

NBER WORKING PAPER SERIES

ORGANIZATION AND INEQUALITY IN A
KNOWLEDGE ECONOMY

Luis Garicano
Esteban Rossi-Hansberg

Working Paper 11458
<http://www.nber.org/papers/w11458>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2005

We thank the editor P.-A. Chiappori and two anonymous referees for very helpful comments. We have also benefitted from discussions with D. Acemoglu, F. Alvarez, K. Bernhardt, J. Crémer, W. Dessein, M. Dewatripont, D. Diamond, L. Einav, D. Ferreira, A. Gavilán, R. Gertner, R. Gibbons, R. Hall, T. Holmes, T. Hubbard, N. Kocherlakota, F. Kubler, K. M. Murphy, C. Prendergast, L. Rayo, R. Rogerson, T. Santos, I. Segal, N. Stokey, S. Tadelis, I. Werning and seminar participants at various institutions. We thank D. Card and K. J. Murphy for providing us with data on wages and on CEO pay respectively. We are particularly grateful to S. Rosen, on whose work this paper builds. Garicano: luis.garicano@gsb.uchicago.edu. <http://gsb.uchicago.edu/fac/luis.garicano>. (773) 8342862. 1101 E 58th Street, Chicago, IL 60637, USA. Rossi-Hansberg: erossi@stanford.edu. <http://www.stanford.edu/~erossi>. (650)7241427. 579 Serra Mall, Stanford, CA 94305-6072, USA. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2005 by Luis Garicano and Esteban Rossi-Hansberg. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Organization and Inequality in a Knowledge Economy
Luis Garicano and Esteban Rossi-Hansberg
NBER Working Paper No. 11458
June 2005
JEL No. D2, J3, L2

ABSTRACT

We present a theory of the organization of work in an economy where knowledge is an essential input in production: a knowledge economy. In this economy a continuum of agents with heterogeneous skills must choose how much knowledge to acquire and may produce on their own or in organizations. Our theory generates an assignment of workers to positions, a wage structure, and a continuum of knowledge-based hierarchies. Organization allows low skill agents to ask others for directions. Thus, they acquire less knowledge than in isolation. In contrast, organization allows high skill agents to leverage their knowledge through large teams. Hence, they acquire more knowledge than on their own. As a result, organization decreases wage inequality within workers, but increases income inequality among the highest skill agents. We also show that equilibrium assignments and earnings can be interpreted as the outcome of alternative market institutions such as firms, or consulting and referral markets. We use our theory to study the impact of information and communication technology, and contrast its predictions with US evidence.

Luis Garicano
Graduate School of Business
University of Chicago
1101 E 58th Street
Chicago, IL 60637
luis.garicano@gsb.uchicago.edu

Esteban Rossi-Hansberg
Stanford University
Department of Economics
579 Serra Mall
Stanford, CA 94305-6072
and NBER
erossi@stanford.edu

1 Introduction

Production of most goods and services depends on the skills of agents involved in production. Organizations allow individuals to combine and leverage their knowledge, thereby effectively expanding the time of able agents by allowing them to use the help of less able agents in production. Recent improvements in information technology have affected how knowledge is acquired and communicated, and thus how production is organized in the economy. Yet the idea of organizing to optimize the use of knowledge has been virtually absent from economic theory. In this paper, we study an economy where knowledge is an essential input in production, and where differences in individual skills to acquire knowledge are the main driver of earnings inequality. We explore how, in equilibrium, organization allows individuals with different skills to combine their knowledge, the roles that individuals play in production, and the associated structure of earnings.

We introduce a theory of a knowledge economy that has four main elements: production requires knowledge; individuals are heterogeneous in cognitive skill; communication is possible; and organization is hierarchical. On her own, an agent encounters one problem per unit of time and she only produces if she can solve it, which happens whenever the difficulty of the problem is below her level of knowledge. Problems are drawn from a known probability distribution. Agents can acquire knowledge at a cost, and decide optimally how much knowledge to acquire. They differ in how costly it is for them to acquire knowledge – in particular there exists a continuum of agents of different cognitive skill.¹ Agents can communicate at a cost and thus help each other solve problems. This allows for the formation of organizations in which agents specialize in either production or problem solving. Production workers draw problems and learn how to solve routine problems, while problem solvers learn to solve the more difficult problems, and contribute their knowledge to production workers as needed. Multiple layers of increasingly more skilled problem solvers may be needed in order to solve the hardest problems and effectively leverage the knowledge of the most skilled. That is, organizations take the form of ‘knowledge-based hierarchies.’² The

¹Our focus on individual skill heterogeneity as the basis for organizational heterogeneity is consistent with recent empirical work on the structure of wages. In particular, recent work using matched firm-worker data (e.g. Abowd, Kramarz and Margolis 1999) finds that ‘virtually all of the interindustry wage differentials is accounted for by variation in average individual heterogeneity between sectors.’ Moreover, the empirical literature on skill differentials has found cognitive skills to be a very good predictor of wage differentials (see Murnane, Willett and Levy, 1995).

²As in Garicano (2000) this organizational structure is optimal when matching problems and solutions is costly.

economy-wide problem is an assignment problem in which agents must acquire knowledge, solve problems for each other and be assigned to an occupation and to a particular layer of one of many possible hierarchical structures.

In this economy wages play an allocative role. Production is not only determined by the total amount of units of knowledge hired by a given team, but also by how these units are distributed among team members. The role of wages is then to provide agents with incentives to join particular production teams depending on their ability. The resulting earnings structure compensates agents for increases in talent more than proportionally. To see this, consider an individual that is more skilled than others, so that in autarchy she can earn more. Suppose now that she is allowed to help others solve their problems. Her earnings relative to others with less knowledge will be higher than in autarchy. First, holding knowledge constant, she spends a larger fraction of her time on the problems she knows and others do not – she leverages her knowledge over a larger amount of problems. Second, her knowledge about unusual problems is now used more often, thus raising the marginal value of learning and leading her to acquire more knowledge – further increasing her earnings.

An equilibrium in this economy exists, is unique, and is efficient. Moreover, it displays three features of particular empirical interest. First, it displays positive sorting, in the sense that higher ability agents share their knowledge with higher ability subordinates (production workers or lower level managers). The reason is that efficiency requires that the expertise of more skilled problem solvers be shielded from easy questions, and thus those posing questions to them must themselves be among the most skilled in their rank or layer. Second, earnings grow faster than individual's ability to solve problems. More skilled individuals solve problems for larger teams and are also assigned to higher layers in the hierarchy. This implies that hierarchies allow the more skilled agents to leverage their talent more, thereby increasing their returns to knowledge. Thus, organization accentuates the ability differentials between the most and least talented individuals. Third, regardless of the distribution of skills, individuals are segmented by cognitive skill in the sense that the original continuous agent skill set is partitioned in layers in a systematic way. In particular, the less skilled agents are production workers, the following agents by cognitive skill are first level problem solvers, the next are second level problem solvers, and so forth. That is, in the language of the matching literature, the economy does not display perfect segregation, where all those in the team have the precise same skill (e.g. Kremer 1993), but rather 'skill stratification' – those with similar skill are in the same position (technically, the skill set in an occupation is connected).

Our analysis shows that the equilibrium in our model results from three alternative decentralizations, corresponding to three existing institutions: first, hierarchies structured as firms with production workers and managers, where production workers deal with routine problems and managers with the exceptions;³ second, consulting markets, where those involved in production pay fees to more knowledgeable problem solvers for help in solving their problem; third, referral markets, where agents can sell the problems whose solution they do not know to other agents at a price that decreases in the difficulty of the problem, as revealed by the skill of the agents who tried to solve it but could not. Our analysis thus pins down the boundary of the problem-solving hierarchy, but not the boundary of the firm. We view the ability of the model to predict transactions of knowledge in hierarchies within firms, or across them, as a richness of the model, since this duality is also a feature of the real world.⁴

Our model sheds some light on the impact of information technology (IT) on wages and organization. The analysis distinguishes two aspects of IT: its impact on the cost of communication among agents (e.g., e-mail and mobile technologies) and its ability to improve access to stored information through improvements in data storage and searching (e.g., decreases in the cost of information processing). We show that these two different aspects of information technology affect wage inequality and organization in different ways.

