# Knowledge-Based Hierarchies: Using Organizations to Understand the Economy

Luis Garicano[1] and Esteban Rossi-Hansberg[2]

[1]Department of Management and Department of Economics, London School of Economics, London WC2A 2AE, United Kingdom; email: luis.garicano@gmail.com

[2]Department of Economics and Woodrow Wilson School, Princeton University, Princeton, New Jersey 08544-1021; email: erossi@princeton.edu

Annu. Rev. Econ. 2015. 7:1–30

First published online as a Review in Advance on February 6, 2015

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi: 10.1146/annurev-economics-080614-115748



JEL codes: D31, L23, D83

## Keywords

firms, inequality, skill premium, span of control, wages

## Abstract

Incorporating the decision of how to organize the acquisition, use, and communication of knowledge into economic models is essential to understand a wide variety of economic phenomena. We survey the literature that has used knowledge-based hierarchies to study issues such as the evolution of wage inequality, the growth and productivity of firms, economic development, and the gains from international trade, as well as offshoring and the formation of international production teams. We also review the nascent empirical literature that has, so far, confirmed the importance of organizational decisions and many of their more salient implications.

# 1. INTRODUCTION

Most human endeavors—from the construction of great infrastructure projects to the development of key ideas and innovations, even the production of a simple pencil (as in Friedman's famous example)—require the collaboration of teams of individuals. Understanding the organization of the skill, knowledge, and time of individuals into teams with common production goals is essential to comprehend virtually all human activities. However, most mainstream economic models still abstract from modeling the organizational problem that is necessarily embedded in any production process. Typically, these frameworks jump directly to the formulation of a production function that depends on total quantities of a predetermined and inflexible set of inputs. In this review, we argue that this simplification, although often practical, leaves important issues aside.

The importance of understanding organizations has been acknowledged at least since Knight's (1921) discussion of the role of managers in making decisions under uncertainty.[1] Some early attempts were made by Mayer (1960) and Ijiri & Simon (1964), who sought to explain the skewness in the distribution of pay and the firm size distribution. Team theory, following Radner (1993), examined the role of organizations as information processing devices: By processing information in parallel, hierarchy can reduce delay. Others, such as Geanakoplos & Milgrom (1991), studied decentralized resource allocation with limited information on unit costs. A line of research, following Calvo & Wellisz (1978), also investigated hierarchies in which managers at each level engaged in monitoring their subordinates. All these papers examine organization in isolation from the market and do not allow for the study of the interaction between organization and the economy, as they involve neither general equilibrium nor heterogeneity. Lucas (1978) and Rosen (1982) introduced models of equilibrium assignment and thus allowed for the joint study of the wage and firm size distribution.[2] Consequently, the literature on knowledge-based hierarchies that we and others have developed, and we review here, aims to incorporate organization into mainstream economics frameworks.

Knowledge is an indispensable input in production. So the ability to efficiently manage its acquisition, communication, and use is a determinant of the productive efficiency of an organization (e.g., a team, a firm, or even an economy as a whole). We conceptualize knowledge as the ability to solve the problems that naturally arise in any production process. The organizational problem arises because this knowledge is embedded in individuals who have limited time to work. Hence, although the knowledge they posses can be used repeatedly as an input in production, the individual who possesses it faces a time constraint that limits in practice how often this knowledge can be used. The theory of knowledge-based hierarchies recognizes that one way to relax this time constraint is to work in teams, in which less knowledgeable workers deal with routine production tasks, thereby economizing on the time of experts and allowing them to specialize in giving directions on the harder tasks. The organizational problem then becomes the problem of determining who knows what, who communicates with whom, and how many workers of each type are required to minimize the cost of producing a certain output. That is, it becomes the problem of determining how to use knowledge efficiently.

Because the organizational problem jointly optimizes the knowledge and number of members of a production team, it links the marginal return of a given worker to the characteristics of the other workers in his or her team, and more broadly to the distribution of knowledge and wages in the population. In fact, labor market and firm organization are two sides of the same coin. The link

[1] Garicano & Van Zandt (2013) present a recent review of the literature on hierarchies.

[2] Prescott & Visscher (1980) and Prescott & Boyd (1987) also incorporate organizations into equilibrium models.

between the two is the scale of operations effect, in which the marginal value of an agent's ability is given by the number of workers he or she manages. So one important application of the organizational problem described above is to understand the impact of economy-wide changes in technologies that affect the acquisition and communication of knowledge [information and communication technology (ICT)] on the distribution of wages in the economy.

Take, for example, the well-known fact of wage polarization. The labor literature has argued that, during the past few decades, inequality at the top of the wage distribution has increased dramatically, while it has concurrently declined at the bottom. Simple theories that explain increases in wage inequality as a result of increases in the price of skill have a hard time reconciling these facts. They are, however, easily rationalized by the theory of knowledge-based hierarchies as a result of large improvements in communication technology since the late 1980s. Better communication technology effectively relaxes the time constraint of experts by allowing them to deal with more tasks. This implies that less knowledgeable workers, who earn less, become less differentiated because the purpose of their knowledge is to relax a constraint that is now less binding. The result is that experts, the superstars of the knowledge economy, earn a lot more, whereas less knowledgeable workers become more equal as their knowledge becomes less useful. Moreover, communication technology allows superstars to leverage their expertise by hiring many workers who know little, thereby casting a shadow on the best workers who used to be the ones exclusively working with them. We call this the shadow of superstars.

In Section 2, we discuss this example in the context of the existing literature as way of motivation. The rest of this article presents a more systematic survey of the work that we and others have done to incorporate organizations into equilibrium frameworks as well as the application of these models to a variety of issues, including firm heterogeneity, economic development, and international trade and offshoring.

Thus, in Section 3, we introduce the basic technology for knowledge utilization that relies on the use of hierarchies. We capture the idea, best expressed by Demsetz (1988, p. 157), that "those who are to produce on the basis of this knowledge, but not be possessed of it themselves, must have their activities directed by those who possess (more of) the knowledge. Direction substitutes for education (that is, for the transfer of the knowledge)." Instead of educating all agents, which is very costly, those in production learn the most routine tasks and ask for directions whenever they need help on less routine tasks. We start by setting up such a model and by reviewing some of the basic results that carry on throughout.

Organizations can be introduced in either an economy in which agents are ex ante identical or one in which they are ex ante heterogeneous. We reserve the concept of skill (as opposed to knowledge) for an innate (or early acquired) exogenous individual characteristic. In frameworks in which agents are ex ante identical, skill is completely abstracted from. In models in which agents are ex ante heterogeneous, either skill can be synonymous with knowledge, when the acquisition of knowledge is not explicitly modeled, or it can determine the cost of acquiring knowledge, when knowledge acquisition is incorporated.

Knowledge-based hierarchies are efficient in either of these cases. When individuals are ex ante homogeneous, they increase the utilization rate of knowledge. All agents obtain identical utility, although they acquire different knowledge depending on their roles in the organization—wages just compensate workers for the cost of knowledge. But instead of having all production workers learn how to do something, most of them specialize in routine tasks, and only a few agents specialize in exceptions and communicate this specialized knowledge to producers as needed. This ex post heterogeneity is the outcome of the organizational problem. In Section 3, we review the basic frameworks that use homogeneous agents, starting from Garicano (2000), who sets up the basic technology and the circumstances under which knowledge-based hierarchies are optimal. We then discuss Caliendo & Rossi-Hansberg (2012), who embed this basic technology in a classic cost-minimization problem and then incorporate it into an equilibrium framework with heterogeneous firms. We finish Section 3 with a brief review of the dynamic problem in Garicano & Rossi-Hansberg (2012) of deciding when to exploit knowledge by using hierarchies to deepen knowledge in existing ideas, versus starting a new field, which requires discarding existing organizations.

In frameworks with ex ante heterogeneous agents, hierarchical organization also becomes a device to allocate heterogeneous agents into positions in which they can optimally leverage their skills. Hence, these models involve an assortative matching problem between agents with different skill levels. The wage function that results from this problem does not simply compensate agents for their cost of education, but it values the specific skills they possess relative to the rest of the individuals in the economy. As such, these frameworks can address the facts on the evolution of the labor market discussed above. Even though the resulting matching problem can be illuminating, it also comes at an analytical cost. So the benefits of using this variant of the framework certainly depend on the application one has in mind. In Section 4, we first discuss Garicano & Rossi-Hansberg (2004), who have ex ante heterogeneous skilled agents but only two layers and no investment in knowledge. We then proceed to discuss Garicano & Rossi-Hansberg (2006), who allow multiple layers, ex ante heterogeneity, and investment in knowledge. We conclude the section by studying the problems that appear when hierarchies operate in markets with double-sided asymmetric information, as in Fuchs et al. (2015).

In Section 5, we move on to applications of these frameworks that examine international trade and the international organization of production. In particular, we discuss how the gains from a trade liberalization are affected by the changes in organization they trigger. Because these changes in organization are mediated by the level of ICT, these technologies affect the gains from openness and the impact of international trade on wage distributions and firm productivity. We then turn to the implications of the theory of knowledge-based hierarchies on offshoring. Specifically, Antràs et al. (2006) show that the matching of individuals in international production teams can affect wages in countries with different skill distributions in ways that are consistent with the empirical evidence.

Section 6 describes the existing empirical work related to organization in knowledge-based hierarchies and its implications. One of the premises underlying the organizational problem is that the characteristics of production teams are not fixed or predetermined by some exogenous force but are active decisions by firms or market participants. As such, these decisions should change when, for example, firms decide to grow or shrink. Caliendo et al. (2015) show that, when French firms grow, wages at each layer react very differently depending on whether firms reorganize by adding new layers of management. Another central implication of the theories of knowledge-based hierarchies is that, within layers, more knowledgeable workers are matched with more knowledgeable managers, so more knowledgeable managers also lead larger teams. Garicano & Hubbard (2012) show that this is in fact the case in US law firms. We finish the review with Section 7, which points the reader to what we believe are good directions for further research.

## 2. THE NEW ORGANIZATION OF WORK: WORKERS IN THE SHADOW OF SUPERSTARS

The distribution of wages and occupations has experienced dramatic changes over the past few decades.[3] In fact, the evolution of the wage structure provides us with a powerful tool to

[3]Acemoglu & Autor (2011) present a systematic assessment of the main trends in the data and the existing attempts to explain them with a unified framework.

discriminate between alternative theories about the fundamental forces driving changes in the labor market, as well as the way heterogeneous workers interact in the labor market.[4]

The first central fact is that there has been a clear increase in the skill premium for workers between 1973 and 2009. Furthermore, the increase is concentrated on highly educated workers. For less educated workers, the increase in the slope is small or perhaps reversed.

Beyond education, many other characteristics are important in determining the knowledge and skill levels of a worker. Perhaps the best measure of the marketable knowledge and skills of an agent is his or her wage. Hence, the labor literature has proposed the use of wage percentiles, summarizing workers' ranks in the overall distribution, as a measure of the rank of skill levels. The evidence shows that the winners from changes in the wage distribution between 1988 and 2008 were those in the bottom and top percentiles. The same was not true during the previous 14 years during which workers at the bottom of the distribution suffered important reductions in their wages. This phenomenon has been labeled wage polarization. The finding, together with the convexification of the relationship between wages and schooling, indicates that during the past three decades, the mechanism causing these changes must have affected workers in the middle of the skill (or knowledge) distribution negatively, whereas it should have affected other workers positively. **Figure 1** show this unequivocally. After an initial increase, starting in 1986, inequality between the 50th and 10th percentile of the distribution has decreased slightly, while inequality in the top half of the distribution has increased markedly.[5]

Simple theories of changes in the return to skill will clearly fail to explain these facts. First, whatever theory one postulates to explain the evolution of the wage distribution needs to predict that the return to skill of workers with less units of skill behaves differently than that of workers with units of skill above the median. That is, we need a theory in which the price of a unit of skill depends on the number the individual possesses. Second, we need a theory in which the baseline wage increases, but the return to skill decreases for the bottom percentiles, whereas it increases for the top ones. Finally, the mechanism we propose has to be operative starting in the mid- to late 1980s, but not before.

