < f(\mu_1)\delta(\bar{L}-1)$.

Thus, $y^*_{\bar{L}}(\mu_1,\ldots,\mu_{\bar{L}}) < F(\mu_1) - c\mu_1 + f(\mu_1)\delta(\bar{L}-1) = F(\mu_1) - c\mu_1 + \frac{f(\mu_1)(\bar{L}-1)\eta^2}{f(0)} < F(\mu_1) - c\mu_1 + \eta \leq y_1 + \eta$ since $\frac{f(\mu_1)}{f(0)} \leq 1$, $\bar{L}-1 < \frac{1}{\eta}$ and $F(\mu_1) - c\mu_1 \leq y_1$. ■

Define $\xi_1 = \frac{chf(\bar{\omega})\eta^2}{f(0)}$; then $\xi_1 > 0$ by (A1). Define also $\xi_2 = \pi f(\bar{\omega}) y_1$; then $y_1 > 0$ by (A2) and, thus, $\xi_2 > 0$ by (A1). Then define $\bar{\xi} = \min\{\xi_1,\xi_2\}$.

Suppose that O is an optimal organization, that $0 < \xi < \bar{\xi}$ and that $L \geq 2$. Lemma A.18 states that the set of problems that layer 1 screens but does not solve are those that neither layer 1 nor layer 2 solve. Its argument consists in showing that if there is some problems that are neither screened nor solved by layer 1, nor solved by layer 2, then it pays to have layer 1 screen them. This is so since, at the very least, the expected amount of time α_2 spent by layer 2 helping layer 1 decreases and, thus, some fraction of people can be transferred from layer 2 to layer 1, increasing output.

Lemma A.18 $B_1 \setminus A_1 = \Omega \setminus (A_1 \cup A_2)$.

Proof. Suppose not; then there is $a \in \Omega$ and $\varepsilon > 0$ such that $[a, a+\varepsilon) \subseteq (A_1 \cup A_2 \cup B_1 \setminus A_1)^c$. Consider an organization $\hat{O}$ equal to O except that $\hat{B}_1 \setminus \hat{A}_1 = (B_1 \setminus A_1) \cup [a, a+\varepsilon)$. We have that $\alpha_2 = (h-\pi)F(A_2) + \pi(1-F(B_1))$ by part 3 of Lemma 1 and Lemma A.11. In addition, $\hat{B}_1 \setminus \hat{A}_2 = (B_1 \cup [a,a+\varepsilon)) \setminus A_2 = B_1 \cup [a, a+\varepsilon)$, where $B_1 \cap [a, a+\varepsilon) = \emptyset$. Thus, $\hat{\alpha}_2(\varepsilon) = \alpha_2 - \pi F([a, a+\varepsilon))$ and $\hat{\alpha}'_2(0) = -\pi f(a)$. Since $(\cup_{j<i} B_j) \setminus A_i \subseteq (\cup_{j<i} \hat{B}_j) \setminus \hat{A}_i$, it follows by Lemma A.11 that $\hat{\alpha}_i(\varepsilon) \leq \alpha_i$ and, thus, $\hat{\alpha}'_i(0) = \lim_{\varepsilon\to 0} \frac{\hat{\alpha}_i(\varepsilon)-\alpha_i}{\varepsilon} \leq 0$ for each $i > 2$. Hence, $\hat{\theta}(\varepsilon) = \theta - \xi\varepsilon - \sum_{i=2}^{L} (\hat{\alpha}_i(\varepsilon) - \alpha_i)\nu_i$, where, recall, $\nu_i = c\mu(A_i) + \xi\mu(B_i \setminus A_i)$ for each $i \in L$. It follows that

$$\gamma \hat{y}'(0) = -\xi + \pi f(a)\nu_2 + \sum_{i>2} (-\hat{\alpha}'_i(0))\nu_i + \pi f(a) y + y \sum_{i>2} (-\hat{\alpha}'_i(0)) > \pi f(\bar{\omega}) y - \xi$$

since $f(a) > f(\bar{\omega})$. We then have that $y \geq y_1$ since $L \geq 2$ and $\pi f(\bar{\omega})y - \xi \geq \pi f(\bar{\omega})y_1 - \xi = \xi_2 - \xi > 0$ since $\xi < \bar{\xi}$. It then follows that there is $\varepsilon > 0$ such that $\hat{y}(\varepsilon) > y$, a contradiction to the optimality of O. This contradiction shows that $B_1 \setminus A_1 = \Omega \setminus (A_1 \cup A_2)$. ■

Lemma A.19 shows that each layer other than 1 does not screen problems that it does not solve. Indeed, layer 1 screens all problems other than those that layer 2 does not solve by Lemma A.18.

Lemma A.19 *$B_l \setminus A_l = \emptyset$ for each $l > 1$.*

Proof. It follows by Lemma A.18 that $\Omega = (B_1 \setminus A_1) \cup A_1 \cup A_2 = B_1 \cup A_2$. Thus, for each $l > 1$, $B_l \setminus A_l = ((B_l \setminus A_l) \cap B_1) \cup ((B_l \setminus A_l) \cap A_2) = \emptyset$ by part 1 of Lemma 1. ■

Lemma A.20 characterizes the expected amount of time α_l that layer l spends helping layer 1.

Lemma A.20 *If O is an organization such that $A_l \cap A_k = \emptyset$ for each $k, l \in L$ with $k \neq l$ and $B_1 \setminus A_1 = (A_1 \cup A_2)^c$, then $\alpha_l = hF(A_l)$ for each $l \geq 2$.*

Proof. By Lemma A.11, $\alpha_l = (h - \pi)F(A_l) + \pi(1 - F(\cup_{i<l} B_i \setminus A_l))$ for each $l \geq 2$. We have that $\alpha_2 = (h - \pi)F(A_2) + \pi(1 - F(A_1) - F(B_1 \setminus A_1)) = (h - \pi)F(A_2) + \pi F(A_2) = hF(A_2)$ since $A_2 = \Omega \setminus (A_1 \cup (B_1 \setminus A_1))$. For each $l > 2$, $\Omega \setminus A_l = (A_1 \cup (B_1 \setminus A_1) \cup A_2) \setminus A_l \subseteq (\cup_{i<l} B_i) \setminus A_l \subseteq \Omega \setminus A_l$. Hence, $\alpha_l = (h - \pi)F(A_l) + \pi(1 - (1 - F(A_l))) = hF(A_l)$. ■

Lemma A.21 to characterize expected output.

Lemma A.21 *The output of the an optimal organization is*

$$y = \frac{\sum_{i=1}^{L} F(A_i) - c\mu(A_1) - ch \sum_{i=2}^{L} F(A_i)\mu(A_i) - \xi(\bar{\omega} - \mu(A_1) - \mu(A_2))}{1 + h \sum_{i=2}^{L} F(A_i)}.$$

Proof. We have that $y = \frac{\theta}{\gamma}$; hence, the conclusion follows by Lemmas A.18–A.20. ■

Lemma A.22 states that the size $\mu(A_2)$ of the knowledge set of layer 2 is the highest among those of the managerial layers. This follows by the formula for output provided by Lemma A.21 since swapping A_2 and A_l for some $l > 2$ affects only the cost of screening $\xi\mu(B_1 \setminus A_1) = \xi(1 - \mu(A_1) - \mu(A_2))$. Thus, the costs of screening decline by maximizing $\mu(A_2)$.

Lemma A.22 *$\mu(A_2) = \max_{2 \leq l \leq L} \mu(A_l)$.*

Proof. Suppose not; let $l \in \{3, \dots, L\}$ be such that $\mu(A_l) > \mu(A_2)$. Consider an organization $\hat{O}$ equal to O except that $\hat{A}_2 = A_l$, $\hat{A}_l = A_2$ and $\hat{B}_1 \setminus \hat{A}_1 = \Omega \setminus (A_1 \cup \hat{A}_2)$. We then have that $\hat{\alpha}_i = hF(\hat{A}_i)$ for each $i \geq 2$ by Lemma A.20. Clearly, whenever $J \subseteq L$ is such that $2, l \in J$, $\sum_{j \in J} F(\hat{A}_j) = \sum_{j \in J} F(A_j)$ and $\sum_{j \in J} F(\hat{A}_j)\mu(\hat{A}_j) = \sum_{j \in J} F(A_j)\mu(A_j)$. Then $\hat{\gamma} = 1 + h\sum_{j=2}^{L} F(A_j) = \gamma$ and

$$\begin{aligned}\hat{\theta} \ &= \sum_{j \in L} F(A_j) - c\mu(A_1) - \sum_{j=2}^{L} chF(A_j)\mu(A_j) - \xi(1 - \mu(A_1) - \mu(\hat{A}_2)) \\ &> \sum_{j \in L} F(A_j) - c\mu(A_1) - \sum_{j=2}^{L} chF(A_j)\mu(A_j) - \xi(1 - \mu(A_1) - \mu(A_2)) = \theta.\end{aligned}$$

Thus, $\hat{y} > y$, a contradiction to the optimality of O. ■

We are heading towards the characterization of the order of the sets in $\mathcal{C}$. Lemma A.23 shows that the cost c_{A_2} of A_2 is no greater than the cost $c_{A_l \cap (B_1 \setminus A_1)}$ of $A_l \cap (B_1 \setminus A_1)$ for each $l > 2$, which implies that there is an optimal organization in which $A_l = A_l \cap (B_1 \setminus A_1) < A_2$. Later on we will strengthen this conclusion by showing that $A_l < A_2$ in any optimal organization.