First, decreases in the cost of communication lead teams to rely more on problem solvers, increasing the centralization of the economy – more problems are solved at the top of the hierarchy. In other words, they decrease the knowledge-content of production work. As a result, wage inequality among workers decreases, but wage inequality among top managers, and between them and workers, increases. Organizations also change as spans of control increase, as does the number of layers of problem solvers. Intuitively, improvements in communication technology generate a type of ‘superstar’ (Rosen, 1981) effect, whereby each top level manager can better leverage their knowledge.

In contrast, reductions in the cost of accessing knowledge increase the number of prob-

³Empirically, such hierarchies are commonly observed within functional or product areas in organizations. In a calling center, a low skill agent takes a call, and if she cannot solve it she passes it to her more experienced manager; in a manufacturing firm, a production worker tries to solve a problem first and if she can not passes it to the skilled shop steward; in a consulting firm, associates and analysts try to solve the easier problems and partners deal with the harder ones.

⁴Although we choose not to settle this indeterminacy formally, arguments from the broader theory of the firm (e.g., Grossman and Hart, 1986) may apply to the question of when should knowledge transactions take place within firms. These arguments could potentially be incorporated in our framework. Empirically, as long as the determinants of the boundaries of the firm are not affected significantly by exogenous variables, changes in the size of hierarchies result in corresponding changes in the size of firms.

lems solved by agents at all organizational layers. This technological change results in an increase in the knowledge content of production work and thus to an increase in wage inequality within a given layer, as well as wage inequality in the economy as a whole. It also leads to an increase in spans of control and, if communication costs are high enough, to a decrease in the number of layers. We find the distinction between information and communication technology helpful in interpreting the available evidence on the impact of technological change in the past decades.

This paper contributes to three strands in the literature: the one that studies the impact of technology on the labor market; the one on hierarchies; and the one on continuous matching models. We describe briefly its relation to these literatures in turn. First, a recent theoretical literature has studied the impact of technology on the labor market, motivated by the recent changes in wage inequality. Within this literature, Saint Paul (2000) explores the impact of knowledge on the structure of earnings, studying team formation and selection when there are spillovers between worker's skills. Other theoretical contributions studying the link between technology and the structure of earnings have focused on issues such as the possibility of reverse causation from skills to technology (Acemoglu, 1998), the impact on within wage inequality of within group ability distinctions (Galor and Moav, 2000), the role of worker heterogeneity in the ability to implement new technologies (Galor and Tsiddon, 1997), capital-skill complementarity (Krusell, et. al. (2000)), or the impact of market size on specialization (Mobius, 2000). None of these papers embed the analysis of the impact of technology on the labor market in a model of hierarchical organization, and thus none of them delivers implications for the organizational structure of teams and firms, or for the occupational distribution of the labor force, which are a focus of this paper.

Second, a large recent literature has studied the internal organization of hierarchies, that is, how hierarchies process information, monitor performance, allocate resources, or facilitate knowledge acquisition.⁵ Papers in this literature solve the firm optimization problem without embedding it in an equilibrium framework. In particular, previous work has ignored the interactions between the hierarchy's design problem and the labor market that are the focus of this paper. Two exceptions are Lucas (1978) and Rosen (1982). Lucas (1978) presents the first equilibrium model of occupational choice: agents of different managerial skill control teams of homogeneous production workers – thus no within worker wage

⁵See for example Van Zandt (1998), Bolton and Dewatripont (1994), Geanakoplos and Milgrom (1991), Qian (1994), and Garicano (2000).

inequality or matching is possible. Rosen (1982) develops an equilibrium model of hierarchies in which workers are differentiated. However, his analysis has three limitations that our paper addresses. First, in his model there is perfect substitution between quantity and quality of workers – only efficiency units matter. As a result, matching plays no role. This has the implication that the assignment of agents to teams is undetermined. Second, Rosen does not characterize fully the equilibrium: issues such as whether perfect stratification takes place or not are left open. Third, the generic multiplicative technology he assumes is silent about the actual role of managers and workers, and as a result does not allow for the study of the impact of specific technological changes.

A third strand to which this paper contributes is the literature on frictionless matching models, started by Tinbergen (1956). A key result in this literature is that when production functions are symmetric and supermodular, the equilibrium assignment of individuals to teams (Kremer, 1993) or couples (Becker, 1981) involves self-matching or segregation. However, as Kremer and Maskin (1996) have remarked, such symmetric production functions face important empirical challenges: self-matching implies little or no within-firm heterogeneity in individuals’ ability or earnings, and cannot account for some key empirical facts about wage inequality. They argue that to generate these empirical patterns, as well as outcomes in which individuals’ positions correspond closely to their ability, production functions must involve complementarity between tasks, imperfect substitutability between skills (so that matching matters), and asymmetric sensitivity to ability.⁶ The production function we suggest has these three properties and thus advances this literature in three respects. First, we allow for hierarchical, that is many to one, matching. Second, occupational choice is endogenous. Third, agents may be in multiple layer hierarchies, and thus matching is multiple-sided.

The rest of the paper is structured as follows. Section 2 presents the model. Section 3 defines and characterizes a competitive equilibrium. Section 4 studies the equilibrium impact of IT. Section 5 contrasts our theory to US evidence and Section 6 concludes. All proofs are relegated to the Appendix.

⁶See Sattinger (1993) for a general literature review and Legros and Newman (2002) for a theoretical review that unifies existing results and delivers some new ones. An alternative approach within this literature is, rather than matching individuals to each other, like in the papers we discuss in the text, to start from an exogenous distribution of worker types and tasks and study how individuals are assigned to tasks in equilibrium (Teulings, 1995 and 2005). An advantage of our approach is that it allows us to make inferences on the organization of work (spans, layers, and knowledge content of the jobs) and the occupational distribution, as well as on wage inequality.

2 The Model

We model an economy in which agents of heterogeneous ability learn to solve tasks, choose an occupation, and a team to join. The available occupations are different layers of problem solving and production. Agents supply a unit of time, which may be used in production or in helping others solve problems. The equilibrium allocation in the economy determines wages for all agents, the tasks they perform, and the composition and structure of teams.

2.1 Production and Knowledge

Production requires labor and knowledge. Agents spend time in production and must solve the problems they confront in order to produce. In particular, agents draw one problem per unit of time spent in production. Output is 1 if the problem is solved, and 0 otherwise. A problem can be solved instantaneously by an agent who has enough knowledge. Some problems are more common than others. Problems are ranked by the likelihood that they will be confronted, so that problem Z is associated with a continuous density $f(Z)$ and c.d.f. $F(Z)$, where $f'(Z) < 0$.

Solving problems requires knowledge. All agents must learn the most common problems before learning the less common ones, so that more knowledgeable agents know everything that less knowledgeable ones do, and more. That is, knowledge is cumulative. The knowledge of an agent is then characterized by a number $\tilde{z} \in \mathbb{R}_+$, signifying that an agent can solve all problems $Z \in [0, \tilde{z}]$. To simplify the discussion and notation, we define the proportion of problems a worker can solve as $q = F(\tilde{z})$. Then $\tilde{z} = z(q)$, where $z(\cdot) = F^{-1}(\cdot)$, and so $z' > 0, z'' > 0$ (by the properties of $f(\cdot)$). Thus, $z(q)$ denotes the knowledge an agent needs to acquire in order to solve a proportion q of problems.