The so-called canonical model explains the changes in the wage distribution as a race between technology and education [see the encyclopedic and informative book by Goldin & Katz (2009) for an account of many of these facts using this approach]. In this view, pioneered by Katz & Murphy (1992), as technology evolves, it progressively becomes more skill intensive (i.e., technological change is skill biased). This raises the return to skill as long as the increase in the supply of skilled workers does not compensate. In a series of influential papers,[6] Acemoglu gives precise conditions under which changes in technology will lead to an increase in the supply of workers without eliminating the increase in the returns to skill.

These explanations are successful in accounting for the increase in the skill premium in the past few decades. However, because they speak only about the price of a unit of skill, rather than the price of a unit of skill as a function of the number of units an individual possesses, they cannot account for the distinct behavior of the wage distribution for low-, medium-, and high-skill individuals.

---

[4]Katz & Murphy (1992), Acemoglu (1999), Autor et al. (2003, 2006, 2008), Lemieux (2006), and Acemoglu & Autor (2011) present all the key facts.

[5]The figure relies on the Current Population Survey data for the US economy during the past few decades, as cleaned up and provided by David Autor.

[6]Acemoglu (2002) provides a nice survey of the skill-biased technological change literature.

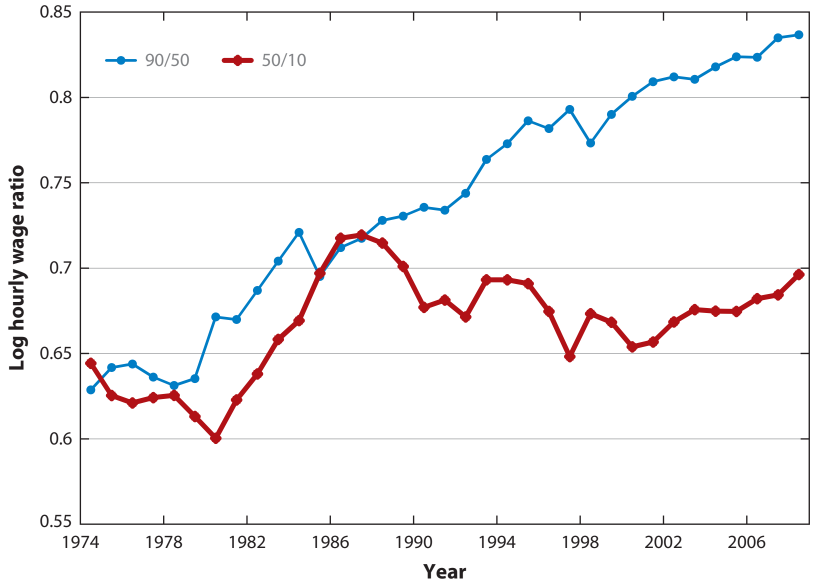


**Figure 1**

Polarization in wages, showing 90/50 and 50/10 wage inequality.

Clearly, the salient fact in **Figure 1** cannot safely be ignored, so the literature has evolved to develop explanations that point to a different evolution of the price of skill across individuals. Two such explanations have been developed. First, there is the task-based approach initially proposed by Autor et al. (2003, 2006) and forcefully pushed as the central explanation of these facts by Acemoglu & Autor (2011). Second, there are theories emphasizing the returns to managerial activities, such as those by Gabaix & Landier (2008) and initially by Lucas (1978) and Rosen (1982). The latter set of papers includes theories of organization, so we discuss them more fully below. Here, we note only that they fail to account for the behavior of the returns to skill for workers with low levels of skill. That is, these are theories that explain how changes in technology have increased the returns to managers but cannot explain why they have decreased the returns to skill at the bottom of the distribution. As such, these theories are interesting as explanations of the emergence of superstar earners (Rosen 1981) but not as explanations of the shadow they seem to have cast on medium-skill workers.

The task-based explanation is the best known explanation of these patterns. These theories do not introduce organization explicitly. They postulate a structure of substitution between distinct tasks in the production process and exploit it to generate complementarity between tasks. A reduction in the cost of producing computers, for example, will reduce the need of some tasks in favor of others depending on the postulated substitution patterns assumed in the production function. A production function with constant elasticity of substitution between tasks, with a nesting structure that makes capital and routine tasks highly substitutable, is common.

One advantage of this approach is that the elasticities of substitution in the production function can be estimated. Harder to estimate and specify is the nesting structure between different tasks and factors in the production function, not to mention the arbitrariness in defining the relevant tasks to include in the production function. A standard story in this literature is that the decline in the cost of capital equipment has decreased the price of routine tasks that are highly substitutable with this form of capital. In contrast, analytic and manual tasks are less substitutable, so the improvement in technology leads to higher demand and higher employment of workers performing those tasks.

A similar story can be told to explain the impact of offshoring. The tasks that are easily offshored are the ones that can be specified in a set of well-defined rules. These are again the routine tasks done by workers in the middle of the wage distribution. The result is wage and employment polarization, as illustrated in **Figure 1**.

The organization-based explanation of these facts that we propose does not contradict this task-based approach. It is, to some extent, a distinct specification of the task-based approach in which tasks are hierarchical and the production function that links the different tasks is based on an explicit organizational problem. The flexible hierarchical nature of this theory has the potential to explain the set of distinct inputs used in production as well as their characteristics. Thus, it can explain, for example, why the medium-skill agents are the ones who have lost in the past few decades. Other task-based approaches also link these losses to agents specializing in routine tasks but do not explain why these are the agents with intermediate incomes. In that sense, in the theories reviewed by Acemoglu & Autor (2011), the fact that medium-skill agents are in the shadow of superstars is hardwired in the theory, not endogenous to the evolution of the economy. Furthermore, because the theory of knowledge-based hierarchies provides an explicit organizational theory, it can distinguish in a precise way between different types of ICT changes and can explain their distinct effects on workers' returns and firm performance.

## 2.1. Superstar Managers and Increases in the Scale of Operations

A well-known explanation for the increasing concentration of rewards at the top of the earnings distribution has to do with the increase in the effective market size of top talent: the scale of operations effect.[7] The idea is that a good decision by a better manager increases the productivity of all of his or her subordinates. As Rosen (1982) argues, a good soldier is useless in the wrong war. In this context, large returns to talent can be generated by the competition for better managers between organizations that can leverage the manager's knowledge by assigning him or her large teams. This argument can explain the phenomenon of superstar pay. As first pointed out by Manne (1965), a better concert pianist has higher earnings not only because consumers are willing to pay more for better performances, but also because he or she plays in larger halls.[8]

The first formal treatment that tries to capture these arguments is provided by Lucas (1978). Lucas's (1978, p. 510) starting point is Manne's (1965) suggestion that "observed size distribution is a solution to the problem: allocate productive factors over managers of different ability so as to maximize output." Lucas solves exactly that problem. He departs minimally from the neoclassical production model in that, essentially, a manager is required to run a firm formed by a set of undifferentiated workers and capital. A manager with skill $\alpha$ uses labor $n$ to produce $\alpha f(n)$ units of

[7]Mayer (1960) provides the first account of this effect. Chandler (1977) offers a historical account of the growth of scale in the US corporation since the end of the nineteenth century and the appearance of hierarchy, which he argues substituted for the old dominance of the market.

[8]Rosen's (1981) superstar model provides a formal model and a complete analysis of this scale of resources effects in the labor market. Even a top economist, who would be laboring in obscurity, can become an overnight star selling books worldwide thanks to the reach of information technology. This can happen even to economists who do not appear to believe in the importance of this mechanism, such as Piketty (2014).

output.[9] Lucas's aim is not so much to have a theory of organization, but to have a theory of why organization matters.

This model generates a superstar effect for the best managers through increases in the scale of operations effects as a result of drops in coordination or communication costs. To see this, let $f(n) = \alpha n^{\theta}$, where $\theta$ governs the decreasing returns that result from larger spans of control. As illustrated in **Figure 2**, an increase in $\theta$ leads each entrepreneur to want a larger team so that $n(\alpha)$ shifts upward. As a result, equilibrium requires that both the equilibrium worker's wage, $w$, and the cutoff skill between workers and managers, $\alpha^*$, move upward, to $w'$ and $\alpha^{*\prime}$. The top managers can command potentially very large teams, so the marginal value of skill $[f(n(\alpha)]$ at that high point in the distribution can be very large. Thus, competition for those top managers can lead to very high earnings—the managerial earnings function, $R$, becomes steeper as technology increases the marginal returns to managerial skill.

In contrast, as a result of the technological change, nothing happens to wage inequality among workers, because there is none. In this model, what is good for managers is also good for workers, who uniformly earn more (with the potential exception of managers who change occupation and become workers). That is, in this model, superstars cast no shadow.

For wage inequality among workers to exist, we need to model how their skill matters. Rosen (1982) develops a model of hierarchy in which managers increase the productivity of the team, like in Lucas (1978), but they also need to spend time supervising each worker. In this model, worker skill (and not just manager skill) matters, and under constant returns to scale, it enters in efficiency units. Thus, managers here care not about the specific characteristics of each worker, but only about the aggregate amount of skills they hire. In this respect, Rosen shows that this technology can be seen as a generalization of Lucas's technology (Rosen also has a multilayer hierarchy). Namely, one obtains $y = \alpha f(Q)$, where now $Q$ is the total number of efficiency units of labor supervised by a manager with skill $\alpha$.

Assume again that $f(Q) = Q^{\theta}$. **Figure 3** illustrates the impact of an increase in $\theta$. As a result of this technological change, which increases the effect of scale, all managers supervise a larger team in terms of $Q$. Moreover, the price of skill increases (as the demand for it goes up), and thus inequality in differences (although not in ratios, unlike in the empirical evidence) among workers increases. All workers benefit from the new technology, so, as in Lucas (1978), superstars cast a light, not a shadow, on others.

Can these types of models explain the patterns of earnings at the top of the distribution? Gabaix & Landier (2008) develop and calibrate a version of these superstar models of Lucas and Rosen, relying on extreme value theory, in which the production of the firm is the product of talent and firm size, as measured by the market value of the firm. The key implications of their model are that CEO pay is proportional to firm size, both cross sectionally and longitudinally, and that, for a given firm size, CEO pay varies with median firm size across countries. Their empirical evidence is consistent with the model. They find that the firm's market value is correlated with CEO pay in the cross section. Most strikingly, in the time series, they show that managerial compensation is almost perfectly correlated with mean market capitalization.

Can these scale theories account for the evolution of pay in the rest of the distribution? The answer is clearly no. In Lucas (1978), all workers gain, but there is no inequality among workers. The assumption in Rosen's (1982) model that workers are perfect substitutes (only total skill matters) implies, counterfactually, that in equilibrium managers are indifferent between working with a few relatively skilled workers and working with many unskilled workers. Hence, assignment patterns between individual managers and workers are indeterminate. Crucially, this

[9] In the original formulation of Lucas (1978), $n$ refers to labor per unit of capital.