Lemma A.23 $\alpha_2 c \leq \alpha_l c + \xi$ *for each* $l > 2$.

Proof. Let $l \in L \setminus \{1, 2\}$ and suppose that $\alpha_2 c > \alpha_l c + \xi$. Then $A_2 < A_l \cap (B_1 \setminus A_1)$ by Theorem 3 since $A_2 \in \mathcal{C}$ and $A_l \cap (B_1 \setminus A_1) \in \mathcal{C}$. Since $A_l \subseteq B_1 \setminus A_1$ by Lemma A.18, it follows that $A_2 < A_l$.

Let $a, a' \in \Omega$ and $\varepsilon > 0$ be such that $a + \varepsilon < a'$, $[a, a + \varepsilon) \subseteq A_2$ and $[a', a' + \varepsilon) \subseteq A_l$. Consider an organization $\hat{O}$ equal to O except that $\hat{A}_2 = (A_2 \setminus [a, a + \varepsilon)) \cup [a', a' + \varepsilon)$, $\hat{A}_l = (A_l \setminus [a', a' + \varepsilon)) \cup [a, a + \varepsilon)$ and $\hat{B}_1 \setminus \hat{A}_1 = \Omega \setminus (A_1 \cup \hat{A}_2)$. Let $\varepsilon' = F(A_2) - F(\hat{A}_2) = F([a, a + \varepsilon)) - F([a', a' + \varepsilon)) > 0$, where the inequality follows by Lemma A.12. Then $F(\hat{A}_2) < F(A_2)$, $F(\hat{A}_2) + F(\hat{A}_l) = F(A_2) + F(A_l)$ and $F(\hat{A}_l) > F(A_l)$. Furthermore, $\mu(\hat{A}_2) = \mu(A_2)$ and $\mu(\hat{A}_l) = \mu(A_l)$.

We have that $\hat{\alpha}_2 = hF(\hat{A}_2) = \alpha_2 - h\varepsilon'$ and $\hat{\alpha}_l = \alpha_l + h\varepsilon'$. As $\varepsilon' \to 0$ when $\varepsilon \to 0$, pick $\varepsilon > 0$ such that $\hat{\alpha}_2 c > \hat{\alpha}_l c + \xi$. Since $\hat{A}_2 < \hat{A}_l \cap (\hat{B}_1 \setminus \hat{A}_1)$ does not hold, it follows by Theorem 3 that $\hat{O}$ is not optimal and, therefore, $\hat{y} < y$ since $\hat{L} = L$ and $\hat{l}_i = l_i$ for each $i \in L$. But $\hat{\gamma} = 1 + h\sum_{i=2}^{L} F(A_i) = \gamma$ and, due to $\mu(A_2) = \max_{2 \leq l \leq L} \mu(A_l)$ by Lemma A.22, $\hat{\theta} = \theta + ch\varepsilon'(\mu(A_2) - \mu(A_l)) \geq \theta$, it follows that $\hat{y} \geq y$, a contradiction. This contradiction shows that $\alpha_2 c \leq \alpha_l c + \xi$. ■

The knowledge set A_1 of layer 1 contains the most frequent problems by Corollary 2. Lemma A.24, which is a direct consequence of Theorem 3, adds to this by showing that the problems that are screened by layer 1 but not solved by any layer other than layer 2 are the least frequent problems (if any).

Lemma A.24 *$A_1 < A_l < (B_1 \setminus A_1) \cap (\cap_{j>2} A_j^c)$ for each $l > 2$.*

Proof. Let $l > 2$, $C = (B_1 \setminus A_1) \cap (\cap_{j>2} A_j^c)$ for convenience and note that, by Lemma A.18, $A_l = A_l \cap (B_1 \setminus A_1) \in \mathcal{C}$. Since $\alpha_l > 0$ by Theorem 1, $c_{A_l \cap (B_1 \setminus A_1)} = c\alpha_l + \xi > \xi = c_C$ and it follows by Theorem 3 that $A_l < C$. That $A_1 < A_l$ follows by Corollary 2. ■

The conclusion of Lemma A.25 follows from the first order conditions of the problem of maximizing output.

Lemma A.25 *$F(A_i) \neq F(A_j)$ for each $i, j \in \{3, \ldots, L\}$ such that $i \neq j$.*

Proof. Let $g : \{2, \ldots, L\} \to \{2, \ldots, L\}$ be such that $g(L) = 2$, $F(A_{g(2)}) = \max_{3 \leq l \leq L} F(A_l)$ and

$$F(A_{g(l)}) = \max\{F(A_j) : j \in \{3, \ldots, L\} \setminus \{g(2), \ldots, g(l-1)\}\}.$$

Then g is a bijection, $g(\{2, \ldots, L-1\}) = \{3, \ldots, L\}$ and $F(A_{g(2)}) \geq \cdots \geq F(A_{g(L-1)})$. Thus, by Lemma A.20,

$$c_{A_{g(2)} \cap (B_1 \setminus A_1)} \geq \cdots \geq c_{A_{g(L-1)} \cap (B_1 \setminus A_1)}.$$

Since $c_{A_{g(L-1)} \cap (B_1 \setminus A_1)} \geq c_{A_2}$ by Lemma A.23, it follows by Theorem 3 that there is an optimal organization $\hat{O}$ such that $\hat{A}_{g(2)} < \cdots < \hat{A}_{g(L-1)} < \hat{A}_{g(L)}$ and $F(\hat{C}) = F(C)$ for each $C \in \mathcal{C}$, thus, in particular, $F(\hat{A}_l) = F(A_l)$ for all $l \in L$; in addition, $\hat{B}_1 \setminus \hat{A}_1 = (\hat{A}_1 \cup \hat{A}_2)^c$ since $F((\hat{B}_1 \setminus \hat{A}_1) \cup \hat{A}_1 \cup \hat{A}_2) = F((B_1 \setminus A_1) \cup A_1 \cup A_2) = 0$ and $(\hat{B}_1 \setminus \hat{A}_1) \cup \hat{A}_1 \cup \hat{A}_2 \in \mathcal{I}$ imply that $(\hat{B}_1 \setminus \hat{A}_1) \cup \hat{A}_1 \cup \hat{A}_2 = \emptyset$. Thus, $\hat{A}_{g(L-1)} < (\hat{B}_1 \setminus \hat{A}_1) \cap (\cap_{j>2} \hat{A}_j^c)$ by Lemma A.24.

Fix $l \in \{2, \ldots, L-2\}$. By Theorem 2, there exists $a, m, b \in \Omega$ such that $a < m < b$, $\hat{A}_{g(l)} = [a, m)$ and $\hat{A}_{g(l+1)} = [m, b)$. Consider $\varepsilon < 0$ and an organization $\tilde{O}$ equal to $\hat{O}$ except that $\tilde{A}_{g(l)} = [a, m+\varepsilon)$ and $\tilde{A}_{g(l+1)} = [m+\varepsilon, b)$. We then have that

$$\tilde{\alpha}_{g(l)}(\varepsilon) = \hat{\alpha}_{g(l)} - hF([m+\varepsilon, m)), \quad \tilde{\alpha}'_{g(l)}(0) = hf(m),$$

$$\tilde{\alpha}_{g(l+1)}(\varepsilon) = \hat{\alpha}_{g(l)} + hF([m+\varepsilon, m)), \quad \tilde{\alpha}'_{g(l+1)}(0) = -hf(m),$$

$$\tilde{\theta}(\varepsilon) = \hat{\theta} + c\tilde{\alpha}_{g(l+1)}(\varepsilon)\varepsilon - c\tilde{\alpha}_{g(l)}(\varepsilon)\varepsilon - \sum_{i=l}^{l+1} (\tilde{\alpha}_{g(i)}(\varepsilon) - \hat{\alpha}_{g(i)}) c\mu(\hat{A}_{g(i)}) \text{ and}$$

$$\hat{\gamma}\tilde{y}'(0) = c\hat{\alpha}_{g(l+1)} - c\hat{\alpha}_{g(l)} + chf(m)(\mu(\hat{A}_{g(l+1)}) - \mu(\hat{A}_{g(l)})) \geq 0,$$

the latter because $\varepsilon < 0$ and $\tilde{y}(\varepsilon) \leq \hat{y}$ since $\hat{O}$ is optimal.