Agents differ in their cognitive ability so that higher ability agents incur lower learning costs. We assume the distribution of ability in the population can be described by a continuous density function, $\alpha \sim \phi(\alpha)$, with support in $[0, 1]$. In particular, we define ability so that the cost of learning to solve an interval of problems of length 1 is given by⁷

$$c(\alpha; t) = t - \alpha. \tag{1}$$

Note that (1) implies log supermodularity of the cost of knowledge acquisition $c(\alpha; t)z$ in ability α and knowledge z , as required for comparative advantage: high ability types have

⁷The linearity of the learning cost function $c(\cdot)$ in α and the limitation of the support to $[0, 1]$ is without loss of generality, since we can always scale α to fit these restrictions.

a comparative advantage in knowledge acquisition (Sattinger, 1975, and Teulings, 1995). A decrease in t represents an improvement in information technology that decreases the cost of learning (i.e., a technology that decreases the cost of accessing knowledge, such as cheaper data-base storage and search).

2.2 Communication and Organization

Agents can communicate their knowledge to others, and thus help them solve problems. The possibility of offering help to others allows agents to form organizations in which several individuals combine their time and knowledge to produce together. These organizations take the form of ‘knowledge-based hierarchies.’ On the lowest layer of these teams is a set of equally knowledgeable production workers, who learn the most routine problems and spend all of their time in production, and thus generate one problem per worker. Above them are one or multiple layers of managers, or specialized problem solvers, who in addition learn the exceptions (the less frequent problems). These managers do not engage in production, and thus do not draw problems. Production then proceeds as follows. Workers draw a problem per unit of time. If they can solve it, they produce; otherwise, they ask for help to the managers in the layer immediately above them, in which case these managers incur a communication cost of $h < 1$ units of time.⁸ If these managers know how to solve the problem they solve it; otherwise, they pass it on to the layer immediately above them, and so on, until the problem is solved or it reaches the highest layer in the organization, L . Teams have a pyramidal structure: each higher layer has a smaller number of agents than the previous one, since only a fraction of problems are passed on.⁹

Thus the organization of production is characterized by (i) agents specialized in production or in management; (ii) production workers learn the more common problems and problem solvers learn in addition exceptions, more exceptional the higher up the hierarchy; (iii) a sequential path of the problems up the hierarchy; and (iv) a pyramidal shape. All these characteristics are optimal under the assumption that agents do not know who may

⁸If $h \geq 1$ organization is never optimal and so all agents prefer to be self-employed. We assume (as in Garicano 2000) that this cost h results whether those asked know the answer to the problem themselves or not, since the communication and diagnosis must still take place.

⁹Note that we assume that no substantial heterogeneity exists in communication skills among workers. Although it could be that those with better cognitive ability are better able to communicate, it could also be that these two variables are related in other, more complicated, ways. If communication cost are a decreasing function of the ability of individuals, the comparative advantage of high skill agents in management will be further emphasized and we would expect the equilibrium allocation to exhibit similar qualitatively properties as in our framework. A more general study of the relation between learning costs and communication costs is worth pursuing, but is beyond the scope of this paper.

know the solution to problems they cannot solve, as Garicano (2000) shows in a model with homogeneous workers.¹⁰ The purpose of the hierarchy is to protect the knowledge of those who are more knowledgeable from easy questions that others can solve. If it were, instead, easy to diagnose and ‘label’ a problem that one does not know, then production workers would directly go with their unsolved problem to those with the exact specialized knowledge, and the organization would not be hierarchical, but there would be instead a one-to-one correspondence between problems solved and skill type.

Consider an organization with n_0 production workers with knowledge $q_0 = F(z_0)$; and n_l problem solving managers in layers $l = 1, \dots, L$, with knowledge q_l . Workers in production draw one problem each, and solve in expectation a fraction q_0 of them. Hence they pass on a fraction $(1 - q_0)$ of all problems. Managers in layer 1 are thus asked to solve $n_0(1 - q_0)$ problems, which they can address in $n_0(1 - q_0)h$ units of time. Optimally, managers join teams with precisely the right number of production workers so that they use all their time. Since all agents have one unit of time available, the number of managers in layer 1 is $n_0h(1 - q_0) = n_1$. The time constraint implies that the span of the manager is limited by the knowledge of their subordinates. If subordinates acquire more knowledge, they will require help less often, and managers will be able to supervise larger teams.¹¹

Managers in layer 1 can only solve a fraction q_1 of problems, and so pass up to the next layer $n_0(1 - q_1)$ problems. Thus the number of managers in layer 2 is given by $n_0h(1 - q_1) = n_2$. In general, managers of layer l are asked $n_0(1 - q_{l-1})$ times, since they are asked for help on all the problems that managers in layer $l - 1$ where not able to solve. Hence, the number of managers in layer l satisfies

$$n_0h(1 - q_{l-1}) = n_l.$$

Note that, as we pointed out before, the organization is pyramidal, $n_0 > n_1 > \dots > n_L$. Output is produced whenever any of the managers or workers can solve the problem, that is,

¹⁰Garicano (2000) studies the organization of a single team composed of homogeneous workers and studies under what conditions a ‘knowledge based hierarchy,’ is optimal. In contrast to this paper, that paper is neither concerned with the equilibrium matches among agents, nor is it concerned with wages.

¹¹Of course all of this holds in expectation. In principle, the interpretation of our technology given in the text requires us to address the stochastic element in the arrival of problems, which could result in congestion and queuing. An alternative interpretation, that circumvents the need to address these issues, is that each worker draws a continuum of problems of measure one with distribution $F(\cdot)$. Workers then solve the problems that they can, given their skill level, and ask managers for help on the rest. Then, h would be interpreted as the time cost for a manager of helping on a unit mass of problems. Output of a production worker, in turn, would be the mass of problems solved. We do not use this language in the text since it leads to a more convoluted exposition.

a problem is solved with probability q_L . Expected total output produced by the organization is then given by

$$y = q_L n_0. \quad (2)$$

Note the source of complementarity between skills in our model: An able top managers increases the productivity of all workers in the team. The more knowledgeable subordinates, the larger the team and the more can managers leverage their knowledge.¹²

2.3 Agents' Problem

Agents are income maximizers. Their problem is to choose their occupation to maximize income, given the available job opportunities. Available jobs are indexed by α' . A job α' pays a wage plus learning costs given by $w(\alpha') + c(\alpha'; t)z(q(\alpha'))$ and requires agents to know how to solve a proportion $q(\alpha')$ of problems. The problem of an agent with ability α is to choose a job α' that maximizes her income minus actual learning costs, $c(\alpha; t)z(q(\alpha'))$, so

$$U(\alpha) = \max_{\alpha'} [w(\alpha') + c(\alpha'; t)z(q(\alpha'))] - c(\alpha; t)z(q(\alpha')). \quad (3)$$

Therefore, agents can either work for jobs designed for their ability α or for jobs designed for different abilities α' . The first order condition yields

$$w'(\alpha^*) = -c'(\alpha^*; t)z(q(\alpha^*)) - z'(q(\alpha^*))q'(\alpha^*)(c(\alpha^*; t) - c(\alpha; t))$$

Notice that $w'(\alpha^*) = -c'(\alpha^*; t)z(q(\alpha^*))$ when $\alpha^* = \alpha$ which, as we show below, is the case in equilibrium. Hence, the slope of the wage function is equal to the decrease in learning costs as ability increases. We turn now to the problem of a team, which will allow us to establish that agents have incentives to choose the job designed for their own ability.