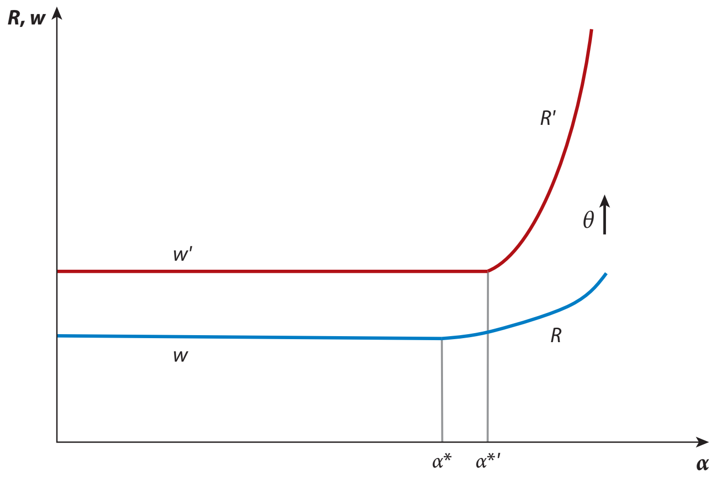


**Figure 2**

Hierarchy and inequality when worker skill is irrelevant (Lucas 1978).

substitutability also means that the price of skill moves in a uniform direction for all agents, in the face of an increase in the scale of operations effect. This implication is clearly counterfactual given the evidence in **Figure 1** and motivates a model in which the units of knowledge possessed by workers with different levels of knowledge are not perfect substitutes.

## 2.2. Workers in the Shadow of Superstars

Work involves solving problems. Confronting a hard problem requires a sufficiently knowledgeable worker. Multiple less knowledgeable workers will not solve the problem if none of them

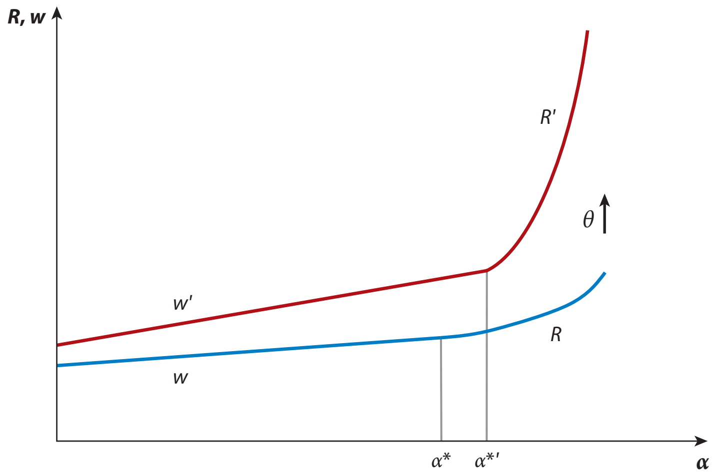


**Figure 3**

Hierarchy and inequality when only total efficiency units matter (Rosen 1982).

knows the answer. In this sense, the basic nature of knowledge implies its imperfect substitutability across heterogeneous individuals. If one worker on his or her own does not know how to program a robot, a team of 10 similar workers will also fail.

This type of problem solving production function was proposed by Garicano (2000). He shows that, if matching problems to solutions is hard,[10] it leads to a knowledge-based hierarchy. Garicano & Rossi-Hansberg (2004, 2006) embed the model of knowledge-based hierarchies in Garicano (2000) in a context in which workers have heterogeneous skill, like in Lucas (1978). In this model, workers are involved in production and managers in problem solving. The key difference is that the production function displays complementarities: Workers benefit from having good managers (they solve more problems and make their own time more productive), and managers benefit from having good workers (they need less help and allow managers to have a larger team). Thus, like in Lucas, manager skill determines the productivity of the team. Unlike in that work, here worker skill determines a manager's team size. The result is an equilibrium that exhibits positive one-to-many sorting between managers and workers.

In our model, because worker skill affects the scale of operations, the same mechanism that drives increasing inequality at the top reduces inequality at the bottom. Namely, superstars cast a shadow on workers, particularly on the ones that barely fail to become managers. To understand this, we note that increases in the scale of operations are direct, through the mechanism proposed by Lucas, the return to skill for managers. The marginal value of managerial skill is team size, and thus any mechanism that increases team size necessarily also increases the marginal return to skill. The result, as discussed above, is wage inequality among managers and the concentration of rewards at the very top. It also makes it more demanding to be a manager (raises the required cutoff), further contributing to the inequality between workers and managers.

The main novelty concerns inequality among workers. In a small-scale world, a good worker is managed by a much better manager than is a bad worker. The marginal product of the good worker is consequently much higher. In a world with large teams, workers are matched with only the best managers, which are similarly skilled, resulting in smaller differences in their marginal product. Hence, as the scale of operations increases, the marginal productivity of differently skilled workers becomes more homogeneous, so the inequality between them declines.

An example may help. Think of a world with very small firms. The best managers, for example, Steve Jobs or Alfred Sloan, run small workshops. They match with the best worker in the economy and produce the best products. Now pick up a worker who is only marginally worse. This worker is assigned to a worse manager, and together they make worse (or fewer) products. Thus, the worker's marginal product is much lower. Now increase the scale of operations so that Jobs or Sloan manage both workers. In this new world, the marginal product of both workers is differentiated only by their talent, not by the talent of their boss. The difference between them is smaller, so wage inequality is lower. In relative terms, the better worker loses from the enlarged shadow of the superstars.

**Figure 4** illustrates the patterns of wage inequality that follow from an increase in the scale of operations (in this model, a decrease in $h$, which parameterizes communication costs, as explained in detail in the next section). Managers, like in Lucas (1978), experience a superstar effect. Workers are now matched with a smaller range of managers and experience a reduction in inequality among them. Also, note the kink in the graph. It implies that the middle classes, the mid-skilled types, could

[10]That is, workers "don't know what they don't know." When an agent sees a problem that he or she cannot solve, the agent only knows that he or she does not know the solution.

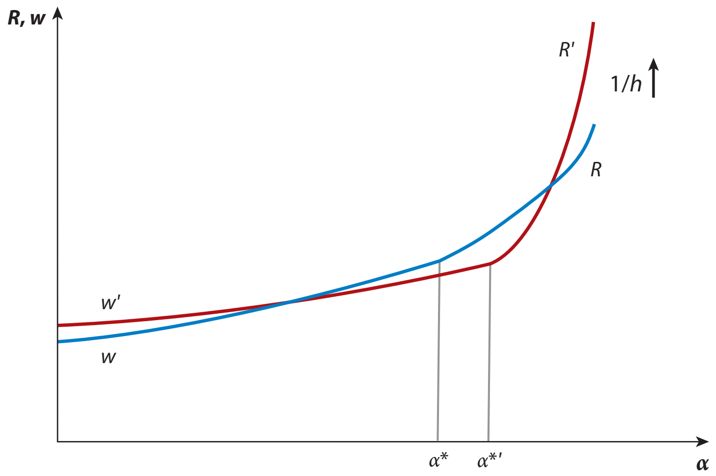


**Figure 4**

Hierarchy and inequality in Garicano & Rossi-Hansberg (2004).

actually be worse off in levels, not only in relative terms, as a result of the technological change, which is exactly what happened in the past few decades, as shown above.[11]

Under what conditions will this mechanism be operational? Essentially, when better workers allow managers to lead larger teams. This requires a production function in which managers' and workers' skills are complementary so that positive sorting follows, one with asymmetric skill sensitivity like in Kremer & Maskin (1996), in which workers of different skills are not perfect substitutes for one another, as well as scale effects like in Lucas (1978). As discussed next, knowledge-based hierarchies have precisely these characteristics: Production takes place in teams, as problems are solved either by workers or by their managers (management by exception); there is positive sorting because more skilled workers ask harder questions and thus require more knowledgeable managers; there are also scale effects as better workers use up less of their managers' time and thus allow for larger teams; and stratified sorting equilibria naturally result.

We move next to the formal part of our review, in which we set up and solve these types of problem solving production function, derive its properties, and show a variety of formulations and applications. We come back to the issues of inequality, with more formal treatments of the ideas above, in Section 4.

# 3. ORGANIZATION WITH EX ANTE HOMOGENEOUS WORKERS

## 3.1. The Basic Model of Knowledge-Based Hierarchies

We start by developing a basic model of an economy in which ex ante identical agents acquire knowledge and select into different roles in the organization. Because all agents are identical before they acquire an education, they obtain the same utility independently of their final knowledge acquisition and level in the hierarchy. Thus, this model without ex ante heterogeneity is inadequate to talk about heterogeneous welfare effects but allows us to study the structure of the hierarchy and

[11]We develop this model formally below and generalize it along multiple dimensions, including introducing several layers.

the knowledge acquisition decisions, and therefore wages, of the individuals that form them. Because individuals have no innate characteristics, there is no matching problem, so we simply need to determine the (hierarchical) position of each individual and the knowledge they acquire.

**3.1.1. Production with problem solving.** We introduce here the basic production technology proposed by Garicano (2000). Production requires time and knowledge. In order to produce, a worker must spend his or her unit of time and solve a problem. If the worker can solve it, he or she produces, and otherwise does not. The worker can only solve the problems he or she knows, either because of innate skills or because he or she has learnt them by incurring a cost. Thus, if a worker draws a problem $\zeta$ and his or her knowledge is in interval $Z \in R^+$, he or she produces if $\zeta \in Z$.

Suppose that the cost of acquiring a knowledge set $Z$ (learning all the problems in $Z$) is proportional to its size, that is, the number of problems in it (formally, its Lebesgue measure). Let the cost of learning a unit length of problems be $c$. For example, it costs $cz$ to learn $Z = [0, z]$.

Some problems are more common than others. The relative frequency of different problems is governed by a continuous distribution function $F$ with density $f$. The expected output $q$ of a worker is the probability that the worker confronts a problem he or she knows, so $q = \int_Z dF$. We assume without loss of generality that problems are indexed so that $f'(z) < 0$. Namely, the most common problems have a lower index. Then, a worker who can draw one problem per unit of time maximizes his or her expected net output, which is given by

$$q = \Pr(\zeta \leq z) - cz = \int_0^z f(\zeta)d\zeta - cz. \tag{1}$$

The optimal solution is implicitly given by $f(z^*) = c$, which equates the marginal value of acquiring knowledge to the marginal cost: The worker learns those problems that are common enough to justify investing in them.[12]

**3.1.2. Communication and production in teams.** Organization allows different workers to acquire different intervals of knowledge and communicate it as required. This has the advantage of allowing workers to increase the utilization rate of knowledge, decreasing the per capita learning cost. The trade-off is that it is costly to match the problem with the worker who knows the solution and to communicate the answer. In this context, hierarchies appear endogenously when matching problems to those who know how to solve them (or labeling the problems) is very costly. Then, the optimal way to obtain help is to ask progressively more knowledgeable workers for help until someone knows the solution, or until it is sufficiently unlikely that someone can solve the problem.[13]

The communication cost, or helping cost $h$, is the time spent away from production by workers communicating how to solve the problem. Garicano (2000) makes three assumptions about this cost. First, all the communication losses are incurred by the agent who is being asked and helps with the solution. Second, communication costs are incurred even when the agent asked does not know the answer because the agent must figure out if he or she knows the answer and communicate

[12]Throughout this article, we assume that the regularity conditions for the existence of a unique optimum are satisfied. If the density function $f(z)$ is nonincreasing, the second-order condition is always satisfied, and the solution is unique.