Considering now $\varepsilon > 0$, we obtain that

$$c\hat{\alpha}_{g(l+1)} - c\hat{\alpha}_{g(l)} + chf(m)(\mu(\hat{A}_{g(l+1)}) - \mu(\hat{A}_{g(l)})) \leq 0.$$

Since $\hat{\alpha}_{g(i)} = hF(\hat{A}_{g(i)})$, it follows that

$$F(\hat{A}_{g(l)}) - F(\hat{A}_{g(l+1)}) = f(m)(\mu(\hat{A}_{g(l+1)}) - \mu(\hat{A}_{g(l)})).$$

We have that $F(\hat{A}_{g(l)}) \geq F(\hat{A}_{g(l+1)})$, which implies that $\mu(\hat{A}_{g(l+1)}) \geq \mu(\hat{A}_{g(l)})$; if $\mu(\hat{A}_{g(l+1)}) = \mu(\hat{A}_{g(l)})$, then $F(\hat{A}_{g(l)}) = F(\hat{A}_{g(l+1)})$ which, together with $\hat{A}_{g(l)} < \hat{A}_{g(l+1)}$, contradicts Lemma A.12. This contradiction shows that $\mu(\hat{A}_{g(l+1)}) > \mu(\hat{A}_{g(l)})$ and $F(\hat{A}_{g(l)}) > F(\hat{A}_{g(l+1)})$.

Thus, $F(\hat{A}_l) = F(A_l)$ for all $l \in L$, together with the above conclusion, implies that $F(A_{g(2)}) > \cdots > F(A_{g(L-1)})$ from which the conclusion of the lemma follows. ■

Let $g(3), \ldots, g(L) \in \{3, \ldots, L\}$ be such that $\alpha_{g(3)} > \cdots > \alpha_{g(L)}$; such bijection $g : \{3, \ldots, L\} \to \{3, \ldots, L\}$ exists by Lemma A.25. Thus, $A_{g(3)} < \ldots < A_{g(L)}$. In addition, as shown in the proof of Lemma A.25, for each $l \in \{3, \ldots, L-1\}$,

$$F(A_{g(l)}) - F(A_{g(l+1)}) = f(b_l)(\mu(A_{g(l+1)}) - \mu(A_{g(l)}))$$

for some $b_l \in \Omega$. Since $F(A_{g(l)}) > F(A_{g(l+1)})$, it follows that $\mu(A_{g(l+1)}) > \mu(A_{g(l)})$.

Lemma A.26 strengthens Lemma A.23 by showing that layer 2 solves the least frequent problems. It also establishes an upper bound on the probability of A_2.

Lemma A.26 *$A_{g(L)} < A_2$ and $F(A_2) < F(A_{g(L)}) + \frac{\xi}{ch}$.*

Proof. We have that $c\alpha_2 \leq c\alpha_{g(L)} + \xi$ by Lemma A.23. Assume first that $c\alpha_2 = c\alpha_{g(L)} + \xi$. Then, by Theorem 1, $c\alpha_2 = c\alpha_{g(L)} + \xi > \xi$ and $A_2 < (B_1 \setminus A_1) \cap (\cap_{l>2} A_l^c)$ by Theorem 3. Theorem 3 also implies that there is an optimal organization $\hat{O}$ such that $\hat{A}_{g(L)} < \hat{A}_2$, $\hat{\mathcal{C}} = \{\hat{C} : C \in \mathcal{C}\}$ and $F(\hat{C}) = F(C)$ for all $C \in \mathcal{C}$. Thus, $\hat{B}_1 \setminus \hat{A}_1 = (\hat{A}_1 \cup \hat{A}_2)^c$ as in the proof of Lemma A.25, $F(\hat{A}_l) = F(A_l)$ and $\hat{\alpha}_l = \alpha_l$ for all $l \in L$.

Thus, it follows by the above and by Theorem 2 that there exists $a, m, b \in \Omega$ such that $a < m < b$, $\hat{A}_{g(L)} = [a, m)$ and $\hat{A}_2 = [m, b)$. Arguing as in the proof of Lemma A.25, we obtain that

$$c\alpha_{g(L)} + \xi - c\alpha_2 = chf(m)(\mu(\hat{A}_2) - \mu(\hat{A}_{g(L)})).$$

Since $c\alpha_{g(L)}+\xi = c\alpha_2$, it follows that $\mu(\hat{A}_2) = \mu(\hat{A}_{g(L)})$ and $F(\hat{A}_{g(L)})-F(\hat{A}_2) = -\frac{\xi}{ch} < 0$, a contradiction to Lemma A.12 since $\hat{A}_{g(L)} < A_2$. This contradiction shows that $c\alpha_{g(L)}+\xi > c\alpha_2$. Hence, $F(A_2) < F(A_{g(L)}) + \frac{\xi}{ch}$ and, by Theorem 3, $A_{g(L)} < A_2$. ■

From now onwards, we assume that O *is an* η*-optimal organization.* Lemma A.27 formalizes the claim that the function $y^*_{\bar{L}}$ gives the output of the optimal organization.

Lemma A.27 $y = y^*_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}})$ *where* $(\mu_1, \ldots, \mu_{\bar{L}}) \in X_{\bar{L}}$ *is such that* $\mu_1 = \mu(A_1)$, $\mu_2 = \mu(A_2)$, $\mu_l = \mu(A_{g(l)})$ *for each* $2 < l \leq L$, *and* $\mu_l = 0$ *for each* $L < l \leq \bar{L}$.

Proof. We have that $y \leq 1$. Hence, if $L > \bar{L}$, then $L \geq \frac{1+\eta}{\eta} \Leftrightarrow L - 1 \geq \frac{1}{\eta}$ and $Y = y - \eta(L - 1) \leq 1 - 1 = 0$, a contradiction. Thus, $L \leq \bar{L}$.

Let $C = (B_1 \setminus A_1) \cap (\cap_{j>2} A^c_j)$ for convenience. We have $A_1 < A_{g(3)} < \cdots < A_{g(L)}$, $A_{g(L)} < A_2$ and $A_{g(L)} < C$, in addition to $\cup^L_{l=1} B_l = \Omega$. Let $\mu_1 = \mu(A_1)$, $\mu_2 = \mu(A_2)$, $\mu_i = \mu(A_{g(i)})$ for each $2 < i \leq L$ and $\mu_i = 0$ for each $L < i \leq \bar{L}$. Then $y = y_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}})$ if $A_2 < C$ and $y = \tilde{y}_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}})$ if $C < A_2$. ■

Lemma A.28 establishes a lower bound on the size $\mu(A_2) = \max_{2\leq l\leq \bar{L}} \mu_l$ of the knowledge set of layer 2. It is a consequence of Lemma A.17 and the assumption that $L \geq 2$ in the optimal organization.

Lemma A.28 $\max_{2\leq l\leq \bar{L}} \mu_l \geq \frac{\eta^2}{f(0)}$.

Proof. Let, by Lemma A.27, $(\mu_1, \ldots, \mu_{\bar{L}}) \in X_{\bar{L}}$ be such that $y = y^*_{\bar{L}}(\mu_1, \ldots, \mu_{\bar{L}})$. If $\max_{2\leq l\leq \bar{L}} \mu_l < \frac{\eta^2}{f(0)}$, then $y^*_{\bar{L}}(\mu_1, \ldots, \mu_L) < y_1 + \eta$ by Lemma A.17. But, as $L \geq 2$, it follows that $Y = y - (L - 1)\eta \leq y - \eta < y_1$, a contradiction to the optimality of O. ■

Lemma A.29 concludes the characterization of the order of the set in $\mathcal{C}$ by showing that the problems that are screened by layer 1 but not solved by layers 1 and 2 are the least frequent problems (if any). It uses the lower bound on μ_2 from Lemma A.28 to show that $c_{A_2} > c_{(B_1\setminus A_1)\cap(\cap_{l>2} A^c_l)}$, from which the conclusion follows by Theorem 3.

Lemma A.29 $A_2 < (B_1 \setminus A_1) \cap (\cap_{l>2} A^c_l)$.