2.4 Firms' Problem

In this section we assume that a hierarchy is integrated in a firm, and study the problem of a firm with a given number of layers. In Section 3, we will determine the different number of layers of the universe of hierarchies that operate in equilibrium. Profits of a hierarchy are given by production minus labor costs, since we normalize the price of output to unity. Thus, the problem of a hierarchy of L layers that faces a wage schedule, $w(\alpha)$, is to choose

¹²Given this technology, note that in order for agents to organize in hierarchies it must be the case that $h < 1$. Production in a two layer hierarchy is given by $q_m/h(1 - q_p)$. However, if these agents work on their own they get $q_m + q_p/h(1 - q_p)$. The second term is larger than the first one if $h \geq 1$.

the ability, knowledge, and number of agents in each layer of the team. The expected profit of the hierarchy is then given by

$$\Pi(L) = \max_{\{q_l, n_l, \alpha_l\}_{l=0}^L} q_L n_0 - \sum_{l=0}^L n_l [c(\alpha_l; t) z(q_l) + w(\alpha_l)] \quad (4)$$

subject to time constraints of the different layers of managers,

$$\begin{aligned} h n_0 (1 - q_{L-1}) &= n_L \equiv 1, \\ h n_0 (1 - q_{L-2}) &= n_{L-1} \\ &\vdots \\ h n_0 (1 - q_0) &= n_1. \end{aligned} \quad (5)$$

That is, profits are given by output minus wages, $w(\alpha)$, and learning costs, $n_l c_l z(q_l)$. We call the manager in the highest layer an entrepreneur, and normalize their number, n_L , to 1.¹³ The choice of the ability of subordinates leads to one immediate result. The first order conditions with respect to α_l yield

$$w'(\alpha) = -c'(\alpha; t) z(q). \quad (6)$$

This implies that in equilibrium wages will be such that agents choose the job designed for their own ability and so in equilibrium the solution to (3) is $\alpha^* = \alpha$. Hence, in equilibrium the slope of the wage function is given by (6). Thus, we ignore this optimization for the rest of this section and eliminate the choice of α from the notation. We come back to it when we discuss the equilibrium allocation of this economy.

In the next section, and before moving to characterizing the equilibrium, we reformulate and study the maximization problem of the firm. This will allow us to substantially simplify the equilibrium characterization, and will also allow us to interpret the knowledge transactions in equilibrium, not only as taking place within firms, but also in referral or consulting markets.

2.5 Knowledge Transactions

In this section we formulate (4) in a convenient recursive form. The objective is to show that the decision making process in the hierarchy can be decentralized by allowing intermediate managers to make decisions about the knowledge of their subordinates. This will allow us

¹³Since there is no interaction between agents in the same layer of a given hierarchy, this normalization is without loss of generality.

to understand knowledge transactions and the role of wage differences within the hierarchy. Wages in this economy allocate agents to teams and encourage them to perform a suitable role in the team. In this sense we can think about wages as compensating members of the team not only for the problems that they actually solve, but also for passing problems to the upper layers.

First note that the total fraction of problems solved by the hierarchy is the sum of the fraction of problems solved at each layer, namely,

$$q_L = [q_L - q_{L-1}] + q_{L-1} = [q_L - q_{L-1}] + [q_{L-1} - q_{L-2}] + \dots + [q_1 - q_0] + q_0. \quad (7)$$

Hence, we can re-write (4) as

$$\Pi(L) = \max_{\{q_l, n_l\}_{l=0}^L} n_0 \left[\sum_{l=0}^{L-1} \left[(q_{l+1} - q_l) - \frac{n_{l+1}}{n_0} [c(\alpha_{l+1}; t)z(q_{l+1}) + w(\alpha_{l+1})] \right] + [q_0 - (c(\alpha_0; t)z(q_0) + w(\alpha_0))] \right]. \quad (8)$$

Second, note that by the constraints in (5), $n_l/n_0 = h(1 - q_{l-1})$, and recall that $n_0 = 1/h(1 - q_{L-1})$. Hence, we can write the maximization problem as

$$\begin{aligned} \Pi(L) = & \max_{q_{L-1}, q_L} \frac{1}{h(1 - q_{L-1})} \{ (q_L - q_{L-1}) - h(1 - q_{L-1})(c(\alpha_L; t)z(q_L) + w(\alpha_L)) + \\ & + \max_{q_{L-2}} \{ [(q_{L-1} - q_{L-2}) - h(1 - q_{L-2})(c(\alpha_{L-1}; t)z(q_{L-1}) + w(\alpha_{L-1}))] + \\ & + \max_{q_{L-3}} \{ [(q_{L-2} - q_{L-3}) - h(1 - q_{L-3})(c(\alpha_{L-2}; t)z(q_{L-2}) + w(\alpha_{L-2}))] + \\ & \vdots \\ & + \max_{q_0} \{ [(q_1 - q_0) - h(1 - q_0)(c(\alpha_1; t)z(q_1) + w(\alpha_1))] + \\ & + [q_0 - (c(\alpha_0; t)z(q_0) + w(\alpha_0))] \} \dots \} \}, \end{aligned}$$

which allows us to define, starting from the last term in the previous equation,

$$p(q_0, w) \equiv q_0 - (c(\alpha_0; t)z(q_0) + w(\alpha_0)), \quad (9)$$

and for the one before the last term,

$$p(q_1; w) \equiv \max_{q_0} [q_1 - q_0 - h(1 - q_0)(c(\alpha_1; t)z(q_1) + w(\alpha_1))] + p_0(q_0, w) \quad (10)$$

and, generically for all intermediate layers, $l = 1, \dots, L - 1$,

$$p(q_l; w) \equiv \max_{q_{l-1}} [q_l - q_{l-1} - h(1 - q_{l-1}) [c(\alpha_l; t)z(q_l) + w(\alpha_l)] + p(q_{l-1}; w)]. \quad (11)$$

The overall profit maximization of the firm is then given by

$$\Pi(L) \equiv \max_{q_L, q_{L-1}} \frac{q_L - q_{L-1} - h(1 - q_{L-1}) [c(\alpha_L; t)z(q_L) + w(\alpha_L)] + p(q_{L-1}; w)}{h(1 - q_{L-1})}$$

and so the first order conditions become

$$c(\alpha_L; t)z'(q_L^*) = \frac{1}{h(1 - q_{L-1}^*)}, \quad \text{and} \quad (12)$$

$$p'(q_{L-1}^*; \cdot) = \frac{1 - q_L^* - p(q_{L-1}^*; \cdot)}{1 - q_{L-1}^*}. \quad (13)$$

The first order conditions of the intermediate layer maximizations are given, from (11), by

$$p'(q_l^*; \cdot) = 1 - h [w(\alpha_{l+1}) + c(\alpha_{l+1}; t)z(q_{l+1}^*)] > 0, \quad \text{for } l = 0, \dots, L - 2, \quad (14)$$

where the inequality follows from the fact that wages and learning costs cannot exceed the maximum gain from solving a problem, which is equal to 1. Notice that using (11) these first order conditions imply that

$$p'(q_l^*; \cdot) = \frac{1 - q_{l+1}^* + p(q_{l+1}^*; \cdot) - p(q_l^*; \cdot)}{1 - q_l^*}. \quad (15)$$

Finally, equation (11) leads to the following envelope conditions

$$p'(q_l^*; \cdot) = 1 - h(1 - q_{l-1}^*)c(\alpha_l; t)z'(q_l^*), \quad \text{for } l = 1, \dots, L - 1. \quad (16)$$

and for $l = 0$, (9) implies that

$$p'(q_0^*; \cdot) = 1 - c(\alpha_0; t)z'(q_0^*). \quad (17)$$

Eliminating $p'(q_l^*; \cdot)$ by combining conditions (16) and (17) with (14) generates an Euler equation that can be readily interpreted: the marginal cost of an increase in the knowledge of layer l workers must equal the gain (as given by the saving in wages and training) associated with reducing the number of managers in the layer $l + 1$ when less questions are asked.

These conditions fully characterize the solution to this problem. To interpret them note that the wages of a manager of layer $0 < l < L$ can be written, from (11), as

$$w(\alpha_l) = \frac{q_l^* - q_{l-1}^* + p(q_{l-1}^*; \cdot) - p(q_l^*; \cdot)}{h(1 - q_{l-1}^*)} - c(\alpha_l; t)z(q_l^*). \quad (18)$$

This suggests a ready interpretation for the function p . It is the fee or transfer that managers receive from lower level managers or workers to deal with the problems that they cannot solve. We will show that in equilibrium this fee is negative. The wage structure of a firm can thus be interpreted as a transfer system in which managers pay workers a fee to pass problem to them and managers keep the output associated with the problems that

they solve. In turn these managers receive fees from the managers above them for passing the problems that they cannot solve.