[13]The assumption that matching problems to knowledge is hard is realistic in situations in which knowledge is hard to codify, as, for example, when knowledge is tacit. Endogenous codes, in which agents have to choose where to set up the boundaries between events, are studied by Crémer et al. (2007).

it to the agent who asked.[14] Third, one finds that $h < 1$, capturing the idea that agents communicating how to solve the problem are not directly involved, and therefore do not spend time, in producing on the basis of that knowledge.

The problem of the organization is then to determine the number of distinct classes of workers (or layers), as well as their measure, their knowledge, a sequence of classes to ask sequentially for help, and production and helping time so as to maximize output per capita, subject to the time constraint that workers use their whole unit of time either helping or producing. Garicano (2000) shows that in a given organization, any arbitrary original allocation of workers, knowledge, communication, and time can be improved, and thus is not optimal, unless it has the following characteristics:

1. Agents specialize in either production or solving problems. Furthermore, only one class is specialized in production.
2. Knowledge acquired by different classes does not overlap.
3. Production workers learn to solve the most common problems; problem solvers learn the exceptions. Moreover, the sequence of classes of agents to ask for help is ordered sequentially from the bottom to the top of the hierarchy.
4. The organization has a pyramidal structure, with each layer of a smaller size than the previous one.

The key characteristic of this structure is management by exception. Production workers know solutions to common problems and successively ask problem solvers who know increasingly exceptional problems. To understand this result, note that any arrangement in which agents in some layer know problems in an interval $[z, z + \varepsilon]$ that come up more often than the problems known by those asking them questions, $[z', z' + \varepsilon]$, where $f(z) > f(z')$, can be improved upon. Simply swap these equally long intervals. Learning costs are unchanged, as all agents learn the same mass of problems as before. Production is unchanged, as the total amount of knowledge is unchanged. But communication costs are lower, as those asking questions are now less likely to confront a problem they do not know. Thus, the crucial insight is that, by relying on management by exception, specialization in knowledge can be attained while minimizing communication costs.

Given these results, we can simply let the knowledge of workers in layer 0 be given by $[0, z_0]$, that of layer 1 managers by $[z_0, z_0 + z_1]$, and, generically, that of agents in layer $\ell$ by $[Z_{\ell-1}, Z_{\ell-1} + z_\ell]$, where $Z_\ell = \sum_{l=0}^{\ell} z_l$ is the total knowledge in the organization up to layer $\ell$, and $z_\ell$ is the measure of knowledge an agent in layer $\ell$ acquires.

For example, output in an organization with one layer of management and $n_0$ workers is simply given by $q = F(Z_1)n_0$. In other words, production combines the (nonoverlapping) knowledge of the organization, $F(z_0 + z_1)$, with the production time of the $n_0$ workers with knowledge $z_0$. In choosing the size of his or her team, the manager faces a time constraint given by

$$1 \geq hn_0\big(1 - F(z_0)\big). \tag{2}$$

Namely, the unit of managerial time has to be sufficient to help the $n_0$ workers in the manager's team who ask with probability $1 - F(z_0)$ at a cost of $h$ units of his or her time. So better workers are more costly as they possess more knowledge, but they facilitate having a larger team, thereby leveraging the knowledge of the manager. Of course, the optimal organization might involve many

[14]The same problem comes up again with probability zero. Thus, there is no learning involved in communication.

layers of management, not just one as in this example. The next section characterizes the general problem using a cost-minimization approach.

## 3.2. Optimal Scale, Knowledge, and the Number of Layers

In the previous section, we introduce the basic technology of knowledge-based hierarchies and some of its key properties and illustrate the way the technology works with a simple two-layer hierarchy. Garicano (2000) goes further and studies the problem of a single organization that has potentially many layers of management. This problem is reformulated in a classic cost function approach by Caliendo & Rossi-Hansberg (2012), who then embed it in a standard heterogeneous firm model based on Melitz (2003).

Consider the cost minimization of a firm that aims to produce $q$ units of output with an organization similar to the one described above. Let the market price of a unit of work time be given by $w$, so the wage of an agent with one unit of time and $z$ units of knowledge is $cz + w$. The cost of producing $q$ units of output with an organization with $L$ layers, the wage bill of such an organization, is given by

$$C^L(q) = \min_{\{n_l^L, z_l^L\}_{l=0}^{L}} \sum_{l=0}^{L} n_l^L \left(cz_l^L + w\right), \tag{3}$$

subject to the organization producing at least $q$ units of output,

$$F\left(Z_L^L\right) n_0^L \geq q, \tag{4}$$

and the time constraints of managers at each layer,

$$n_l^L = h\left(1 - F\left(Z_{l-1}^{l-1}\right)\right) n_0^L \text{ for } l \leq L. \tag{5}$$

The last set of constraints denote that the number of units of time of managers in layer $l$ is given by the number of problems (or workers), $n_0^L$, times the fraction of them that has not been solved up to layer $\ell$, $\left(1 - F\left(Z_{l-1}^L\right)\right)$, times the amount of time managers need to spend with each problem, $h$.

Given a number of layers, the solution to this problem determines the structure of the efficient organization to produce $q$ units of output. Practically, it determines the knowledge levels and the number of agents in each of them that the organization needs to employ. Note that the problem above does not impose a constraint on the number of managers at the top of the organization, $n_L^L$. Absent such a constraint, the cost function is linear in $q$. This means that, given the number of layers, there is an optimal organization that is simply replicated to produce the desired level of output. A small corner shop will use a few minutes of the same organization used by a large corporation such as Walmart. This is clearly an unappealing feature.

Larger organizations use more complex organization with more layers in their corporate hierarchy. As we review in Section 6, Caliendo et al. (2015) present systematic evidence that larger French firms in fact use more complex hierarchies. One natural way of incorporating this realistic feature is to add an integer constraint on the number of managers at the very top of the organization. Because of their tacit knowledge and skills, it is hard for CEOs to replicate themselves. For example, we can assume, as in Caliendo & Rossi-Hansberg (2012), that each organization requires one agent at the top, so

$$n_L^L = 1. \tag{6}$$

With this constraint in place, one can show that the cost function is U-shaped in output, with a minimum efficient scale that increases with the number of layers and a total wage bill at the minimum efficient scale that decreases with the number of layers. So organizations that produce more use more layers and typically produce at a lower cost. Size gives an advantage by making a complex organization with many layers the efficient way to produce. In this sense, independently of any exogenous cost advantage firms might have, larger firms (e.g., owing to higher demand for a better product) will also have an endogenous cost advantage as a result of their more efficient organization. Clearly, such a large organization is only efficient if the firm produces a large quantity. For small production levels, such an organization is extremely inefficient. More formally, the cost function is given by the envelope of $C^L(q)$ for all $L$, namely,

$$C(q) = \min_{L \geq 0} C^L(q). \tag{7}$$

In reduced form, one could think of this organizational problem as choosing a technology from a menu of pairs of fixed and marginal costs. Large firms or organizations choose options with larger fixed costs and lower marginal costs, and small firms do the inverse. The organizational problem above goes beyond this reduced-form choice by specifying the characteristic of the workforce within the organization that leads to each particular combination.[15]

**3.2.1. Empirical implications: how firms grow.** A full characterization of this problem for the case of an exponential distribution of problems, $F(z) = 1 - e^{-\lambda z}$, can be found in Caliendo & Rossi-Hansberg (2012). Suppose the managerial integer constraint at the top, $n_L^L = 1$, determines the boundary of the firm so that organizations form as firms. Then, one central implication of the organizational problem above is that when firms reorganize to grow, for example, owing to a positive demand shock—by adding a layer of management—they reduce the knowledge and wages of all employees in pre-existing layers. Namely, they employ workers who are less knowledgeable. In contrast, when they grow without adding layers, they upgrade the knowledge of their employees and pay them more. The intuition is simple. When the organization grows substantially, it prefers to add one or a few experts (a new layer) rather than making all the existing workers learn more. The latter is too expensive because it involves training a potentially large number of workers. Hence, instead, the firm adds a layer of management and economizes on its wage bill by reducing the knowledge of everyone else. In contrast, when no layer is added, the firm has to increase the knowledge of everyone to increase its scale without violating the time constraints of its managers. Caliendo et al. (2015) present evidence that these implications are very much in line with the evidence for French firms, as discussed in Section 6.

**3.2.2. Empirical implications: information and communication technology and the organization of production.** A second set of implications concerns the impact of ICT on the hierarchy and knowledge acquired. The theory clearly distinguishes between these two aspects of technology. The key trade-off is between acquiring knowledge and asking others for help. Technologies that make acquiring knowledge cheaper reduce communication and allow workers to do more without

[15] Qian (1994) presents an alternative hierarchical model with many layers using a theory of monitoring rather than knowledge-based hierarchies. Chen (2014) provides an equilibrium version of that model with heterogeneous firms.

relying on help, whereas technologies that make communicating information cheaper increase the hierarchy.

As expected, a decline in both $c$ and $h$ shifts down the cost function, making production more efficient. Obviously, $h$ only affects the cost function when the organization consists of more than just a layer of workers because in that case no communication is required. So both these changes are technological improvements. However, they have distinct effects on the equilibrium organizations use. A decline in $c$ makes knowledge cheaper and therefore less useful to build a complex organization.[16] After all, the whole point of the organization is to economize on the wage bill by using knowledge efficiently. Hence, the model predicts that such a decline would increase the scope of decision making by lower-level workers, increase the span of control of supervisors, increase the ratio of production workers to problem solvers, and reduce the number of layers of management.

In contrast, a decline in $h$ implies that the mechanism to economize on wages by using a complex organization is more effective, so more layers of managers are used. This makes some agents learn and earn less, as discussed above, but some others gain as they manage larger teams and can leverage their knowledge more. We discuss in Section 6 some evidence that ICTs do indeed have different effects.

### 3.3. Growth: Exploiting and Exploring Knowledge

The models above study the hierarchical organization from a timeless perspective. Of course, in reality, hierarchies are set up slowly over time. Garicano & Rossi-Hansberg (2012) extend the model above to analyze how this slow evolution affects technology use and innovation. We differentiate between two knowledge-generating activities: exploiting existing technology and innovating to develop new technologies. First, exploitation takes place as organizations undertake production over time and add new layers (new referral markets) of experts. That is, we model this process as the sequential emergence of a collection of markets for expert services in which agents sell the problems they cannot solve to other agents. By allowing these new experts to leverage their knowledge about unusual problems, the new layers allow for more knowledge to be acquired and make production more efficient under the current technology. This process exhibits decreasing returns, as eventually most problems are well known, and the knowledge acquired is increasingly less valuable.

Second, innovation is the result of agents' decisions about how much to invest to create radically new technologies. This investment process exhibits convex costs so that the investment, if it happens, takes place smoothly over time. Of course, the economy's ability to exploit the new technology through organization determines the profitability of innovation investments. The rate of innovation, the extent of exploitation, and the amount of organization in the economy are jointly determined in the theory and depend on the cost of acquiring and communicating knowledge.