Proof. It follows by Lemmas A.22 and A.28 that $\mu_2 = \max_{2\leq l\leq L} \mu_l = \max_{2\leq \mu\leq \bar{L}} \mu_l \geq \frac{\eta^2}{f(0)}$. Hence, by Lemma A.20,

$$c_{A_2} = c\alpha_2 = chF(A_2) \geq chf(\bar{\omega})\mu_2 \geq \frac{chf(\bar{\omega})\eta^2}{f(0)} = \xi_1 > \xi = c_{(B_1\setminus A_1)\cap(\cap_{l>2} A^c_l)}.$$

Therefore, the conclusion follows from Theorem 3. ■

Lemma A.30 shows that the optimal organization is pyramidal in the sense of having more workers than managers.

Lemma A.30 *$\beta_1 > \sum_{l=2}^{L} \beta_l$ if $h < 1$.*

Proof. We have that $\sum_{l=2}^{L} \alpha_l = hF(\cup_{l=2}^{L} A_l) \leq h(1 - F(A_1)) < 1$. Hence, it follows that $\beta_1 = \frac{1}{\gamma} > \frac{\sum_{l=2}^{L} \alpha_l}{\gamma} = \sum_{l=2}^{L} \beta_l$. ■

Lemma A.31 concludes the proof of part 5 of the theorem.

Lemma A.31 *$F(A_{g(L)}) > \eta$ and $\mu(A_2) > \mu(A_{g(L)})$.*

Proof. Consider an organization $\hat{O}$ equal to O except that $\hat{L} = L \backslash \{g(L)\}$ and $\hat{\beta}_l = \frac{\hat{\alpha}_l}{\hat{\gamma}}$ for each $l \in \hat{L}$. We have that $\hat{\alpha}_l = hF(A_l) = \alpha_l$ for each $l \in \hat{L}$, $\hat{\gamma} = 1 + \sum_{l \in L \backslash \{g(L)\}} \alpha_l < \gamma$ and $\hat{\theta} = \theta - F(A_{g(L)}) + chF(A_{g(L)})\mu(A_{g(L)})$. Then,

$$\frac{\theta - F(A_{g(L)})}{\gamma} \leq \frac{\hat{\theta}}{\gamma} < \frac{\hat{\theta}}{\hat{\gamma}} \leq \frac{\theta}{\gamma} - \eta,$$

where the last inequality follows because O is η-optimal. Hence, $F(A_{g(L)}) > \gamma\eta > \eta$ since $\gamma > 1$.

It follows by Lemmas A.26 and A.29 and by Theorem 2 that there exists $a, m, b \in \Omega$ such that $a < m < b$, $A_{g(L)} = [a, m)$ and $A_2 = [m, b)$. Arguing as in the proof of Lemma A.25, we obtain that

$$c\alpha_{g(L)} + \xi - c\alpha_2 = chf(m)(\mu(A_2) - \mu(A_{g(L)})).$$

Since $c\alpha_{g(L)} + \xi = chF(A_{g(L)}) + \xi > chF(A_2) = c\alpha_2$, it follows that $\mu(A_2) > \mu(A_{g(L)})$. ■

To conclude the proof of Theorem 5, recall from Lemma A.21 that

$$y = \frac{\sum_{i=1}^{L} F(A_i) - c\mu(A_1) - ch\sum_{i=2}^{L} F(A_i)\mu(A_i) - \xi(\bar{\omega} - \mu(A_1) - \mu(A_2))}{1 + h\sum_{i=2}^{L} F(A_i)}.$$

Let $\hat{g} : \{3, \ldots, L\} \to \{3, \ldots, L\}$ be a bijection and consider an organization $\hat{O}$ equal to O except that $\hat{A}_{\hat{g}(l)} = A_{g(l)}$ for each $3 \leq l \leq L$. Then $\hat{y} = y$ and $\hat{O}$ is an optimal organization.

A.13 Proof of Theorem 6

Part 1: Let O be an optimal organization and L be the number of its layers. Suppose, in order to reach a contradiction, that $F(\cup_{l \in L} A_l) < 1$. Then there is $a, b \in \Omega$ such that

$[a,b) \subseteq \left(\cup_{l=1}^{L} A_l\right)^c$ and $a < b$. Let $0 < \varepsilon < b - a$ and consider the organization $\hat{O}$ with L layers equal to O except that $\hat{B}_L = \hat{A}_L = A_L \cup [a, a+\varepsilon)$. Letting $y(\varepsilon)$ be the resulting output, it follows that

$$y(\varepsilon) = \frac{\theta + F([a, a+\varepsilon)) - \alpha_L(\varepsilon)c\varepsilon}{\gamma + \alpha_L(\varepsilon) - \alpha_L},$$

where, by Lemma A.11,

$$\alpha_L(\varepsilon) = hF(A_L) + hF([a, a+\varepsilon)) + \pi(1 - F(A_L \cup (\cup_{j<L} B_j) \cup [a, a+\varepsilon))).$$

We have that $\alpha_L(\varepsilon) \leq \alpha_L + hF([a, a+\varepsilon))$ and, by defining

$$\hat{y}(\varepsilon) = \frac{\theta + F([a, a+\varepsilon)) - \alpha_L c\varepsilon - chF([a, a+\varepsilon))\varepsilon}{\gamma + hF([a, a+\varepsilon))},$$

we obtain that $y(\varepsilon) \geq \hat{y}(\varepsilon)$.

We have that

$$\gamma \hat{y}'(0) = f(a) - \alpha_L c - hyf(a).$$

Since $f(0) \geq f(a) \geq f(\bar{\omega}) > 0$ and $\alpha_L \leq h[F(A_L) + 1 - F(A_L \cup (\cup_{j<L} B_j))] \leq h$, it follows that $\gamma \hat{y}'(0) \geq f(\bar{\omega}) - h(c + f(0))$ and that $\hat{y}'(0) > 0$ whenever h is sufficiently small. Therefore, there exists $\varepsilon > 0$ such that $y(\varepsilon) \geq \hat{y}(\varepsilon) > \hat{y}(0) = y$. But this contradicts the optimality of O and shows that $F(\cup_{l \in L} A_l) = 1$.

Part 2: Let O be an optimal organization and L be the number of its layers. Note first that $L > 1$ if h is sufficiently small; indeed, if $L = 1$, then output is $y_1 < 1$ whereas the output of a hierarchy $\tilde{O}$ with two layers and $(\tilde{\mu}_1, \tilde{\mu}_2) = (0, \bar{\omega})$ is $\frac{1-ch\bar{\omega}}{1+h}$; since $\frac{1-ch\bar{\omega}}{1+h} > y_1$ whenever h is sufficiently small, this shows that $L > 1$.

Suppose, in order to reach a contradiction, that $B_1 \neq \emptyset$. Then there is $a, b \in \Omega$ such that $[a,b) \subseteq B_1$ and $a < b$. By adjusting b if necessary, we may assume that (i) $[a,b) \subseteq A_1$ or (ii) $[a,b) \subseteq B_1 \setminus A_1$; the former is possible if $a \in A_1$ and the latter if $a \in B_1 \setminus A_1$. Let $0 < \varepsilon < b - a$ and consider the organization $\hat{O}$ with L layers equal to O except that $\hat{A}_2 = A_2 \cup [a, a+\varepsilon)$, $\hat{B}_2 = B_2 \cup [a, a+\varepsilon)$, $\hat{A}_1 = A_1 \setminus [a, a+\varepsilon)$ and $\hat{B}_1 = B_1 \setminus [a, a+\varepsilon)$ in case (i) and $\hat{B}_2 = B_2 \cup [a, a+\varepsilon)$ and $\hat{B}_1 = B_1 \setminus [a, a+\varepsilon)$ in case (ii). Letting $y(\varepsilon)$ be the resulting output, it follows that

$$y(\varepsilon) = \frac{F(\cup_{l \in L} A_l) - \sum_{l \in L} \alpha_l(\varepsilon)\nu_l(\varepsilon)}{\sum_{l \in L} \alpha_l(\varepsilon)}$$

and

$$\gamma y'(0) = -\sum_{l \in L} \alpha_l'(0)\nu_l - \sum_{l \in L} \alpha_l \nu_l'(0) - y \sum_{l \in L} \alpha_l'(0).$$

Using Lemma A.11, we have that

$$\alpha_l(\varepsilon) = hF(\hat{A}_l) + \pi(1 - F(\hat{A}_l \cup (\cup_{j<l}\hat{B}_j))) \text{ for each } l > 1, \text{ and}$$
$$\nu_l(\varepsilon) = c\mu(\hat{A}_l) + \xi\mu(\hat{B}_l \setminus \hat{A}_l) \text{ for each } l \in L.$$

Then $\alpha_l'(0) = \nu_l'(0) = 0$ for each $l > 2$ and $\alpha_1'(0) = 0$ since $\alpha_1(\varepsilon) = 1$.