The earnings of a particular manager, $w(\alpha_l)$, for each of the $1/h$ problems she can deal with, are given by the conditional probability that she can solve the problem given that those below her could not solve it, $(q_l^* - q_{l-1}^*) / (1 - q_{l-1}^*)$, plus the (negative) conditional fee that she receives from those who pass her the problem, $p(q_{l-1}^*; \cdot)$, minus the (negative) conditional fee that she pays to pass the problem to a higher layer, $p(q_l^*; \cdot)$, minus her training costs.

The above conditions suggest the following way to write the problem, which in fact can be shown to be equivalent. Suppose each manager chooses her knowledge and the knowledge of those below her sequentially, so

$$w(\alpha_l) = \max_{q_l, q_{l-1}} \left[\frac{q_l - q_{l-1} + p(q_{l-1}; \cdot) - p(q_l; \cdot)}{h(1 - q_{l-1})} - c(\alpha_l; t)z(q_l) \right]. \quad (19)$$

Since workers do not have subordinates, let them choose only their own knowledge to maximize their income, namely,

$$w(\alpha_0) = \max_{q_0} [q_0 - p(q_0; \cdot) - c(\alpha_0; t)z(q_0)]. \quad (20)$$

Finally, top managers do not pass problems to higher layers. If we define $p(q_L^*; \cdot)$ as the transfer that entrepreneurs would have to pay to pass a problem to a higher layer, it must be the case that

$$p(q_L^*; \cdot) \geq 0. \quad (21)$$

Otherwise they would prefer to pass the problem further. Hence, their earning are given by

$$w(\alpha_L) = \max_{q_L, q_{L-1}} \left[\frac{q_L - q_{L-1} + p(q_{L-1}; \cdot)}{h(1 - q_{L-1})} - c(\alpha_L; t)z(q_L) \right]. \quad (22)$$

It is easy to check that the problem in (19 – 22) is equivalent to the problem in (4) if we impose zero profits ($\Pi(L) = 0$). It is also easy to check that the first order conditions of problem (19) are the same as the first order conditions (12 – 14) plus the envelope conditions (16). Hence using this transfer function we can interpret the problem as one in which agents choose sequentially their knowledge and the knowledge of those below them so as to maximize their own earnings.

In fact, these transfers can be interpreted as prices and so the problem above can be understood as one in which knowledge transactions take place in the market instead of within firms. In particular, as the next section points out, we can think either about

agents selling the problems that they cannot solve, as in the case of referrals, or agents hiring consultants to ask them the solution to problems for which they do not have enough knowledge. We now turn to a description of these markets and show that they are simply reinterpretations of the knowledge transactions within firms described above.

2.6 Alternative Formulations: Knowledge Transactions in the Market

Referrals and the market for problems: Consider now a market in which there are two types of occupations: production workers and problem solvers. Production workers with skill α_0 draw a problem per unit of time and use their knowledge q_0 to try to solve it. If they can solve the problem, they do so and earn 1; if they cannot solve it, they sell it in the market at a price $\tilde{p}(q_0)$. They incur a training cost $c(\alpha_0; t)z(q_0)$. Their choice of knowledge is the solution to their earning maximization problem, as given by

$$w(\alpha_0) = \max_{q_0} q_0 + \tilde{p}(q_0)(1 - q_0) - c(\alpha_0; t)z(q_0).$$

There are (possibly) multiple layers of problem solvers, $l = 1, \dots, L$. Problem solvers in layer 1 buy problems from the pool of problems left unsolved by production workers at price $(\tilde{p}(q_0))$. These problems are communicated to them at a cost of h units of time per problem. Problem solvers of layer 2 buy problems from the pool of problems left unsolved by those in layer 1 at a price $p(q_1)$, and so on. In general, the earnings of these problem solvers are given by

$$w(\alpha_l) = \max_{q_l, q_{l-1}} \frac{1}{h} \left(\frac{q_l - q_{l-1} + \tilde{p}(q_l)(1 - q_l)}{1 - q_{l-1}} - \tilde{p}(q_{l-1}) \right) - c(\alpha_l; t)z(q_l),$$

for $l = 1, \dots, L$, where $\tilde{p}(q_l) \leq 0$ for those problem solvers who do not sell the problems further in equilibrium. Problem solvers at layer l can deal with $1/h$ problems per unit of time; they buy problems from agents with knowledge q_{l-1} and solve it with probability $(q_l - q_{l-1})/(1 - q_{l-1})$. If they cannot solve them (which happens with probability $(1 - q_l)/(1 - q_{l-1})$), they sell them at a price $\tilde{p}(q_l)$. They unconditionally pay $\tilde{p}(q_{l-1})$ for each problem they get from the lower layer problem solvers.

This formulation is just a reinterpretation of the problem above if we let

$$\tilde{p}(q_l) = -\frac{p(q_l; w)}{1 - q_{l-1}}, \tag{23}$$

where w denotes the equilibrium wage function. To see that the problem is the same, substitute (23) in equation (20) and in equation (19).

This type of institution, where agents sell problems to other agents in exchange for a ‘referral’ fee, is often observed empirically. One example is the market for legal claims. In

this market, when a lawyer faces a client whose case she cannot solve, she refers that client to another lawyer in exchange for a fee.¹⁴

The market of consultant services: Consider a second alternative market institution. Production workers draw a problem per unit of time, and keep ownership of the production associated with solving the problem (rather than selling it when they do not know how to solve it); they pay a fee per problem to other agents for their advice. If production workers know the solution to the problem, they solve it; if not, they pay a fee $\hat{p}(q_1)$ to the problem solvers in layer 1. If these cannot solve them, then they pay a fee to problem solvers in layer 2, and so on. Workers earnings are then given by

$$w(\alpha_0) = \max_{\{q_l\}_{l=0}^L} q_L - c(\alpha_0; t)z(q_0) - \sum_{l=1}^L (1 - q_{l-1}) \hat{p}(q_l). \quad (24)$$

So workers' earnings are given by the probability that the problem is ultimately solved, minus the training costs, and minus the expected consulting fees paid (given by the probability that a given consultant l is hired times the consulting fee of that consultant, $\hat{p}(q_l)$). Consultant earnings are the expected consulting fee earned, as given by the fee per service times the expected number of services they can provide (given that each one requires h units of time), minus their training costs, namely,

$$w(\alpha_l) = \max_{q_l} \frac{\hat{p}(q_l)}{h} - c(\alpha_l; t)z(q_l). \quad (25)$$

Again, this is just a reinterpretation of the setup in the previous section if we let

$$\hat{p}(q_l) = \frac{q_l - p(q_l; w) - [q_{l-1}^* - p(q_{l-1}^*; w)]}{1 - q_{l-1}^*}, \quad (26)$$

as well as

$$\hat{p}(q_L) = \frac{q_L - [q_{L-1}^* - p(q_{L-1}^*; w)]}{1 - q_{L-1}^*}, \quad (27)$$

where as before w represent the equilibrium wage function. It is easy to check that substituting (26) and (27) in (25) and (24), we again obtain (19).

This type of institution, where agents pay others a consulting fee, are prevalent in medicine and consulting services for firms. An agent goes to the doctor, pays for advice, and receives a treatment. If the treatment is not the right solution the patient goes to another, more expensive and knowledgeable, doctor and so on.

¹⁴See Spurr (1988). Garicano and Santos (2004) analyze a referral market under asymmetric information.