If it happens, progress in our model takes place in leaps and bounds. After a new technology is adopted, investment in innovation decreases, and agents concentrate on exploitation as they acquire first the more productive pieces of knowledge about this technology and then the rarer ones. Radical innovation will not take place again until the current innovation has been exploited to a certain degree. Both the timing of the switch to a new technology and the size of the jump in the technology are endogenous, as agents must choose how much and when to invest in radical innovation. As long as the value of continuing with an existing technology is sufficiently high, the

---

[16]In fact, for the exponential case, the only relevant parameter is $c/\lambda$.

switch to the new technological generation does not take place. As in Arrow (1974), organizations are specific to a particular technology, so adopting the new technology makes the knowledge acquired about the previous technology obsolete. It requires agents to start accumulating new knowledge and start building new organizations. Hence, inherent in new knowledge is a process of creative destruction (Schumpeter 1942) in which adopting a radical innovation makes the existing organization obsolete.[17]

Progress may also come to a halt if agents decide not to invest in radical innovation. Specifically, the payoff of exploiting existing technologies may be such that agents optimally create very large organizations, comprising a large set of referral markets and a large number of different specialized occupations. Such organizations take a long time to build, and thus agents choose to postpone forever the moment at which a new technology would be exploited. The radical innovation process never gets started. When the current technology is already well exploited, agents do not have any development of the alternative radical technology to build on and prefer to exert their efforts on small advances of the existing technology. The result is stagnation.

**3.3.1. Empirical implications: information and communication technology and economic growth.** ICTs determine the depth to which an innovation is exploited and thus the rate of growth. Consider first information technology. The main benefit of organization is that individuals can leverage their cost of acquiring knowledge over a larger set of problems, increasing the utilization rate of knowledge. Information technology advances (e.g., databases) decrease the need for organizational complexity, shorten the exploitation process, and unambiguously increase growth.

Better communication technology unambiguously increases welfare but has an ambiguous effect on growth: For intermediate communication costs, and thus for intermediate costs of exploiting a particular technology, organizations may get stuck in an old technology forever. The intuition is quite clear. When communication technology is expensive, organizations do not grow very large, so the amount of organizational learning for a given technology is small; thus, organization does not give old technologies a considerable advantage over new technologies. Similarly, when communication technology is inexpensive, technologies are extensively exploited via organization, so technological improvements are very valuable. Thus, switching to a new technology is attractive and frequent. For intermediate values of communication costs, large organizations are useful for exploiting a technology, and the potential improvement in technology might not be valuable enough to induce switching. Hence, making organizations more efficient, by reducing communication costs from high to intermediate levels, shifts the balance of economic activity from investing in new innovations to exploiting more fully existing innovations, and this may reduce long-term economic growth, potentially all the way down to zero.

## 4. ORGANIZATION WITH EX ANTE HETEROGENEOUS AGENTS

### 4.1. Sorting and Inequality

We introduce here organizations that utilize the skills of heterogeneous workers and retake, more formally, the issue of wage inequality. We start with the simplest possible model (Garicano & Rossi-Hansberg 2004), one in which hierarchies have only two layers, and agents have

[17]Previous models of creative destruction, following the pioneering work by Aghion & Howitt (1992) and Grossman & Helpman (1991), do not take organizations into account—new products substitute for the old, but organizations play no role.

exogenously given knowledge or, equivalently because here it is exogenously given, skill. This model generates the results described previously in Section 2.2.

A firm has $n$ low-skill agents (workers), with skill $z_p$, and one high-skill agent (a manager), with skill $z_m$. Agents are organized in a knowledge hierarchy as explained above. Specifically, the $n$ workers attempt to solve one problem each. If they fail, they request help or directions from the manager and reattempt their problem using those directions. Normalize $z$ so that the distribution of problems $F(z)$ is uniform in [0, 1].[18] Let the distribution of skill in the population be given by $\phi(z)$.

The output produced by a team is then simply $z_m n$, where $z_m$ is the knowledge or skill of its manager. This is subject to the time constraint of the managers, which, as in Section 3 above, is given by $hn(1 - z_p) = 1$. The span of the manager is limited by the knowledge of production workers; if production workers are more knowledgeable, they will need help less often, and managers will be able to supervise larger teams.[19] A manager with skill $z_m$ maximizes his or her earnings $R = z_m n - w(z_p)n$ by his or her choice of the skill of the workers he or she employs, namely,

$$R = \max_{z_p} \frac{z_m - w(z_p)}{h(1 - z_p)}. \tag{8}$$

Solving for an equilibrium in this economy is a continuous assignment problem (see Sattinger 1993) with two twists relative to standard assignment problems. First, who is assigned to whom is not a given but an equilibrium outcome. In standard assignment models, this identity of the agents is assumed (e.g., by assuming one-to-one matching of prespecified masses of agents with distinct types, or between agents and machines, industries, etc.). In contrast, here we are marrying a mass of workers with a mass of managers—where those roles and masses are not given by assumption. This property is the result of integrating an occupational choice, in which agents can decide freely what job to take, and many-to-one matching, in which the span of control of managers is an equilibrium outcome itself. Second, agents can decide not to be matched and produce on their own; their output in that case is just $F(z) = z$.

Optimality requires positive sorting; that is, workers with more skill must be assigned to managers with more skill. The reason is that the production function is supermodular in both skills. A better manager must leverage his or her higher knowledge over a larger number of workers, and this requires workers to be more skilled so that they can deal with more problems on their own. Technically, this supermodularity is guaranteed by a positive cross derivative of output $z_m/(h(1 - z_p))$ with respect to $z_m$ and $z_p$.

To characterize the equilibrium in this economy, we need to describe three objects: the allocation of agents to occupations (i.e., production workers, managers, and self-employed), the team composition (i.e., the matching between workers and managers and the optimal spans of control of managers), and the earnings function.[20]

Note that the slope of the wage equation must be such that the first-order condition holds. So for all workers ($z_p \leq z^*$, as we show below), one obtains

[18] This normalization is not in the original exposition in Garicano & Rossi-Hansberg (2004) but simplifies the exposition.

[19] Note that there are no decreasing returns to scale in this production function directly, but only through the congestion in the time constraint of the manager. We thus have a trade-off between congestion and skill: A manager can increase the team size if he or she so desires but needs to hire more skilled workers to do so. This implies that matching matters.

[20] Occupational choice works as in Lucas (1978) and Rosen (1982), but here there is a matching function that, together with the wage function, is pinned down in equilibrium.

$$w'(z_p) = hn(z_p)(z_m - w(z_p)) = \frac{z_m - w(z_p)}{1 - z_p} \text{ for all } z_p \leq z^*. \tag{9}$$

Together with a boundary condition, Equation 9 determines a unique wage function, $w(z)$, specifying the wage of each individual $z_p$ for a given match $(z_p, z_m)$. The wage slope for those agents who do not join in a hierarchy (self-employed) is given by $w_I'(z) = 1$. Clearly, one finds that $w'(z_p) < w_I'(z_p) = 1$ for all workers [because for agents who are workers $w(z_p) > z_p$, so $w'(z_p) < (z_m - z_p)/(1 - z_p) < 1$]. Using the envelope theorem, we find that the earnings function of managers satisfies

$$R'(z_m) = n(z_p) > 1 = w_I'(z_m). \tag{10}$$

This ordering of the slopes implies that an equilibrium with self-employed agents will be characterized by a pair of thresholds $(z^*, z^{**})$ such that all agents with skill $z < z^*$ become production workers, all agents with skill $z > z^{**}$ become managers, and those in between are self-employed. Furthermore, optimal occupational choice requires that for a marginal worker, $w(z^*) = z^*$; for a marginal manager, $R(z^{**}) = z^*$; and $R(1) = 1/h$. Equations 9 and 10, together with these three boundary conditions, determine the full earnings function and one threshold, for example, $z^*$.

Consider now the labor market clearing condition. Again, here the fact that one-to-many worker–manager matching takes place in equilibrium makes a difference. Instead of summing over all workers and over all managers, we need pointwise equalization of supply and demand. Let $m(z_p)$ denote the skill of the manager assigned to workers with skill $z_p$, which by positive sorting is an increasing function. The labor market clearing condition is then given by

$$\int_0^{z_p} \frac{1}{n(z)} \phi(z) dz = \int_{z^{**}}^{m(z_p)} \phi(z) dz \text{ for all } z_p \leq z^*. \tag{11}$$

Because Equation 11 holds for all $z_p \leq z^*$, and $n(z) = 1/((1 - z)h)$, we can differentiate with respect to $z_p$ to obtain $m'(z_p) = h(1 - z_p)$, which, together with $m(0) = z^{**}$ and $m(z^*) = 1$, determines the equilibrium assignment function $m(z)$ and the remaining threshold $z^{**}$.[21]

**4.1.1. Empirical implications: superstars and their shadow.** The model provides some implications also found in Lucas (1978) and Rosen (1982). The main one is the correspondence between the firm size distribution and distribution of managerial earnings: more specifically, the implication that managerial wages grow faster than ability thanks to the scale effects generated by larger team sizes (although here workers' wages also grow more slowly with skill). The theory also provides some novel implications. It pins down, for the first time we believe, the occupational distribution in the economy: The less skilled agents are production workers, the medium-skilled agents are independent or self-employed, and the most skilled agents are managers or problem solvers. Moreover, the composition of each team is also pinned down. It is not just the case that more skilled managers have more subordinates, but those subordinates are also of better quality. This matching in turn has implications for the wages of workers. As shown below, all three

[21]An equilibrium with only two layers, as well as a positive mass of self-employed agents, exists only in part of the parameter space. For example, $h$ needs to be sufficiently large (see Garicano & Rossi-Hansberg 2004 and Antràs et al. 2006 for details). Fuchs et al. (2015) solve the (equivalent) planner's problem, which has some additional measure of theoretic complications.

properties (stratification, positive sorting, and skill-scale effects) are empirically tested in Garicano & Hubbard (2012).

As discussed extensively but informally in Section 2.2, a reduction in communication costs has the following consequences. First, it increases firm size for all managers, as team size satisfies the manager's time constraint. Second, this leads to an increase in the managerial superstar effect because the marginal return to managerial talent is, analogously to Lucas (1978), proportional to team size, $n(z_p)$. Third, and more unique to our framework, it reduces wage inequality among workers, as a result of the sorting effect. Intuitively, for any two workers $z$ and $z'$, the skills of the managers with whom they are matched are closer to one another the larger the team sizes are [formally, $w''(z) = h$].[22]

#### 4.1.2. Empirical implications: organizations and power laws.

Geerolf (2014) develops a version of the theory described above but with exogenous skills and multiple layers. He shows that the distribution of spans of control for agents in firms with $L$ layers is a Pareto distribution with coefficient $2^L/(2^L - 1)$. Hence, under minimal assumptions on the distribution of skills, as the number of layers increases, the coefficient of the distribution of spans of control converges to 1. Using the empirical methodology in Caliendo et al. (2015) to identify layers (described in Section 6), he can account extremely well for the distribution of firm sizes with different numbers of layers in the French data. The key insight from this work is that the matching process with complementarities that we have discussed here can magnify even tiny skill differences and lead to Pareto distributions both in wages and in firm sizes.

### 4.2. Endogenous Skill and Sorting

The previous setting abstracted from a variety of dimensions to gain tractability and cleanly deliver our results. Chief among these abstractions is that knowledge is exogenous, which forces us to abandon the critical distinction between ICT. Furthermore, it simplifies the organization by determining exogenously that teams can have, at the most, a layer of managers and one of workers. Garicano & Rossi-Hansberg (2006) analyze the full-blown model in which agents are ex ante heterogeneous, they can acquire knowledge endogenously, and they can form teams with an arbitrary number of layers.