Furthermore, $\alpha_2(\varepsilon) = hF(\hat{A}_2) + \pi(1 - F(\hat{A}_2) - F(\hat{B}_1))$ since $\hat{B}_1 \cap \hat{A}_2 = \emptyset$ and, hence, $\alpha_2'(0) \leq hf(a)$; indeed, $\alpha_2(\varepsilon) = hF(A_2) + hF([a, a+\varepsilon)) + \pi(1 - F(A_2) - F(B_1))$ and $\alpha_2'(0) = hf(a)$ in case (i), and $\alpha_2(\varepsilon) = hF(A_2) + \pi(1 - F(A_2) - F(B_1) + F([a, a+\varepsilon)))$ and $\alpha_2'(0) = \pi f(a) \leq hf(a)$ in case (ii).

We also have that $\nu_1(\varepsilon) = \nu_1 - c\varepsilon$ and $\nu_1'(0) = -c$ in case (i), and $\nu_1(\varepsilon) = \nu_1 - \xi\varepsilon$ and $\nu_1'(0) = -\xi$ in case (ii); thus, $-\nu_1'(0) \geq \xi$. Analogously, $\nu_2'(0) \leq c$. Using $y \leq 1$, $f(a) \leq f(0)$, $\nu_2 \leq c\bar{\omega}$ and $\alpha_2 \leq h$, it then follows that

$$\gamma y'(0) = -(\nu_2 + y)\alpha_2'(0) - \nu_1'(0) - \nu_2'(0)\alpha_2 \geq \xi - h(f(0)(c\bar{\omega} + 1) + c).$$

Thus, $y'(0) > 0$ whenever h is sufficiently small. Therefore, there exists $\varepsilon > 0$ such that $y(\varepsilon) > y(0) = y$. But this contradicts the optimality of O and shows that $B_1 = \emptyset$.

Part 3: We have that $y \leq 1$ in any organization, hence $y - (L-1)\eta \leq 0$ whenever $L \geq \frac{1+\eta}{\eta}$. Let y^* be the output of O_2^*. Since $1 - \eta > y_1$, $1 - \eta > 1 - (L-1)\eta$ for each $2 < L < \frac{1+\eta}{\eta}$ and $y_2^* = \frac{1 - ch\bar{\omega}}{1+h} \to 1$ as $h \to 0$, it follows that there is $\bar{h} > 0$ such that $y_2^* - \eta > y - (L-1)\eta$ for each organization with output y and $L \neq 2$ layers and such that parts 1 and 2 of the theorem hold. Parts 1 and 2 of the theorem then imply that $y_2^* - \eta > y - \eta$ for each organization with output y and $L = 2$ layers.

A.14 Proof of Theorem 7

We use the following notion in the proof of Theorem 7. Letting $\mu_j = \mu(A_j)$ for each $j \in L$, we then have that optimal hierarchies (i.e. optimal organizations in the $\xi = c$ and $\pi = h$ case) satisfy $A_j = \left[\sum_{i=1}^{j-1} \mu_i, \sum_{i=1}^{j} \mu_i\right)$ for each $1 \leq j \leq L$ and $\mu_i > 0$ for each $i > 1$ (see Lemma A.16 for the latter). The vector $(\mu_1, \ldots, \mu_L)$ is obtained by solving

$$\max_{(\mu_1,\ldots,\mu_L)\in\mathbb{R}_+^L} \frac{F(\sum_{i=1}^{L}\mu_i) - c\mu_1 - ch\sum_{i=2}^{L}\mu_i(1 - F(\sum_{l=1}^{i-1}\mu_l))}{1 + h\sum_{i=2}^{L}(1 - F(\sum_{l=1}^{i-1}\mu_l))} \tag{28}$$

$$\text{subject to } \sum_{i=1}^{L}\mu_i \leq \bar{\omega}. \tag{29}$$

For each $L \in \mathbb{N}$, let $\mu_L = (\mu_{L,1}, \ldots, \mu_{L,L})$ be a solution to (28)–(29), let $y_L(\mu_1, \ldots, \mu_L)$ be the value of the objective function at $(\mu_1, \ldots, \mu_L)$ and let $y_L = \frac{\theta_L}{\gamma_L}$ be the value of the objective function at μ_L, i.e. $y_L = y_L(\mu_L)$.

Part 1: The conclusion follows from the following claim.

Claim 1 *A lexicographically optimal organization exists if (A1), $\xi = c$, $\pi = h$ and $f(0) > \min\{ch, c\}$ hold.*

Proof. We start by showing that if $f(0) > \min\{ch, c\}$ then $\sup_{L \in \mathbb{N}} y_L > 0$. Consider first the case where $\min\{ch, c\} = ch$; in this case, consider the hierarchy with $L = 2$ with $A_1 = \emptyset$ and $A_2 = [0, \mu_2)$. Then $y(\mu_2) = \frac{F(\mu_2) - ch\mu_2}{1+h}$ and $\frac{dy(0)}{d\mu_2} = \frac{f(0)-ch}{1+h} > 0$, implying that there exists $\mu_2^* > 0$ such that $y(\mu_2^*) > 0$. Thus, $y_2 > 0$. In the case $\min\{ch, c\} = c$ consider the hierarchy with $L = 1$ with $A_1 = [0, \mu_1)$. Then $y(\mu_1) = F(\mu_1) - c\mu_1$ and $\frac{dy(0)}{d\mu_1} = f(0) - c > 0$, implying that there exists $\mu_1^* > 0$ such that $y(\mu_1^*) > 0$. Thus, $y_1 > 0$.

Recall that, for each $L \in \mathbb{N}$, $\mu_L = (\mu_{L,1}, \ldots, \mu_{L,L})$ is a solution to (28)–(29), $y_L(\mu)$ is the value of the objective function at μ, $y_L = y_L(\mu_L)$, θ_L is the numerator of y_L and γ_L is the denominator of y_L.

Consider first the case where $\max_{L \in \mathbb{N}} y_L$ exists. In this case, let L^* be the smallest L' such that $y_{L'} = \max_{L \in \mathbb{N}} y_L > 0$. Then the hierarchy O with $L = L^*$ and $A_i = [\sum_{j=1}^{i-1} \mu_{L^*,j}, \sum_{j=1}^{i} \mu_{L^*,j})$ for each $i \in L^*$ is an optimal organization. If not, then there exists a hierarchy $\hat{O}$ such that (i) $\hat{y} > y_{L^*}$ or (ii) $\hat{y} = y_{L^*}$ and $\hat{L} < L^*$ or (iii) $\hat{y} = y_{L^*}$, $\hat{L} = L^*$, $\hat{l}_i \subseteq l_i$ for all $i \in L^*$ and $\hat{l}_j \neq l_j$ for some $j \in L^*$. Case (iii) cannot hold because $l_1 = L^* = \hat{L} = \hat{l}_1$ and $l_j = \{j\} = \hat{l}_j$ for each $j \in L^* \setminus \{1\}$. Since $\hat{O}$ is a hierarchy, $\hat{y} \leq y_{\hat{L}} \leq y_{L^*}$, hence (i) does not hold. Thus, (ii) must hold. But since $L^* = \min\{L' : y_{L'} = \max_{L \in \mathbb{N}} y_L\}$ and $\hat{L} < L^*$, we get $\hat{y} \leq y_{\hat{L}} < y_{L^*}$, a contradiction.

Due to the above, we may now assume that $y_L < \sup_{L' \in \mathbb{N}} y_{L'}$ for each $L \in \mathbb{N}$. Let $\{L_k\}_{k=1}^{\infty}$ be such that $L_1 = 1$, $L_k \geq \max\{k, L_{k-1} + 1\}$ and $y_L < y_{L_k}$ for all $L < L_k$.[38] For convenience, let $y_k = y_{L_k}$, $\theta_k = \theta_{L_k}$, $\gamma_k = \gamma_{L_k}$ and $\mu_k = \mu_{L_k}$ for each $k \in \mathbb{N}$. Then $y_k \uparrow \sup_{L \in \mathbb{N}} y_L$. Furthermore, $\mu_{k,i} > 0$ for all $1 < i \leq L_k$ and $k \in \mathbb{N}$. If not, let $\hat{O}$ be the hierarchy obtained from the hierarchy defined by μ_k by removing the layers $i \in \{2, \ldots, L_k\}$

[38] In more detail, let $L_1 = 1$; assuming that $L_1, \ldots, L_k$ have been defined such that $L_k \geq k$, let L_{k+1} be the smallest $L \geq L_k + 1$ such that $y_L > y_{L_k}$.

with $\mu_{k,i} = 0$, i.e. let $\hat{L} = \{1\} \cup \{l \in L_k \setminus \{1\} : \mu_{k,l} > 0\}$. Then $\hat{L} < L_k$ and $y_{L_k} \le \hat{y}$.[39] But then $y_{L_k} \le \hat{y} \le y_{\hat{L}}$ and $\hat{L} < L_k$, which is a contradiction to $y_L < y_{L_k}$ for all $L < L_k$.