3 Equilibrium

The previous analysis has allowed us to obtain, for a given hierarchy, the proportion of tasks each agent should learn to perform, as well as team sizes, given wages. We now turn to the analysis of an equilibrium in this economy. An equilibrium allocation specifies the sets of agents in different occupations, an assignment of agents to their supervisors, and the wage schedule (or other prices in the different decentralizations) that supports this assignment.¹⁵

Before we define a competitive equilibrium for this economy we need to discuss the labor market equilibrium condition. In order for labor markets to clear we need to guarantee that the supply of workers or managers for any measurable set of abilities at a given layer is equal to the demand for these workers or managers by managers or entrepreneurs at any layer. Let $n(\alpha)$ denote the total number of workers or managers hired as direct subordinates of managers or entrepreneurs with ability α in equilibrium. Let $a(\alpha)$ denote the ability of the manager assigned to an employee of ability α in equilibrium. In order for $a(\alpha)$ to be defined over the whole set of abilities, $[0, 1]$, we set $a(\alpha) = 1$ for all entrepreneurs. Since hierarchies have only one entrepreneur at the top $n(a(\alpha)) = 1$ when agents with ability α are entrepreneurs.¹⁶ Let A_S be the set of agents with subordinates and let A_M denote the set of agents that are not at the top of the hierarchy (all agents that have a manager or entrepreneur above them). Then, labor markets clear if for every $\alpha \in A_M$,¹⁷

$$\int_{[0, \alpha] \cap A_M} \phi(\alpha') d\alpha' = \int_{[a(0), a(\alpha)] \cap A_S} \frac{n(\alpha')}{n(a(\alpha'))} \phi(\alpha') d\alpha'. \quad (28)$$

The left hand side is the supply of employees in the interval $[0, \alpha]$ (the integral of population density over the set of workers and managers, $[0, \alpha] \cap A_M$). The right hand side is the demand for employees by managers and entrepreneurs in the interval $[a(0), a(\alpha)]$: Managers and entrepreneurs of ability α hire $n(\alpha)$ employees and there are $n(a(\alpha))$ of them. The definition of equilibrium in this setup is then given by:

¹⁵Note that the problem we confront is different from Sattinger (1993) and Teulings (1995) (and all other) canonical ('Ricardian') assignment problems. First, rather than matching one worker and one machine, we match here one manager and any number of subordinates (see Fernández and Galí (1999) for an example with borrowing constraints). Second, our economy must not assign given machines to given workers; we have instead to determine which agents are going to be managers, at what layer, and which ones are going to be production workers. Third, the interaction between manager skill and subordinate skill is not direct, but takes place through team size and the knowledge acquired.

¹⁶Given that the best agent in the economy, $\alpha = 1$, may hire more than 1 worker, this implies that $n(1)$ is in general not single-valued.

¹⁷The integrals in this equation are not simply over the sets $[0, \alpha]$ and $[a(0), a(\alpha)]$ since we have not proven yet that the set of workers, managers, and entrepreneurs are connected. In Proposition 4 we prove that this has to be the case in equilibrium.

Definition 1 *A competitive equilibrium is*

- *the set of numbers of layers of hierarchies operating, \bar{L} , where $L \in \bar{L}$ is the number of layers of the highest hierarchy,*
- *a collection of sets $\{A_l = A_{lM} \cup A_{lE}\}_{l=0}^L$ such that for $\alpha \in A_{lM}$ agents become managers of layer l , $l = 1, \dots, L$, or workers for $\alpha \in A_{0M}$, and entrepreneurs of layer l for $\alpha \in A_{lE}$, $l = 0, \dots, L$,*
- *a wage function, $w(\alpha) : [0, 1] \rightarrow \mathbb{R}_+$,*
- *an assignment function, $a(\alpha) : [0, 1] \rightarrow A_S$, where $A_S \equiv [0, 1] \setminus A_{0M}$ and $a(\alpha) = 1$ for $\alpha \in \bigcup_{l=0}^L A_{lE}$,*
- *a knowledge function $q(\alpha) : [0, 1] \rightarrow [0, 1]$ and*
- *a total number of direct subordinates of agents with ability α , $n(\alpha) : A_S \rightarrow \mathbb{R}_+$, such that:*
 - (i) *Agents choose occupations to maximize utility, (3).*
 - (ii) *Firms choose the skill of their employees, their knowledge, and spans of control so as to maximize profits, (4).*
 - (iii) *Firms make zero profits and so $p(q(\alpha); \cdot) \geq 0$, for $\alpha \in A_{lE}$, $l = 0, \dots, L$, where $p(\cdot)$ is given by (11) and (21).*
 - (iv) *Labor markets clear, that is, (28) is satisfied for every $\alpha \in A_M \equiv \bigcup_{l=0}^L A_{lM}$.*

3.1 Assignment

We now turn to the characterization of the assignment function. An important characteristic of the equilibrium assignment is that it exhibits positive sorting, in the sense that better managers are matched with better employees. That is, the assignment function is strictly increasing for $\alpha \in A_M$.¹⁸ This is the result of the complementarity in production between the knowledge of team members. To see this, recall that the output of a given manager, as given by the production component in (18), is $q_l - q_{l-1}/(h(1 - q_{l-1}))$. Better subordinates

¹⁸We will talk about workers, managers and entrepreneurs for simplicity; clearly, these are measure 0 atoms in a continuous distribution. The only appropriate way to think about them are masses of workers and managers at given intervals. The reader should thus not read from the existence of the assignment function that agents are matched one to one.

allow for an increase in the size of the team, while a larger team size increases the marginal value of managerial knowledge.¹⁹

Proposition 1 *Any equilibrium of this economy involves positive sorting.*

High ability managers hire high ability agents so as to be shielded from solving easy and common problems. Hiring better workers allows managers to specialize in solving only the harder problems that lower layer agents cannot solve. With this result in hand, deriving the labor market equilibrium condition (28) implies that

$$\frac{\partial a(\alpha)}{\partial \alpha} = \frac{n(a(a(\alpha)))}{n(a(\alpha))} \frac{\phi(\alpha)}{\phi(a(\alpha))} \text{ for } \alpha \in A_M. \quad (29)$$

From (5), and since $n(\alpha)$ is the total number of direct subordinates of agents with ability α ,

$$\frac{n(a(a(\alpha)))}{n(a(\alpha))} = \frac{1 - q(\alpha)}{1 - q(a^{-1}(\alpha))} \text{ for } \alpha \in A_M \setminus A_{0M},$$

where $a^{-1}(\cdot)$ denotes the inverse of the assignment function $a(\cdot)$, and

$$\frac{n(a(a(\alpha)))}{n(a(\alpha))} = h(1 - q(\alpha)) \text{ for } \alpha \in A_{0M}.$$

Hence, the assignment function is given by

$$\frac{\partial a(\alpha)}{\partial \alpha} = \begin{cases} \frac{1 - q(\alpha)}{1 - q(a^{-1}(\alpha))} \frac{\phi(\alpha)}{\phi(a(\alpha))} & \text{for } \alpha \in A_M \setminus A_{0M} \\ h(1 - q(\alpha)) \frac{\phi(\alpha)}{\phi(a(\alpha))} & \text{for } \alpha \in A_{0M} \end{cases}. \quad (30)$$

Intuitively, suppose that we are assigning managers to workers. Then, the number of managers per subordinate that results from the firm optimization, over the ratio of available agents with the corresponding ability (given by the skill distribution), determines the slope of the assignment function. Equation (30) is a collection of ordinary differential equations that determine the functions $a(\cdot)$ given some initial values. Assume for the moment that the equilibrium is formed by a collection of connected sets, $\{A_l = A_{lM} \cup A_{lE}\}_{l=0}^L$, characterized by a set of real numbers, $\{\alpha_{ll}^*, \alpha_{ll+1}^*\}_{l=0}^L$, with $\alpha_{LL+1}^* = 1$, such that $A_{0M} = [0, \alpha_{00}^*]$, $A_{lM} = [\alpha_{l-1l}^*, \alpha_{ll}^*]$, $A_{lE} = [\alpha_{ll}^*, \alpha_{ll+1}^*]$ and $A_{LE} = [\alpha_{LL}^*, 1]$. That is, a threshold α_{ij}^* determines the boundary in the ability set between layer i and j when $i \neq j$, and between the set of managers and entrepreneurs when $i = j$ (see Figure 1). Different collections of sets may, in

¹⁹Note that it may be the case that multiple workers choose to acquire zero knowledge. In that case the assignment is indeterminate since managers only care about the amount and cost of knowledge of their workers and not their ability. In this case, given that the assignment of these workers is irrelevant for any other feature of an equilibrium, we always choose the assignment that exhibits positive sorting.

principle, be an equilibrium too, however, we prove below that the unique equilibrium of this economy has this form. Given this segmentation of skills into occupations, the boundary conditions that determine the assignment function are given, for $l = 0, \dots, L - 1$, by

$$a(0) = \alpha_{01}^*, \quad a(\alpha_{ll}^*) = \alpha_{l+1l+2}^*, \quad a(\alpha_{ll+1}^*) = \alpha_{l+1l+2}^*, \quad a(\alpha_{L-1L-1}^*) = 1.$$

Therefore, we have written the problem in a way that allows us to determine the equilibrium assignment function if we can determine the knowledge function $q(\cdot)$ and the boundaries between the abilities of the different occupations of agents, $\{\alpha_{ll}^*, \alpha_{ll+1}^*\}_{l=0}^L$.