Garicano & Rossi-Hansberg (2006) assume agents differ in their cognitive ability or skill so that higher-ability agents incur lower learning costs. Specifically, the distribution of ability in the population can be described by a continuous density function, $\alpha \sim \phi(\alpha)$, with support in $[0, 1]$.[23] In particular, ability is defined so that the cost of learning to solve an interval of problems of length 1 is given by[24]

$$c(\alpha;\, t) = t - \alpha. \tag{12}$$

The cost of knowledge acquisition is then $c(\alpha; t)z$. As required for comparative advantage, high-ability types have a comparative advantage in knowledge acquisition. A decrease in $t$ represents an

[22]We discuss this further in Section 5, where we also present the exact mathematical expression for a case that can be solved in closed form.

[23]Consistent with the notation in the previous sections, we reserve $z$ for knowledge and introduce $\alpha$ for the exogenous skill, or ability, of an individual.

[24]The linearity of the learning cost function $c(\cdot)$ in $\alpha$ and the restriction of the support to $[0, 1]$ are without loss of generality, as we can always scale $\alpha$ to fit these restrictions.

improvement in information technology that decreases the cost of learning (e.g., a technology that decreases the cost of accessing knowledge, such as cheaper database storage and search).

The problem is somewhat tricky to solve, as the knowledge acquisition, matching, and number of layers (multiple hierarchies coexist) are endogenous in equilibrium. A given agent, with a given skill, could well be an entrepreneur, a middle manager, or a worker and in each case would be matched with entirely different other agents. We do not go over the details of the construction of the equilibrium; the reader is referred to Garicano & Rossi-Hansberg (2006). It suffices to say that the problem can be solved recursively. First, the first-layer managers choose the knowledge of workers; then the second-layer managers choose the knowledge of first-layer managers, etc. Finally, entrepreneurs or CEOs (the agents for whom it is not profitable to pass problems to other experts) choose the knowledge of the managers one layer below, as well as their own knowledge.

Garicano & Rossi-Hansberg (2006) characterize the optimal organization as follows. First, with regard to equilibrium properties, the equilibrium is unique, and it involves positive sorting so that more skilled workers are matched with more skilled managers and more skilled entrepreneurs. The equilibrium wage function, $w(\alpha){:}[0,1]\rightarrow\mathbb{R}_+$, is increasing and convex. Furthermore, the knowledge function $z(\alpha){:}[0,1]\rightarrow[0,1]$ is increasing. Occupational stratification, as described above, still holds. The lowest skilled agents are workers, then the next ones are middle managers at progressively higher layers, and the top agents in the economy are entrepreneurs. Second, regarding knowledge substitution, organization allows for the substitution of the knowledge of less skilled workers for that of their supervisors. Thus, organization decreases the knowledge (and therefore the value) of less skilled agents. In contrast, organization increases the return from learning about difficult tasks (as the knowledge can be leveraged more). Thus, organization increases the knowledge acquired by more skilled agents. As a result, organization decreases the marginal value of worker skill and increases the marginal value of entrepreneur skill. Finally, with regard to wage inequality and organization, it follows that organization makes low-ability agents (and, in particular, workers) more equal, thereby reducing earnings inequality within this group. In contrast, it makes high-skill agents more unequal, thereby increasing earnings inequality within this group. Finally, it increases the gap between low- and high-skill agents, which results in increases in earnings inequality between these groups.

**4.2.1. Empirical implications: information and communication technology, wage inequality, and knowledge substitution.** The framework above allows for a separate analysis of each form of ICT on wage inequality. Communication technology ($h$) increases inequality at the top and reduces it at the bottom, as emphasized above. The introduction of endogenous knowledge accentuates this effect through the knowledge substitution effect, in which the knowledge of subordinates is substituted by that of their supervisors. In contrast, information processing and knowledge acquisition technology (parameterized by $t$) increase within-class inequality for both low- and high-skill agents. Everyone acquires more knowledge, which creates larger within-class differences. The effects on between-class wage inequality are thus ambiguous, but technological change unambiguously decreases firm size, increases the knowledge (empowerment) of workers, and reduces layers of management. Garicano & Rossi-Hansberg (2006) argue that the timing of innovation in these technologies is consistent with the evidence on the temporal evolution of firm sizes, wage inequality, and CEO/worker earnings ratios.

## 4.3. Contracting with Double-Sided Adverse Selection

In many of the previous discussions, we implicitly, and sometimes explicitly, equate hierarchies with firms. Of course, nothing in our setting requires the exchange of knowledge in a hierarchy to take place within a firm. Garicano & Rossi-Hansberg (2006) show that, in a world with perfect information, consulting and referral markets can support the exact same equilibrium outcomes.

The above allocations put a high informational burden on agents. Absent some a priori knowledge about the persons conducting the exchange, both the difficulty of the questions posed to consultants and the knowledge of those consultants are hard to assess. A consultant could pretend to be more knowledgeable than he or she really is, and the agent asking for help could also misrepresent the difficulty of the question.

Such a setting with double-sided adverse selection is studied by Fuchs et al. (2015). They show that in the presence of these frictions, spot contracting is not efficient. However, they demonstrate that an ex ante, firm-like contractual arrangement uniquely delivers the first best outcome. This arrangement involves hierarchies in which consultants are full residual claimants of output and compensate producers via incentive contracts.

This characterization of the optimal ex ante contract looks very much like some firms in the knowledge-intensive sector (e.g., in professional services). There, the highest skilled agents hold all the equity and hire a number of less knowledgeable associates who are paid a combination of a fixed wage and a bonus for performance. Thus, the existence of these double-sided information problems shows why these knowledge-based hierarchies are often constituted as firms.

Relatedly, Acemoglu et al. (2014) consider the problem of crowdsourcing, in which a principal who owns a set of unsolved problems contracts with agents (either his or her employees or external contractors) who attempt to solve these problems. They show that the first best allocation of talent, which involves an endogenous hierarchy of problem solvers, is implementable regardless of informational asymmetries (see Fuchs et al. 2015 for a comparison of both approaches).

## 5. INTERNATIONAL ORGANIZATIONS: TRADE AND OFFSHORING

The theories above are useful to study a variety of issues concerning international trade and offshoring (the formation of international organizations and production processes). We have studied these issues in a variety of papers, including Caliendo & Rossi-Hansberg (2012), for the case of international trade, and Antràs et al. (2006, 2008), for the case of offshoring.[25]

### 5.1. Trade and Organization

Assessing the impact of the large increases in international trade that the world has experienced in the past several decades involves understanding the effect that reductions in trade costs have had on the characteristics of producers and their labor force. Bernard & Jensen (1997, 1999) initiated a large literature by pointing out the distinct characteristics of exporters relative to other firms. Melitz (2003) then provided the workhorse model to explain these observations using a mechanism that selected the best firms as exporters as the result of fixed exporting costs. In his model, organization does not play a role, so the marginal cost of a firm remains constant as it starts to export. A firm is a technology that is immutable to the markets in which it sells.

This is not the case once we introduce organization into the problem. Parallel to Section 3.2, as an exporter grows, for example, owing to a trade liberalization, it will do so by adding workers,

[25]Gumpert (2014) provides a discussion of multinational firms and how they determine the hierarchical organization of headquarters and affiliate plants. Marin & Verdier (2010, 2008) and Marin et al. (2014) use the monitoring-based theory of organization to study offshoring and the organization of international production. Antràs & Rossi-Hansberg (2009) present a more general review on the large literature on organization and international production processes.

sometimes layers, and will modify the knowledge and wages of its labor force accordingly. Furthermore, it will increase its productivity (but reduce its revenue-based productivity) as a result of that reorganization. The reverse is also true for nonexporters that tend to shrink or exit as a result of a bilateral trade liberalization, and will reduce their productivity.

The above reasoning implies that international trade, by changing the optimal production size of firms and therefore their organizational form, affects the efficiency of production of individual firms as well as aggregate productivity. Caliendo & Rossi-Hansberg (2012) show that the welfare gains from trade are declining in the communication costs, $h$, but increasing in the information costs, $c$. Trade allows firms to grow larger by selling to more markets. The larger scale makes using a large, more complex, organization economical. This channel leads to additional gains from trade. These additional gains are larger the more efficient is the organizational technology (namely, the lower are communication costs). In contrast, the additional gains are smaller if the costs of acquiring knowledge are smaller because, as underscored above, the purpose of organization is to economize on these costs. In fact, gains from trade converge to the ones in a standard Melitz model as $c$ tends toward zero because in that case the organizational problem disappears.

## 5.2. Offshoring and Organization

Antràs et al. (2006) address a different problem related to globalization: the formation of international production teams and its implications. Feenstra & Hanson (1997) and Goldberg & Pavcnik (2007) show that globalization has led to increases in wage inequality among workers in less developed countries. This has been viewed as puzzling in the international trade literature because trade, either in goods or in tasks, should lead to increases in the price of the abundant factor, as dictated by the standard Heckscher-Ohlin logic.[26] This is, however, not necessarily the case once we analyze the matching between workers in global production teams.

The model in Antràs et al. (2006) is simple enough to be solved analytically. Consider the model used in Section 4.1 but with a uniform distribution of skills between 0 and $u$ and the assumption that parameters are such that there are no self-employed agents. In this case, we can solve the system of two differential equations explicitly. The wage function for workers (agents with $z \in [0, z^*(h, u)]$) is given by

$$w(z) = z^*(h,u) - \sigma(h,u)(1-z) + \frac{1}{2}hz^2. \tag{13}$$

The skill premium is just the derivative of Equation 13 with respect to $z$, namely,

$$w'(z) = \sigma(h,u) + hz. \tag{14}$$

The skill premium has two parts. The first term is the baseline price of a unit of skill and is determined by the supply and demand of skills in this economy. This is what we call the competition effect. The second term is increasing in the number of units of skill an individual possesses and increases with communication costs: the complementarity effect. The source of this effect is that having more units of skill allows the workers to be matched to better managers and earn more per unit of skill. As discussed above, this effect declines as teams get larger (which results from an

[26]Grossman & Rossi-Hansberg (2008, 2012) analyze in detail the factor price implications of reductions in task trading costs in frameworks without hierarchical organization.

improvement in communication costs). In the limit, as $h$ goes toward zero, every agent works with the best manager, and this source of inequality disappears.

Globalization allows this economy to form international production teams with other economies that might have different distributions of skills. Call the economy above, with a distribution of skills between 0 and $u$, the South. To economize on parameters, let the North be identical but with $u = 1$.[27] The result is a world economy with a mass of skills equal to 1 for $z > u$ and equal to $1 + 1/u$ for $z < u$.

The ability to form international production teams leads workers in the South to match with better managers in the North, which increases the price of a unit of skill in the South. Furthermore, for the worst workers, the complementarity effect increases as well because of the larger density of workers relative to managers. Finally, more Southern agents become workers, so there is a compositional effect that increases workers' wage inequality. Hence, consistent with empirical observations, wage inequality for Southern workers unambiguously rises. Firms in the South exit as a result of the formation of international teams.

In the North, the formation of international teams also leads to an increase in wage inequality among managers. Among workers, the baseline price of skill decreases (the relative number of unskilled workers in the world is larger than in the North), but they face the same enhanced complementarity effect. Thus, in the North, if communication technology is good enough so the first effect dominates, globalization can lead to a decrease in wage inequality at the bottom together with an increase in inequality at the top. This is exactly what happened in the United States in the past few decades, as discussed in the context of **Figure 1**. So superstars, and the shadow they cast on workers, can also be the result of reductions in offshoring costs.