We have that $F(\sum_{i=1}^{L_k-1} \mu_{k,i}) \to 1$. Indeed, if $F(\sum_{i=1}^{L_k-1} \mu_{k,i}) \not\to 1$, then, taking a subsequence if necessary, there is $\eta > 0$ such that $F(\sum_{i=1}^{L_k-1} \mu_{k,i}) \le 1 - \eta$ for each k. Hence, $1 - F(\sum_{l=1}^{i-1} \mu_{k,l}) \ge 1 - F(\sum_{l=1}^{L_k-1} \mu_{k,l}) \ge \eta$ for each $2 \le i \le L_k$ and, therefore, $0 < y_k \le \frac{1}{1+(L_k-1)h\eta} \to 0$. But this contradicts $y_k \to \sup_{L \in \mathbb{N}} y_L > 0$.

Next we claim that there is $K > 2$ such that $\sum_{i=1}^{L_k} \mu_{k,i} = \bar{\omega}$ for each $k \ge K$. Suppose not; then, taking a subsequence if necessary, $\sum_{i=1}^{L_k} \mu_{k,i} < \bar{\omega}$ for all k. For each k, since $\mu_{k,L_k} > 0$, $\frac{\partial y(\mu_k)}{\partial \mu_{L_k}} = 0$ and, thus

$$f(\sum_{i=1}^{L_k} \mu_{k,i}) = ch(1 - F(\sum_{i=1}^{L_k-1} \mu_{k,i})). \tag{30}$$

Since f is strictly decreasing, $f(\bar{\omega}) > 0$ and $F(\sum_{i=1}^{L_k-1} \mu_{k,i}) \to 1$, we have that

$$f(\sum_{i=1}^{L_k} \mu_{k,i}) > f(\bar{\omega}) > ch(1 - F(\sum_{i=1}^{L_k-1} \mu_{k,i}))$$

for all k sufficiently large, contradicting (30).

For each $k \in \mathbb{N}$ and $1 \le i \le L_k$, let $A_{k,i} = [\sum_{l=1}^{i-1} \mu_{k,i}, \sum_{l=1}^{i} \mu_{k,i})$. We have that $F(A_{k,L_k-1}) \to 0$. Indeed, for each $k \ge K$, $\sum_{l=1}^{L_k} F(A_{k,l}) = F(\cup_{l=1}^{L_k} A_{k,l}) = 1$ (recall that the sets in $\{A_{k,l}\}_{l=1}^{L_k}$ are pairwise disjoint) and, thus,

$$\begin{aligned} \gamma_k &= 1 + h \sum_{i=2}^{L_k} (1 - F(\cup_{l=1}^{i-1} A_{k,l})) = 1 + h \sum_{i=2}^{L_k} \sum_{l=i}^{L_k} F(A_{k,l}) \\ &= 1 + h \sum_{i=2}^{L_k} (i-1) F(A_{k,i}). \end{aligned}$$

Since

$$y_k = \frac{\theta_k}{\gamma_k} \le \frac{1}{1 + h \sum_{i=2}^{L_k} (i-1) F(A_{k,i})}$$

and $\lim_k y_k > 0$, it follows that there is $B > 0$ such that $\sum_{i=2}^{L_k} (i-1) F(A_{k,i}) \le B$ for each $k \ge K$. In particular, $(L_k - 2) F(A_{k,L_k-1}) \le B$ for each $k \ge K$ and, as $L_k \to \infty$, it follows that $F(A_{k,L_k-1}) \to 0$.

[39] For the latter, note first that $\sum_{l \in L_k : l \le i-1} \mu_{k,l} = \sum_{l \in \hat{L} : l \le i-1} \mu_{k,l}$ for each $2 \le i \le L_k + 1$. Hence, $F(\sum_{i=1}^{L_k} \mu_{k,i}) = F(\sum_{l \in \hat{L}} \mu_{k,i})$ and $\sum_{i=2}^{L_k} \mu_{k,i}(1 - F(\sum_{l=1}^{i-1} \mu_{k,l})) = \sum_{i \in \hat{L} : i \ge 2} \mu_{k,i}(1 - F(\sum_{l \in \hat{L} : l \le i-1} \mu_{k,l}))$, implying that $\theta_k = \hat{\theta}$. Furthermore, $\hat{\gamma} = 1 + h \sum_{i \in \hat{L} : i \ge 2} (1 - F(\sum_{l \in \hat{L} : l \le i-1} \mu_{k,l})) = 1 + h \sum_{i \in \hat{L} : i \ge 2} (1 - F(\sum_{l=2}^{i-1} \mu_{k,l})) \le 1 + h \sum_{i=2}^{L_k} (1 - F(\sum_{l=2}^{i-1} \mu_{k,l})) = \gamma_k$. Thus, $\hat{y} = \frac{\hat{\theta}}{\hat{\gamma}} \ge \frac{\theta_k}{\gamma_k} = y_k$.

We now conclude the argument. Fix $k \geq K$ and recall that $\mu_{k,i} > 0$ for all $i > 1$. If $\mu_{k,1} > 0$ as well, then μ_k is a local maximizer of

$$\max_{(\mu_1,\ldots,\mu_{L_k})} \frac{1 - c\mu_1 - ch\sum_{i=2}^{L_k}\mu_i(1 - F(\sum_{l=1}^{i-1}\mu_l))}{1 + h\sum_{i=2}^{L_k}(1 - F(\sum_{l=1}^{i-1}\mu_l))}$$
$$\text{subject to } \sum_{i=1}^{L_k}\mu_i = \bar{\omega};$$

if $\mu_{k,1} = 0$, then μ_k is a local maximizer of

$$\max_{(\mu_2,\ldots,\mu_{L_k})} \frac{1 - ch\sum_{i=2}^{L_k}\mu_i(1 - F(\sum_{l=2}^{i-1}\mu_l))}{1 + h\sum_{i=2}^{L_k}(1 - F(\sum_{l=2}^{i-1}\mu_l))}$$
$$\text{subject to } \sum_{i=2}^{L_k}\mu_i = \bar{\omega}.$$

In either case, by Luenberger and Ye (2008, Theorem, p. 327) (note that the regularity assumption of the theorem is trivially satisfied because there is only one constraint), there is λ_k (the Lagrange multiplier of the constraint $\sum_{i=1}^{L_k}\mu_i = \bar{\omega}$) such that

$$\frac{\partial y(\mu_k)}{\partial \mu_{L_k}} - \lambda_k = 0,$$
$$\frac{\partial y(\mu_k)}{\partial \mu_{L_k - 1}} - \lambda_k = 0.$$

The first of these equations gives

$$-\frac{ch(1 - F(\sum_{i=1}^{L_k-1}\mu_{k,i}))}{\gamma_k} - \lambda_k = 0 \Leftrightarrow \lambda_k = -\frac{ch(1 - F(\bar{\omega} - \mu_{k,L_k}))}{\gamma_k}.$$

The second of these equations gives

$$\lambda_k = \frac{\gamma_k\left[-ch(1 - F(\bar{\omega} - \mu_{k,L_k} - \mu_{k,L_k-1})) + ch\mu_{k,L_k}f(\bar{\omega} - \mu_{k,L_k})\right] + \theta_k hf(\bar{\omega} - \mu_{k,L_k})}{\gamma_k^2}.$$

Putting the two together yields

$$\begin{aligned} -ch(1 - F(\bar{\omega} - \mu_{k,L_k})) = \; & -ch(1 - F(\bar{\omega} - \mu_{k,L_k} - \mu_{k,L_k-1})) + ch\mu_{k,L_k}f(\bar{\omega} - \mu_{k,L_k}) \\ & + y_k hf(\bar{\omega} - \mu_{k,L_k}) \end{aligned}$$

and, hence,

$$y_k = c\left(\frac{F(A_{k,L_k-1})}{f(\bar{\omega} - \mu_{k,L_k})} - \mu_{k,L_k}\right).$$

Since $\mu_{k,L_k} > 0$ and $f(\bar{\omega} - \mu_{k,L_k}) > f(\bar{\omega}) > 0$, it follows that

$$0 < y_k < c\frac{F(A_{k,L_k-1})}{f(\bar{\omega})}.$$

Since $F(A_{k,L_k-1}) \to 0$, it then follows that $y_k \to 0$, a contradiction to $y_k \to \sup_{L\in\mathbb{N}} y_L > 0$. This completes the proof of the claim. ■

We turn now to the proof of parts 2 and 3. Let $\bar{h} > 0$ be such that Theorem 6 and part 1 hold for each $0 < h < \bar{h}$.