Figure 1 illustrates the way the assignment function matches individuals in a hierarchy. It present an example where the economy has a maximum of three layers ($L = 2$) with entrepreneurs in layer 1 and 2 (that is, the set of entrepreneurs is connected, as we show below). The arrows at the top of the line represent the assignment for the hierarchy that hires the worst workers in the economy, which are matched with the worst managers of layer 1, which in turn are matched with the worst entrepreneurs of layer 2 (a hierarchy with 3 layers). The arrows below the line illustrate the assignment in the hierarchy that hires the best workers, which are matched with the best entrepreneurs of layer 1 (a hierarchy with 2 layers).

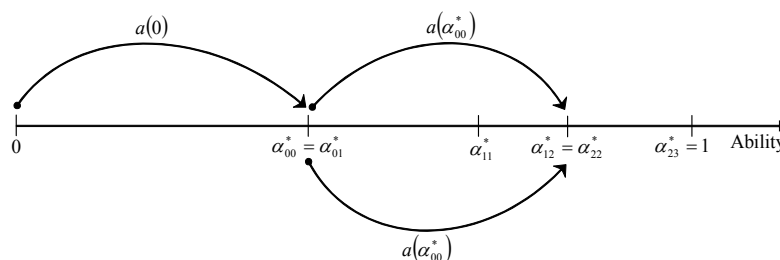


Figure 1: Assignments in equilibrium

3.2 Wages

We now turn to analyze the equilibrium wage schedule. The first characteristic of the wage function is that it is continuous with respect to ability, although in general not differentiable at every skill level. Continuity is the result of the agent's utility maximization problem and the labor market equilibrium condition. To see this note that if the wage function is not continuous, the problem in (3) implies that a strictly positive mass of agents with heterogenous skill will choose to work in a job that requires a particular skill level (a corner solution). This is, however, not consistent with the labor market equilibrium condition that requires these agents to work for managers with different skills (the assignment function

is strictly monotone in A_M). Continuity of $w(\alpha)$ therefore follows for all α , including the boundaries, $\{\alpha_{ll}^*, \alpha_{ll+1}^*\}_{l=0}^L$, between occupations.

The slope of the wage function is given, as we argued above, by (6), namely,

$$w'(\alpha) = -c'(\alpha; t)z(q(\alpha)) = z(q(\alpha)) > 0,$$

where the second equality follows from $c'(\alpha; t) = -1 < 0$ and the inequality from $z(\cdot) > 0$. Thus, since the wage function is continuous, it is increasing. In words, the marginal return to skill is the marginal value of agents skill, which is given by the knowledge the agent acquires. This equation contains a lot of the intuition for what follows. More inequality will result in equilibrium whenever a technology change leads agents to learn more tasks, as then the difference between more and less skilled agents becomes more pronounced.

The wage function is also convex. To see this note that for all α where $w(\cdot)$ is differentiable (all $\alpha \in [0, 1]$ except for the thresholds $\{\alpha_{ll}^*, \alpha_{ll+1}^*\}_{l=0}^L$),

$$w''(\alpha) = z'(q(\alpha))q'(\alpha).$$

The first term is positive given that $z'(\cdot) > 0$. We still need to show that more skilled agents learn more, so $q'(\alpha) > 0$. In Proposition 1 we showed that this is in fact the case under our assumption that $z''(\cdot) > 0$. Of course, the above argument only shows that the wage function is convex within occupations, given that the wage function is not differentiable at the thresholds (since q is not continuous at the thresholds). To show that it is globally convex we need to show that at the thresholds between occupations the slope of the wage function increases, or the knowledge function increases (in general discretely). In Proposition 1 we show that this must be the case in order for the labor market clearing condition to be satisfied. Thus, in our setup wages compensate ability more than proportionally. We formalize this reasoning in the next proposition.

Proposition 2 *Any equilibrium wage function, $w(\alpha) : [0, 1] \rightarrow \mathbb{R}_+$, is increasing and convex. Furthermore, the knowledge function $q(\alpha) : [0, 1] \rightarrow [0, 1]$ is increasing.*

It is illustrative to compare the equilibrium wage schedule, and the underlying knowledge acquisition, to the one that would result in equilibrium absent organization – that is, when all agents are self-employed (e.g., $h \geq 1$). The following proposition shows that organization reduces the knowledge of production workers and increases the knowledge of high level problem solvers. Production workers acquire less knowledge than they would absent organization, since they substitute learning for asking. As a consequence, the marginal

return to worker skill (6) is lower than with self-employment. For the highest level problem solvers, the marginal value of their knowledge is higher than in self-employment, since they can spread such knowledge over a larger team.

Proposition 3 *Relative to an economy without organization, organization increases the knowledge of entrepreneurs and decreases the knowledge of production workers. As a result organization decreases the marginal value of worker skill and increases the marginal value of the entrepreneur's skill.*

For those agents in between the top and the bottom, there are two effects – they can ask for help, which reduces the value of knowledge, and they can answer questions, which increases it. The proposition above thus shows that organization accentuates ability differentials among the best and worse individuals.

3.3 Occupational Stratification

In equilibrium occupations are stratified by ability. The lowest skilled agents are workers, the next set of agents are managers of layer one, then entrepreneurs of layer one, and so on for other layers. That is, agents in the economy are segmented by cognitive skill.

First notice that by equation (14), $p'(q(\alpha); \cdot) > 0$ for all $\alpha < a^{-1}(1)$ and by condition (21) and Proposition 1, $p(q(\alpha); \cdot) \geq 0$ for all $\alpha \geq a^{-1}(1)$. That is, the fee paid for transferring problems to the managers above should be non-negative for all agents that are at the top of hierarchies (do not ask questions) in equilibrium. This implies, since q is an increasing function of α , as shown in Proposition 2, that the fee for passing problems is negative for all $\alpha < a^{-1}(1)$. Managers, in effect, pay workers to pass problems to them. Since p is an increasing function of q , which is an increasing function of α , this implies that the set of agents without supervisors (entrepreneurs) has to be a connected set. Hence the economy can only sustain firms with two different but adjacent number of layers, $\bar{L} = \{L - 1, L\}$.

The occupational stratification by ability is also present among different layers of managers and managers and workers. In particular, the worse agents will be workers, and better agents will be managers of higher layers, and the higher the ability the weakly higher the layer. The result is again the outcome of $q(\alpha)$ being increasing in α and the fact that it is never optimal for a firm to hire subordinates of a manager with more or the same knowledge than the manager. To understand this last statement note that forming a team with subordinates with the same ability is always suboptimal and so agents prefer to be on their

own. The wage of an entrepreneur of a firm with L layers is given by (22), which is positive if there are firms in operation that have L layers. If this entrepreneur instead of choosing optimally the knowledge of her subordinates were to choose a subordinate in layer $L - 1$ with her same knowledge, $q_{L-1} = q_L$, her wage would be smaller than $-c(\alpha_L; t)z(q_L)$ since $p(q_L; \cdot) \geq 0$. Earnings become even more negative if $q_{L-1} > q_L$. A parallel argument goes through for intermediate managers.