## 6. EMPIRICS OF HIERARCHIES

The importance of the hierarchical structure of organizations for worker, firm, and aggregate outcomes is not obvious. Even though the theory discussed above can forcefully make the claim that worker, firm, and economy-wide performance can be affected by the way agents organize in production teams, it could be the case that in reality firms just expand by replicating their operations. The returns to an individual's skill and knowledge could be independent of the other agents he or she works with. That the models above are consistent with the facts presented in Section 2 provides only indirect evidence in favor of these theories. We discuss below some more direct evidence that fails to falsify the mechanisms we have emphasized.

### 6.1. Firm Growth, Reorganization, and Wages

Caliendo et al. (2015) use French firm- and worker-level data to analyze the relationship among firm characteristics, firm growth, and a firm's organization. To do so, they use wages as their measure of worker skill and/or knowledge and divide the firm into layers using hierarchical occupational definitions. The hierarchical definitions identify a maximum of four layers per firm.[28]

[27]Here we only study the implications from offshoring to a particular country. Antràs et al. (2008) study the role of local communication technology and the distribution of skills in making agents in target countries more attractive as production partners.

[28]Caliendo et al. (2015) use a match of the Bénéfice Réel Normal data set, which includes firm balance sheet information, and the Déclarations Annuel des Données Sociales data set, which includes worker-level occupation, hours of work, and earnings of salaried employees.

These data allow the authors to create a picture of the organization of French manufacturing firms between 2002 and 2007. Importantly, their study uses a large institutional data set instead of more specialized data sets of a few firms, as in previous studies.[29] The authors can follow firms, and the wages they pay in each occupation, as they add and drop layers over time (namely, as they hire workers in a hierarchical occupation in which they did not employ anyone before). **Figure 5** presents the resulting representative hierarchies, divided by the number of layers, for 553,125 firm-year observations. The results are quite stark. Firms with more layers are larger in terms of value-added and total hours of work employed and pay higher hourly wages on average. Furthermore, a vast majority of them are hierarchical in that those with higher layers employ less hours and pay higher hourly wages.

The results also show that transitions between layers are quite common, but mostly occur by adding or dropping one layer. Furthermore, the probability of adding (dropping) a layer is increasing (decreasing) in a firm's value-added. Caliendo et al. (2015) also show that when firms expand or contract substantially, they tend to reorganize by adding or dropping layers. Dropping and adding layers are also associated with internal changes in the characteristics and wages of the employees of the firm. **Figure 5** shows that firms with only one layer pay their labor force on average 27.17 2005 euros per hour. In contrast, firms with two layers pay the workers at the bottom only 18.15 euros and firms with three layers pay only 16.91 euros. That is, as firms add layers, they hire workers in pre-existing layers that earn on average less. This is the case for all comparisons in **Figure 5**. Of course, these figures average over firms that are heterogeneous in a variety of dimensions, so these results might be generated by compositional effects. In their paper, Caliendo et al. (2015) show that the reduction in average wages in all pre-existing layers is present for all transitions in firms with any number of layers when one uses time-series evidence with firm fixed effects. Furthermore, as the theory predicts, this reduction (increase) in wages is only present when firms add (drop) layers and not when they expand or contract without reorganizing. In that case, as the theory above predicts, average wages in a layer grow when the firm expands and decline when it contracts.[30]

Overall, these results show that layers of management vary systematically with firm characteristics. Firms seem to actively manage their production hierarchy. It is important then to condition on reorganization when analyzing changes in firm characteristics. Moreover, the impact of any economy-wide change that affects firm size will be mediated by these reorganizations. For example, the routinization of work (as firms grow and add more layers) reduces the knowledge and wages of workers in the middle of the wage distribution the most. Look at the top managers in four-layer firms in **Figure 5**, who make on average 87.66 euros per hour. Their direct subordinates make only 43.6 euros rather than the 57.43 euros agents in the same layer make in three-layer firms.

## 6.2. Sorting, Spans of Control, and the Distribution of Wages

Several papers have studied the effect of larger spans of control on wages. Both Smeets & Warzynski (2008) and Fox (2009) find a positive effect of the number of subordinates, and the level of

[29]The previous literature on organizational form has used very detailed information on a small set of firms. The best examples are Baker et al. (1994) and Baker & Holmstrom (1995), who analyze one firm. Other studies such as Rajan & Wulf (2006) and Guadalupe & Wulf (2010) employ data sets with a few hundred firms. Caroli & Van Reenen (2001) use larger surveys of English and French firms but do not focus on the hierarchical structure of firms.

[30]Tåg (2013) replicates many of these qualitative results for Swedish firms, and ongoing projects have also replicated them for Danish and Portuguese firms.

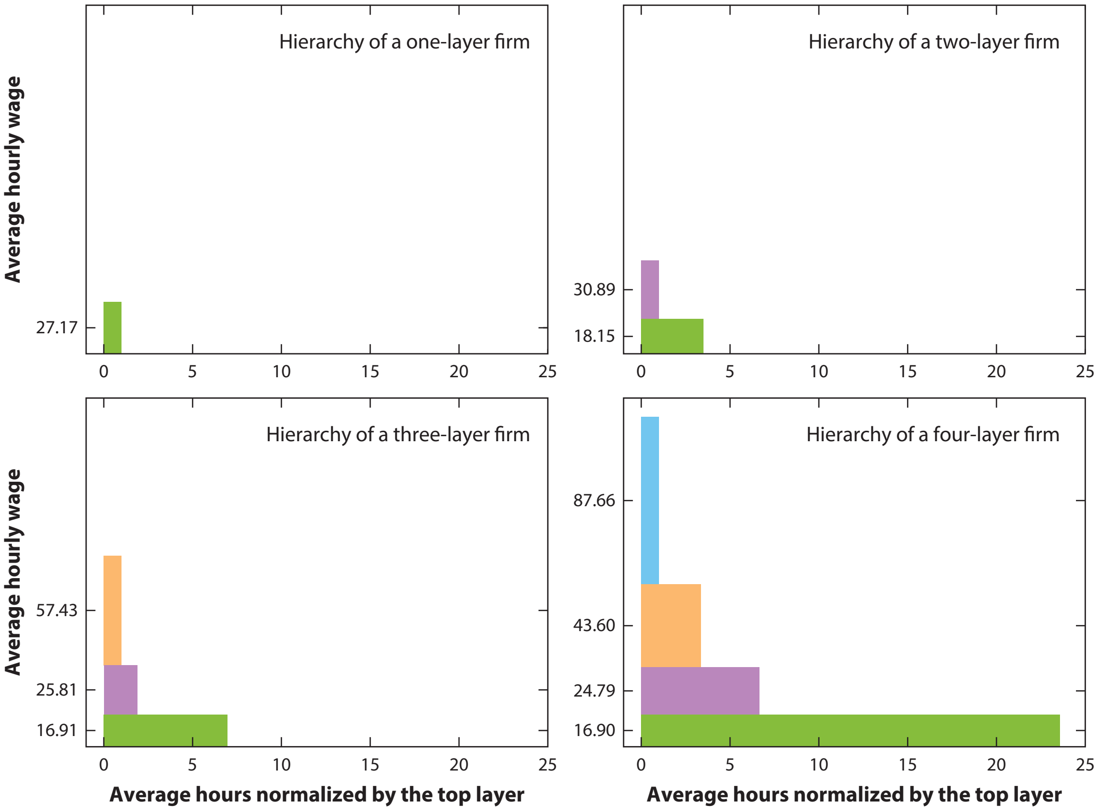


**Figure 5**

Representative hierarchies in France. The figure shows the average wage in 2005 euros in each layer on the vertical axis and the number of hours, normalized by the number of hours in the top layer, on the horizontal axis. Figure reproduced from Caliendo et al. (2015), figure 5.

responsibility, on wages, the former for a large European firm and the latter for Swedish and US firms [see also the classic study by Brown & Medoff (1989) that relates firm size and wages]. Using 1992 census data for US law firms,[31] Garicano & Hubbard (2007, 2012) compare earnings among individuals who ranked the same hierarchically and show that those who work with more associates per partner (i.e., more workers per manager) earn more.

Garicano & Hubbard (2012) go further and present evidence that associate earnings are higher at offices in which partner earnings are higher. Although this does not necessarily reflect that associates' and partners' abilities are positively correlated (it could be driven by office-level demand shocks), this evidence is certainly consistent with equilibria with positive assortative matching. The authors also find that lawyers are stratified as in the model in Section 4.1. Namely, independent lawyers' earnings are in between those of associates and partners in large law firms.[32]

[31]The data cover the entire legal industry in the United States from the 1992 Census of Service Industries, which had a questionnaire specific for law firms.

[32]The exception involves very large cities in which associates of large law firms sometimes can earn more than independent lawyers, which is perhaps a sign that in large cities law firms have more than two layers, with some associates operating in middle-rank positions.

## 6.3. Information Technology and the Organization of Work

Bloom et al. (2014) aim to separate the two distinct components of ICT along the lines of our work here. Their test involves a key prediction of the theory: Lower information costs shift decisions to lower layers, whereas lower communication costs shift them to higher ones. The authors test this using decisions shifted both from the bottom to the middle of the hierarchy and from the middle to the top. They rely on a proprietary firm survey on management practices, which they merge with private sector data on the use of ICT by 1,000 firms. Although they do not make any causal claims for their mechanism, they indeed find that network technologies (a communication technology) result in autonomy shifting from plant managers up to the center and from workers up to plant managers. In contrast, the two information technologies they study lead to more decentralized decision making. This is the first paper that, thanks to the rich data available, can empirically evaluate the direct implications of the knowledge-based hierarchy model concerning the different effects of ICT.[33]

# 7. THE ROAD AHEAD

We argue in this article that understanding equilibrium outcomes through the lens of theories that put at their core knowledge-based hierarchies can illuminate a variety of economic phenomena. The way in which individuals organize in teams to produce determines the productivity of these teams, as well as the rewards that individuals can command for their work. Once we view economics through this organizational lens, the evolution of technology, and its distinct forms, takes on a different meaning. Communication technologies enhance the hierarchy by making it a better technology to economize on knowledge. Information technology makes knowledge cheaper and therefore organization less useful. We argue above that the evolution of wage inequality in the different parts of the income distribution can be well understood as a response to the important changes in ICT we have observed in the past few decades. Communication technology is highlighting the advantages of superstars and is making the less skilled more equal, thereby hurting the middle class. This is what we have termed the shadow of superstars.

Of course, much work is still needed to develop the core theories, implications, and empirical evidence that ensue from this organization-based view of economics and, specifically, the view centered on knowledge-based hierarchies. Particularly promising, we believe, are three avenues. The first one involves dynamics and economic development. Section 3.3 outlines some of our work on the topic, but much more is needed. In particular, we need theories in which agents decide on the timing of the development of heterogeneous hierarchies. Furthermore, we need to link hierarchies with the career paths of individuals and see whether progressive promotions within and across organizations can explain the facts on individual wage dynamics over the life cycle.[34]

Second, these frameworks should be used to evaluate a variety of tax policies, and other distortions, that affect the labor market, firm characteristics, and international transactions. The effect of firm size distortions, as studied by Garicano et al. (2013), and in the context of knowledge-based hierarchies by Torres-Coronado (2014), is a good example. There is room for much more work analyzing the effect of minimum wages, other forms of corporate taxation and subsidies,

[33]Studying the introduction of information technology–based credit scoring models in a Colombian bank, Paravisini & Schoar (2012) also find evidence that improvements in information technology decentralize decisions by making managers dedicate more time to the hardest decisions.