Lemma A.32 *If O is an optimal hierarchy, then*

$$\frac{c(1 - F(\sum_{i=1}^{L-1} \mu_i))}{f(\bar{\omega})} \leq y_L. \tag{31}$$

Proof. Let O be an optimal hierarchy and L be the number of its layers. Suppose, in order to reach a contradiction, that $\frac{c(1-F(\sum_{i=1}^{L-1} \mu_i))}{f(\bar{\omega})} > y_L$, i.e.

$$ch(1 - F(\sum_{i=1}^{L-1} \mu_i)) > y_L h f(\bar{\omega}). \tag{32}$$

Let $\varepsilon > 0$ and consider the hierarchy $\hat{O}$ with $L+1$ layers, $\hat{\mu}_{L+1} = \varepsilon$, $\hat{\mu}_L = \mu_L - \varepsilon$ and $\hat{\mu}_i = \mu_i$ for each $i = 1, \ldots, L-1$. It follows by Theorem 6 that $\sum_{i=1}^{L} \mu_i = \bar{\omega}$. Thus, the output of $\hat{O}$ is

$$\hat{y}(\varepsilon) = \frac{1 - c\mu_1 - ch\sum_{i=2}^{L} \mu_i(1 - F(\sum_{l=1}^{i-1} \mu_l)) + ch\varepsilon(1 - F(\sum_{l=1}^{L-1} \mu_l)) - ch\varepsilon(1 - F(\bar{\omega} - \varepsilon))}{1 + h\sum_{i=2}^{L}(1 - F(\sum_{l=1}^{i-1} \mu_l)) + h(1 - F(\bar{\omega} - \varepsilon))}.$$

Clearly, $\hat{y}(0) = y_L$ and, by (32),

$$\gamma_L \hat{y}'(0) = ch(1 - F(\sum_{i=1}^{L-1} \mu_i)) - y_L h f(\bar{\omega}) > 0.$$

Thus, for some $\varepsilon > 0$, $\hat{y}(\varepsilon) > \hat{y}(0) = y_L$. But this contradicts the optimality of O and shows that (31) holds. ■

Part 2: The necessity part follows by Lemma A.32. Indeed, the optimality of O_2^* implies that $\sup_L y_L = y_2(0, \bar{\omega}) = y_2$ and that (31) holds with $L = 2$ and $\mu_1 = 0$. Thus,

$$\frac{c}{f(\bar{\omega})} = \frac{c(1 - F(\sum_{i=1}^{L-1} \mu_i))}{f(\bar{\omega})} \leq y_2 = \sup_L y_L.$$

Conversely, assume that $\frac{c}{f(\bar{\omega})} \leq \sup_{L'} y_{L'}$. Let O be an optimal organization, which exists by part 1, and L be the number of its layers; thus, $y_L = \sup_{L'} y_{L'}$. Suppose, in order to reach a contradiction, that $L > 2$. Since $\mu_L > 0$ and $\mu_{L-1} > 0$ by Lemma A.16, the first-order conditions imply that $\frac{\partial y_L(\mu_1,\ldots,\mu_L)}{\partial \mu_L} = \frac{\partial y_L(\mu_1,\ldots,\mu_L)}{\partial \mu_{L-1}}$; thus, using $\sum_{i=1}^{L} \mu_i = \bar{\omega}$ by Theorem 6,

$$cF(\bar{\omega} - \mu_L) = c(F(\bar{\omega} - \mu_{L-1} - \mu_L) + \mu_L f(\bar{\omega} - \mu_L)) + y_L f(\bar{\omega} - \mu_L).$$

Since $F(\bar{\omega} - \mu_{L-1} - \mu_L) \geq 0$, it follows that

$$cF(\bar{\omega} - \mu_L) \geq (c\mu_L + y_L)f(\bar{\omega} - \mu_L). \tag{33}$$

Consider the function $g : [0, \bar{\omega}] \to \mathbb{R}$ defined by setting, for each $x \in [0, \bar{\omega}]$,

$$g(x) = cF(\bar{\omega} - x) - (cx + y_L)f(\bar{\omega} - x);$$

condition (33) then states that there is $x^* \in (0, \bar{\omega})$ such that $g(x^*) \geq 0$. But g is strictly decreasing (since F is strictly increasing and f is strictly decreasing) and $g(0) = c - y_L f(\bar{\omega}) \leq 0$, hence no such x^* exists. This contradiction, in turn, implies that $L = 2$.

Theorem 6 together with $L = 2$ then implies that $O = O_2^*$, i.e. that O_2^* is an optimal organization and the only one.

Finally, note that (3) implies that $y_2 = \sup_L y_L$ and, hence, $\frac{c}{f(\bar{\omega})} \leq y_2$. Conversely, the latter implies (3) since we always have $y_2 \leq \sup_L y_L$.

Part 3: Suppose that f is differentiable and that (4) hold. Note first that

$$\frac{\partial \left(\frac{F(x+y) - F(x)}{f(x+y)} \right)}{\partial y} > 0 \tag{34}$$

for each $x, y \in \Omega$ such that $x + y \leq \bar{\omega}$ since

$$\frac{\partial \left(\frac{F(x+y) - F(x)}{f(x+y)} \right)}{\partial y} = \frac{f(x+y)^2 + (-f'(x+y))(F(x+y) - F(x))}{f(x+y)^2}$$

and $f'(x+y) < 0$ since f is strictly decreasing.

For the necessity part, note that if O_3^* is an optimal organization, then $\sup_{L \in \mathbb{N}} y_L = y_3$, and (31) implies $\frac{c(1-F(\mu_2^*))}{f(\bar{\omega})} \leq y_3$. We also that that $\sup_{L \in \mathbb{N}} y_L < \frac{c}{f(\bar{\omega})}$ because otherwise part 2 of this theorem would imply that O_2^* is the unique optimal organization and so O_3^* could not be optimal.

Regarding sufficiency of (5), let O be an optimal organization, which exists by part 1, and L be the number of its layers. Suppose, in order to reach a contradiction, that $L > 3$ (note that $L \notin \{1, 2\}$ by Theorems 6 and part 2) and let $(\mu_1, \ldots, \mu_L)$ be a solution to (28)–(29). By Lemma A.16 and Theorem 6, it follows that $(\mu_2, \ldots, \mu_{L-1})$ is such that $\mu_i > 0$ for each $i = 2, \ldots, L-1$ and it solves

$$y_L = \max_{\mu_2, \mu_3, \ldots, \mu_{L-1}} \frac{1 - ch \sum_{l=2}^{L-1} \mu_l (1 - F(\sum_{j=2}^{l-1} \mu_j)) - ch(\bar{\omega} - \sum_{l=2}^{L-1} \mu_l)(1 - F(\sum_{j=2}^{L-1} \mu_j))}{1 + h \sum_{l=2}^{L} (1 - F(\sum_{j=2}^{l-1} \mu_j))}$$

subject to $\sum_{j=2}^{L-1} \mu_j < \bar{\omega}$. The first-order condition FOC_i with respect to μ_i is, for each $i \in \{2, \ldots, L-1\}$,

$$
\begin{aligned}
&- ch \left(1 - F\left(\sum_{j=2}^{i-1} \mu_j\right)\right) + ch \sum_{l=i+1}^{L-1} \mu_l f\left(\sum_{j=2}^{l-1} \mu_j\right) \\
&+ ch\mu_L f\left(\sum_{j=2}^{L-1} \mu_j\right) + ch\left(1 - F\left(\sum_{j=2}^{L-1} \mu_j\right)\right) + y_L h \sum_{l=i+1}^{L} f\left(\sum_{j=2}^{l-1} \mu_j\right) = 0.
\end{aligned}
$$

Equivalently:

$$
\frac{F(\sum_{j=2}^{i} \mu_j) - F(\sum_{j=2}^{i-1} \mu_j)}{f(\sum_{j=2}^{i} \mu_j)} = \frac{y_L}{c} + \mu_{i+1} \text{ for each } i \in \{2, \ldots, L-2\}, \text{ and} \tag{35}
$$

$$
\frac{F(\sum_{j=2}^{L-1} \mu_j) - F(\sum_{j=2}^{L-2} \mu_j)}{f(\sum_{j=2}^{L-1} \mu_j)} = \frac{y_L}{c} + \bar{\omega} - \sum_{j=2}^{L-1} \mu_j, \tag{36}
$$

where (35) is obtained by subtracting FOC_{i+1} from FOC_i for each $i \in \{2, \ldots, L-2\}$ and (36) is just FOC_{L-1}.

Suppose for a contradiction that $\sum_{j=2}^{L-2} \mu_j < \mu_2^*$. By the definition of μ_2^*, the assumption that $y_L > y_3$, and because $F(\sum_{j=2}^{L-3} \mu_j) \geq 0$, it follows that

$$
\frac{F(\mu_2^*) - F(\sum_{j=2}^{L-3} \mu_j)}{f(\mu_2^*)} \leq \frac{F(\mu_2^*)}{f(\mu_2^*)} = \frac{y_3}{c} + \bar{\omega} - \mu_2^* < \frac{y_L}{c} + \bar{\omega} - \mu_2^*.
$$

Thus, if $\sum_{j=2}^{L-2} \mu_j < \mu_2^*$, (34) and (35) imply

$$
\frac{y_L}{c} + \mu_{L-1} = \frac{F(\sum_{j=2}^{L-2} \mu_j) - F(\sum_{j=2}^{L-3} \mu_j)}{f(\sum_{j=2}^{L-2} \mu_j)} < \frac{F(\mu_2^*) - F(\sum_{j=2}^{L-3} \mu_j)}{f(\mu_2^*)} < \frac{y_L}{c} + \bar{\omega} - \mu_2^*
$$

and, hence, $\mu_{L-1} < \bar{\omega} - \mu_2^*$.

Letting $x = \sum_{j=2}^{L-2} \mu_j$, $y = \mu_{L-1}$ and $x' = \bar{\omega} - \mu_{L-1}$, (4) together with the mean value theorem imply that

$$
\frac{1 - F(\bar{\omega} - \mu_{L-1})}{f(\bar{\omega})} > \frac{F(\sum_{j=2}^{L-2} \mu_j + \mu_{L-1}) - F(\sum_{j=2}^{L-2} \mu_j)}{f(\sum_{j=2}^{L-2} \mu_j + \mu_{L-1})} - (\bar{\omega} - \mu_{L-1} - \sum_{j=2}^{L-2} \mu_j).
$$

This together with $\mu_{L-1} < \bar{\omega} - \mu_2^*$ and (5) imply that

$$
\frac{F(\sum_{j=2}^{L-2} \mu_j + \mu_{L-1}) - F(\sum_{j=2}^{L-2} \mu_j)}{f(\sum_{j=2}^{L-2} \mu_j + \mu_{L-1})} + \sum_{j=2}^{L-2} \mu_j + \mu_{L-1} < \frac{1 - F(\mu_2^*)}{f(\bar{\omega})} + \bar{\omega} \leq \frac{y_L}{c} + \bar{\omega}.
$$

But this contradicts (36).

The above contradiction implies that $\sum_{j=2}^{L-2} \mu_j \geq \mu_2^*$. This together with (34) (use $x = \sum_{j=2}^{L-2} \mu_j$, $y = \mu_{L-1}$ and $y' = \bar{\omega} - \sum_{j=2}^{L-2} \mu_j$), $\sum_{j=2}^{L-1} \mu_j < \bar{\omega}$ and (5) imply that

$$\frac{F(\sum_{j=2}^{L-1} \mu_j) - F(\sum_{j=2}^{L-2} \mu_j)}{f(\sum_{j=2}^{L-1} \mu_j)} + \sum_{j=2}^{L-1} \mu_j < \frac{1 - F(\sum_{j=2}^{L-2} \mu_j)}{f(\bar{\omega})} + \bar{\omega}$$
$$\leq \frac{1 - F(\mu_2^*)}{f(\bar{\omega})} + \bar{\omega} \leq \frac{y_L}{c} + \bar{\omega}.$$

But this contradicts (36). This contradiction shows that we cannot have $L > 3$. Thus, it follows that $L = 3$.

To complete the proof, we need to show that if (μ_1, μ_2, μ_3) solves (28)–(29) for $L = 3$ then $\mu_1 = 0$, $\mu_2 = \mu_2^*$ and $\mu_3 = \bar{\omega} - \mu_2^*$. Note that since $\mu_2 > 0$ and $\mu_3 > 0$, $\frac{\partial y_3(\mu_1,\mu_2,\mu_3)}{\partial \mu_2} = \frac{\partial y_3(\mu_1,\mu_2,\mu_3)}{\partial \mu_3}$. This equality together with $\mu_1 = 0$ and $\mu_3 = \bar{\omega} - \mu_2$ (which follow from Theorem 6) implies that

$$\mu_2 + \frac{F(\mu_2)}{f(\mu_2)} = \frac{y_3}{c} + \bar{\omega}.$$

Note that $\mu_2 = 0$ implies $y_3 = -c\bar{\omega}$ and that $\mu_2 = \bar{\omega}$ implies $y_3 = \frac{c}{f(\bar{\omega})}$; since $0 < y_3 < \frac{c}{f(\bar{\omega})}$ (the latter by (5)), neither $\mu_2 = 0$ nor $\mu_2 = \bar{\omega}$ is a solution to this equation. Since the right hand side is strictly increasing in μ_2, the equation has a unique solution μ_2^*. In conclusion, O_3^* is an optimal organization and the only one.

Finally, note that, (5) implies that $y_3 = \sup_L y_L$ and, hence, $\frac{c(1-F(\mu_2^*))}{f(\bar{\omega})} \leq y_3 < \frac{c}{f(\bar{\omega})}$. Conversely, the latter implies that $\frac{c(1-F(\mu_2^*))}{f(\bar{\omega})} \leq \sup_L y_L$ since we always have $y_3 \leq \sup_L y_L$. If $\sup_L y_L \geq \frac{c}{f(\bar{\omega})}$, then $y_2 = y_2(0, \bar{\omega}) = \sup_L y_L$ by part 2 and, hence, $y_3 = y_3(0, \bar{\omega}, 0) = y_2 = \sup_L y_L$, contradicting $y_3 < \frac{c}{f(\bar{\omega})}$. Thus, $\sup_L y_L < \frac{c}{f(\bar{\omega})}$ and (5) holds.

A.15 Sufficient conditions for (4)

We first show that a sufficient condition for (4) is that

$$f(0) < 2f(\bar{\omega}). \tag{37}$$

Indeed, note that

$$\frac{\partial \left(\frac{F(x+y)-F(x)}{f(x+y)} \right)}{\partial x} = \frac{f(x+y) - f(x)}{f(x+y)} + \frac{(F(x+y) - F(x))(-f'(x+y))}{f(x+y)^2}. \tag{38}$$

The second term is positive, and since $f(0) < 2f(\bar{\omega})$, the first term is bounded below by $\frac{f(\bar{\omega})-f(0)}{f(\bar{\omega})} = 1 - \frac{f(0)}{f(\bar{\omega})} > -1$.

While sufficient, (37) is not necessary for (4) as it can easily be seen for the case where f is affine. Indeed, when $\bar{\omega} = 1$ and $f(z) = a - 2(a-1)z$ for each $z \in \Omega$ and some $1 < a < 2$, $f(0) = a$, $f(\bar{\omega}) = 2 - a$ and $f(0) < 2f(\bar{\omega})$ holds if and only if $a < 4/3$. That (4) holds for any affine f is a consequence of the fact that another sufficient condition for (4) is that

$$f \text{ is concave.} \tag{39}$$

Indeed, we can write (38) as

$$\frac{\partial\left(\frac{F(x+y)-F(x)}{f(x+y)}\right)}{\partial x} = \frac{y}{f(x+y)}\left[\frac{f(x+y)-f(x)}{y} + \frac{F(x+y)-F(x)}{y}\frac{(-f'(x+y))}{f(x+y)}\right].$$

Since $F''(z) = f'(z) < 0$ for each $z \in \Omega$, F is concave and $\frac{F(x+y)-F(x)}{y} \geq F'(x+y) = f(x+y)$. Thus,

$$\frac{\partial\left(\frac{F(x+y)-F(x)}{f(x+y)}\right)}{\partial x} \geq \frac{y}{f(x+y)}\left[\frac{f(x+y)-f(x)}{y} - f'(x+y)\right] \geq 0$$

if f is concave.

As before, (39) is not necessary for (4). This can be seen by considering the (truncated) exponential distribution, e.g. let $\bar{\omega} = 1$, $\lambda > 0$ and $f(z) = \frac{\lambda e^{-\lambda z}}{1-e^{-\lambda}}$ for each $z \in [0,1]$. Then $F(z) = \frac{1-e^{-\lambda z}}{1-e^{-\lambda}}$ for each $z \in [0,1]$ and, for each $0 \leq x \leq x+y \leq 1$,

$$\frac{F(x+y)-F(x)}{f(x+y)} = \frac{e^{-\lambda x}(1-e^{-\lambda y})}{\lambda e^{-\lambda x}e^{-\lambda y}} = \frac{1-e^{-\lambda y}}{\lambda e^{-\lambda y}}.$$

It then follows that $\frac{\partial\left(\frac{F(x+y)-F(x)}{f(x+y)}\right)}{\partial x} = 0$.