[34]Some papers, such as Roys & Seshadri (2014), are starting to make progress on this front.

place-based policies that affect the spatial location of firms and therefore their organization, education and training subsidies, and many others.

Finally, a lot more empirical work is needed. The evidence in Section 6 establishes the basic empirical credibility of this approach. But much more empirical work is required to understand the effect that different economic shocks have on the organizations that we see in equilibrium and the implied distribution of wages. In particular, we need causal evidence of the effect of specific shocks, such as trade liberalizations, policy reforms, or idiosyncratic firm-specific demand and productivity shocks, on organization. We also need better evidence of the effects that the large changes in technology, and specifically ICT, have had on the actual changes in organizations and labor market outcomes that we have observed. Can we establish an unequivocal causal link? Can we assess the magnitudes of these effects?

One could continue this list, of course. Can these organization-based theories explain the productivity gap between firms in developed and developing countries? Can they explain business cycle issues such as slow recoveries and asymmetric cycles? We think that they can and that they will, but we are still far from getting there. In writing an article such as this one, one is always left somewhat encouraged by the road traveled but also awed by the enormity of the task ahead. Undoubtedly, we focus unfairly on the part of this endeavor that we have contributed to. We apologize for this and encourage others to actively participate in this agenda.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We thank Pol Antràs, Lorenzo Caliendo, Bob Lucas, Ferdinando Monte, and Steve Redding for useful comments. Ricardo Reyes-Heroles provided excellent research assistance.

## LITERATURE CITED

Acemoglu D. 1999. Changes in unemployment and wage inequality: an alternative theory and some evidence. *Am. Econ. Rev.* 89:1259–78

Acemoglu D. 2002. Technical change, inequality, and the labor market. *J. Econ. Lit.* 40:7–72

Acemoglu D, Autor D. 2011. Skills, tasks and technologies: implications for employment and earnings. In *Handbook of Labor Economics*, Vol. 4, ed. D Card, O Ashenfelter, pp. 1043–171. Amsterdam: North-Holland

Acemoglu D, Mostagir M, Ozdaglar A. 2014. *Managing innovation in a crowd*. NBER Work. Pap. 19852

Aghion P, Howitt P. 1992. A model of growth through creative destruction. *Econometrica* 60:323–51

Antràs P, Garicano L, Rossi-Hansberg E. 2006. Offshoring in a knowledge economy. *Q. J. Econ.* 121:31–77

Antràs P, Garicano L, Rossi-Hansberg E. 2008. Organizing offshoring: middle managers and communication costs. In *The Organization of Firms in a Global Economy*, ed. E Helpman, D Marin, T Verdier, pp. 311–39. Boston, MA: Harvard Univ. Press

Antràs P, Rossi-Hansberg E. 2009. Organizations and trade. *Annu. Rev. Econ.* 1:43–64

Arrow K. 1974. *The Limits of Organization*. New York: Norton

Autor D, Katz L, Kearney M. 2006. The polarization of the U.S. labor market. *Am. Econ. Rev.* 96:189–94

Autor D, Katz L, Kearney M. 2008. Trends in U.S. wage inequality: revising the revisionists. *Rev. Econ. Stat.* 90:300–23

Autor D, Levy F, Murnane R. 2003. The skill content of recent technological change: an empirical exploration. *Q. J. Econ.* 118:1279–333

Baker G, Gibbs M, Holmstrom B. 1994. The wage policy of a firm. *Q. J. Econ.* 109:921–55
Baker G, Holmstrom B. 1995. Internal labor markets: too many theories, too little evidence. *Am. Econ. Rev.* 85:255–59
Bernard A, Jensen JB. 1997. Exporters, skill upgrading and the wage gap. *J. Int. Econ.* 42:3–31
Bernard A, Jensen JB. 1999. Exceptional exporter performance: cause, effect, or both? *J. Int. Econ.* 47:1–25
Bloom N, Sadun R, Garicano L, Van Reenen J. 2014. The distinct effects of information technology and communication technology on firm organization. *Manag. Sci.* 60:2859–85
Brown C, Medoff J. 1989. The employer size-wage effect. *J. Polit. Econ.* 97:1027–59
Caliendo L, Monte F, Rossi-Hansberg E. 2015. The anatomy of French production hierarchies. *J. Polit. Econ.* In press
Caliendo L, Rossi-Hansberg E. 2012. The impact of trade on organization and productivity. *Q. J. Econ.* 127:1393–467
Calvo G, Wellisz S. 1978. Supervision, loss of control, and the optimum size of the firm. *J. Polit. Econ.* 86:943–52
Caroli E, Van Reenen J. 2001. Skill-biased organizational change? Evidence from a panel of British and French establishments. *Q. J. Econ.* 116:1449–92
Chandler A. 1977. *The Visible Hand: The Managerial Revolution in American Business*. Cambridge, MA: Belknap
Chen C. 2014. *Management quality, firm organization and international trade*. Unpublished manuscript, Univ. Hong Kong
Crémer J, Garicano L, Prat A. 2007. Language and the theory of the firm. *Q. J. Econ.* 122:373–407
Demsetz H. 1988. The theory of the firm revisited. *J. Law Econ. Organ.* 4:141–61
Feenstra R, Hanson G. 1997. Foreign direct investment and relative wages: evidence from Mexico's maquiladoras. *J. Int. Econ.* 42:371–93
Fox J. 2009. Firm-size wage gaps, job responsibility, and hierarchical matching. *J. Labor Econ.* 27:83–126
Fuchs W, Garicano L, Rayo L. 2015. Optimal contracting and the organization of knowledge. *Rev. Econ. Stud.* In press
Gabaix X, Landier A. 2008. Why has CEO pay increased so much? *Q. J. Econ.* 123:49–100
Garicano L. 2000. Hierarchies and the organization of knowledge in production. *J. Polit. Econ.* 108:874–904
Garicano L, Hubbard T. 2007. Managerial leverage is limited by the size of the market: theory and evidence from the legal services industry. *J. Law Econ.* 50:1–44
Garicano L, Hubbard T. 2012. Learning about the nature of production from equilibrium assignment patterns. *J. Econ. Behav. Organ.* 84:136–53
Garicano L, Lelarge C, Van Reenen J. 2013. *Firm size distortions and the productivity distribution: evidence from France*. Unpublished manuscript, London Sch. Econ.
Garicano L, Rossi-Hansberg E. 2004. Inequality and the organization of knowledge. *Am. Econ. Rev.* 94:197–202
Garicano L, Rossi-Hansberg E. 2006. Organization and inequality in a knowledge economy. *Q. J. Econ.* 121:1383–435
Garicano L, Rossi-Hansberg E. 2012. Organizing growth. *J. Econ. Theory* 147:623–56
Garicano L, Van Zandt T. 2013. Hierarchies and the division of labor. In *Handbook of Organizational Economics*, ed. R Gibbons, J Roberts, pp. 604–54. Princeton, NJ: Princeton Univ. Press
Geanakoplos J, Milgrom P. 1991. A theory of hierarchies based on limited managerial attention. *J. Jpn. Int. Econ.* 5:205–25
Geerolf F. 2014. *A static and microfounded theory of Zipf's law for firms and of the top labor income distribution*. Unpublished manuscript, Univ. Calif., Los Angeles
Goldberg P, Pavcnik N. 2007. Distributional effects of globalization in developing countries. *J. Econ. Lit.* 45:39–82
Goldin C, Katz LF. 2009. *The Race Between Education and Technology*. Cambridge, MA: Harvard Univ. Press
Grossman G, Helpman E. 1991. Quality ladders in the theory of growth. *Rev. Econ. Stud.* 58:43–61

Grossman G, Rossi-Hansberg E. 2008. Trading tasks: a simple theory of offshoring. *Am. Econ. Rev.* 98:1978–97

Grossman G, Rossi-Hansberg E. 2012. Task trade between similar countries. *Econometrica* 80:593–629

Guadalupe M, Wulf J. 2010. The flattening firm and product market competition: the effect of trade liberalization on corporate hierarchies. *Am. Econ. J. Appl. Econ.* 2:105–27

Gumpert A. 2014. *The organization of knowledge in multinational firms*. Unpublished manuscript, Univ. Munich

Ijiri Y, Simon H. 1964. Business firm growth and size. *Am. Econ. Rev.* 54:77–89

Katz L, Murphy K. 1992. Changes in relative wages, 1963–1987: supply and demand factors. *Q. J. Econ.* 107:35–78

Knight F. 1921. *Risk, Uncertainty and Profit*. Boston, MA: Hart, Schaffner and Marx

Kremer M, Maskin E. 1996. *Wage inequality and segregation by skill*. NBER Work. Pap. 5718

Lemieux T. 2006. Postsecondary education and increasing wage inequality. *Am. Econ. Rev.* 96:195–99

Lucas R Jr. 1978. On the size distribution of business firms. *Bell J. Econ.* 9:508–23

Manne H. 1965. Mergers and the market for corporate control. *J. Polit. Econ.* 73:110–20

Marin D, Schymik J, Tarasov A. 2014. *Trade in tasks and the organization of firms*. Unpublished manuscript, Univ. Munich

Marin D, Verdier T. 2008. Competing in organizations: firm heterogeneity and international trade. In *The Organization of Firms in a Global Economy*, ed. E Helpman, D Marin, T Verdier, pp. 142–72. Boston, MA: Harvard Univ. Press

Marin D, Verdier T. 2010. *Corporate hierarchies and international trade: theory and evidence*. Unpublished manuscript, Univ. Munich

Mayer T. 1960. The distribution of ability and earnings. *Rev. Econ. Stat.* 42:189–95

Melitz M. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71:1695–725

Paravisini D, Schoar A. 2012. *The information and agency effects of scores: randomized evidence from credit committees*. Unpublished manuscript, London Sch. Econ.

Piketty T. 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Belknap

Prescott EC, Boyd JH. 1987. Dynamic coalitions: engines of growth. *Am. Econ. Rev.* 77:63–67

Prescott EC, Visscher M. 1980. Organization capital. *J. Polit. Econ.* 88:446–61

Qian Y. 1994. Incentives and loss of control in an optimal hierarchy. *Rev. Econ. Stud.* 61:527–44

Radner R. 1993. The organization of decentralized information processing. *Econometrica* 61:1109–46

Rajan R, Wulf J. 2006. The flattening firm: evidence on the changing nature of firm hierarchies from panel data. *Rev. Econ. Stat.* 88:759–73

Rosen S. 1981. The economics of superstars. *Am. Econ. Rev.* 71:845–58

Rosen S. 1982. Authority, control, and the distribution of earnings. *Bell J. Econ.* 13:311–23

Roys N, Seshadri A. 2014. *Economic development and the organization of production*. Unpublished manuscript, Univ. Wisconsin–Madison

Sattinger M. 1993. Assignment models of the distribution of earnings. *J. Econ. Lit.* 31:831–80

Schumpeter J. 1942. *Capitalism, Socialism and Democracy*. New York: Routledge

Smeets V, Warzynski F. 2008. Too many theories, too few facts? What the data tell us about the link between span of control, compensation and career dynamics. *Labour Econ.* 15:687–703

Tåg J. 2013. Production hierarchies in Sweden. *Econ. Lett.* 121:210–13

Torres-Coronado J. 2014. *Size-dependent firm regulations and the return to skill in a many-to-one assignment model*. Unpublished manuscript, ITESM, Mexico City