Proposition 4 *Any equilibrium allocation with a maximum number of $L \geq 0$ layers is characterized by a set of thresholds, $\{\alpha_{ll}^*, \alpha_{ll+1}^*\}_{l=0}^L$, such that $\alpha_{ii} \leq \alpha_{ij} \leq \alpha_{jj}$, for $i < j$, $\alpha_{LL+1}^* = 1$, $[\alpha_{L-1L-1}^*, 1]$ are entrepreneurs of layers $L - 1$ and L , $[\alpha_{00}^*, \alpha_{L-1L-1}^*]$ are managers of layers 1 to $L - 1$, and $[0, \alpha_{00}^*]$ are workers. Hence, $\alpha_{ll}^* = \alpha_{ll+1}^*$ for all $l = 0, \dots, L - 2$, and $\alpha_{L-1L}^* = \alpha_{LL}^*$.*

Note that this ‘stratification’ result holds independently of the skill distribution. This is in contrast with the general class of production functions with complementarities and asymmetric skill sensitivity, where whether such strong stratification of occupations by skill takes place depends on the distribution of skill in the population, as Kremer and Maskin (1997) and Legros and Newman (2002) show. Intuitively, for the production technology we study, if it does not pay for an agent to pass problems, it does not pay for more able agents to pass even harder problems. Conversely, if an agent is worth helping, because the problems she cannot solve are sufficiently valuable, then all agents who are less knowledgeable than her are also worth helping.

3.4 Existence, Uniqueness, and Optimality

An equilibrium of this economy exists and is unique. Before we present this result, it is useful to outline the algorithm to find the equilibrium in this economy. An equilibrium can be constructed as follows:

1. Set $L = 1$ and fix $0 < \alpha_{00}^1 < \alpha_{01}^1 < 1$ and $w(0)$.
2. We can calculate $w(\alpha)$ for all $\alpha \in [0, 1]$ using equation (6) and $w(0)$.
3. Find α_{00}^* such that

$$\alpha_{00}^* = \min \left[\left\{ \alpha : \lim_{\alpha \uparrow \alpha_{00}^1} w(\alpha) = \lim_{\alpha \downarrow \alpha_{00}^1} w(\alpha) \right\}, \alpha_{01}^1 \right],$$

4. Let the supply of workers accumulated in the interval $[0, \alpha_{00}^*]$ be denoted by S_0 , so

$$S_0(w(0)) = \int_0^{\alpha_{00}^*} \phi(\alpha') d\alpha',$$

and the demand for workers accumulated in $[\alpha_{01}^1, 1]$ be denoted by D_0 , so

$$D_0(w(0)) = \int_{\alpha_{01}^*}^{\alpha_{12}^1} \frac{n(\alpha')}{n(a(\alpha'))} \phi(\alpha') d\alpha'.$$

Note that, since we are using (29) to calculate assignments, condition (28) is satisfied in the interior of these intervals. Find $w(0)$ such that

$$S_0(w(0)) = D_0(w(0)).$$

5. Find α_{01}^* such that

$$\alpha_{01}^* = \min \left[\left\{ \alpha : \lim_{\alpha \uparrow \alpha_{01}^*} w(\alpha) = \lim_{\alpha \downarrow \alpha_{01}^*} w(\alpha) \right\}, 1 \right].$$

6. If $0 < \alpha_{00}^* < \alpha_{01}^* < 1$, then $p(\alpha_{00}^*; w) = 0$ and we have found an equilibrium with $L = 1$ layers. If $\alpha_{01}^* = 1$ and so $\alpha_{00}^* = 0$, we found an equilibrium and the equilibrium number of layers is $L = 0$. If $\alpha_{00}^* = \alpha_{01}^*$ then the equilibrium number of layers L is larger than 1. Go to step 1. and repeat the algorithm with $L = 2$.

Figure 2 presents a diagram of an equilibrium wage function with the different thresholds and the conditions that have to be met in equilibrium. At the boundaries $(0, \alpha_{00}^*, \alpha_{01}^*, \alpha_{11}^*, \dots, \alpha_{LL}^*, 1)$ the wage function is not differentiable since the amount of knowledge acquired changes discontinuously. In the example illustrated in the figure, the economy has a maximum of three layers $\{0, 1, L = 2\}$, and so there are no entrepreneurs of layer 0 and no managers of layer 2 ($\alpha_{00}^* = \alpha_{01}^*$, $\alpha_{L-1L}^* = \alpha_{LL}^*$, and so $\bar{L} = \{1, 2\}$).

The next proposition shows that there exists an equilibrium with the characteristics described above. The proof of existence is constructive and so it develops in detailed the algorithm described above.²⁰

Proposition 5 *There exists a unique equilibrium allocation.*

²⁰ As noted in the proof of Proposition 1 it may be the case that multiple workers decide to acquire zero knowledge. In that case the equilibrium allocation is unique up to the choice of the assignment of these workers to managers.

So far we have studied the model without reference to the efficiency properties of the equilibrium allocation. The equilibrium allocation is efficient. The logic behind the result is that the total time endowment of agents is used in production, and the problems that are not eventually solved would be more costly to solve than the benefits that agents may derive from solving them (and therefore $p(\cdot) \geq 0$). Moreover, even though the technology exhibits complementarity between worker skills, all the interactions between agents are priced either through wages within hierarchies or in the market with problem prices. Hence, the First Welfare Theorem applies and the equilibrium allocation is Pareto Optimal. In the proof of the next proposition we set up the social planner problem and show that it is solved by the equilibrium allocation.

Proposition 6 *The equilibrium allocation is Pareto optimal.*

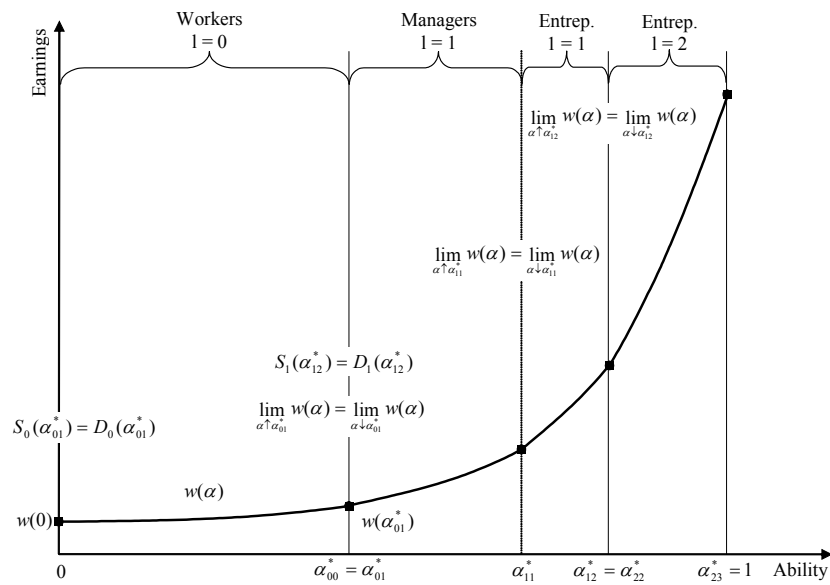


Figure 2: Construction of an equilibrium allocation

4 Effect of IT on wages and organization

To illustrate the qualitative predictions of the theory and advance the characterization of an equilibrium allocation, we compute several numerical examples. We will use these examples to explain the general equilibrium effects of both types of technological changes in our theory.

We study an example with an exponential density of problems, $f(z) = e^{-\lambda z}$ and a uniform distribution of worker talent, $\alpha \sim U[0, 1]$. Moreover, we let $\lambda = 2$ in all exercises.

Figure 3 presents the results of this numerical simulation.

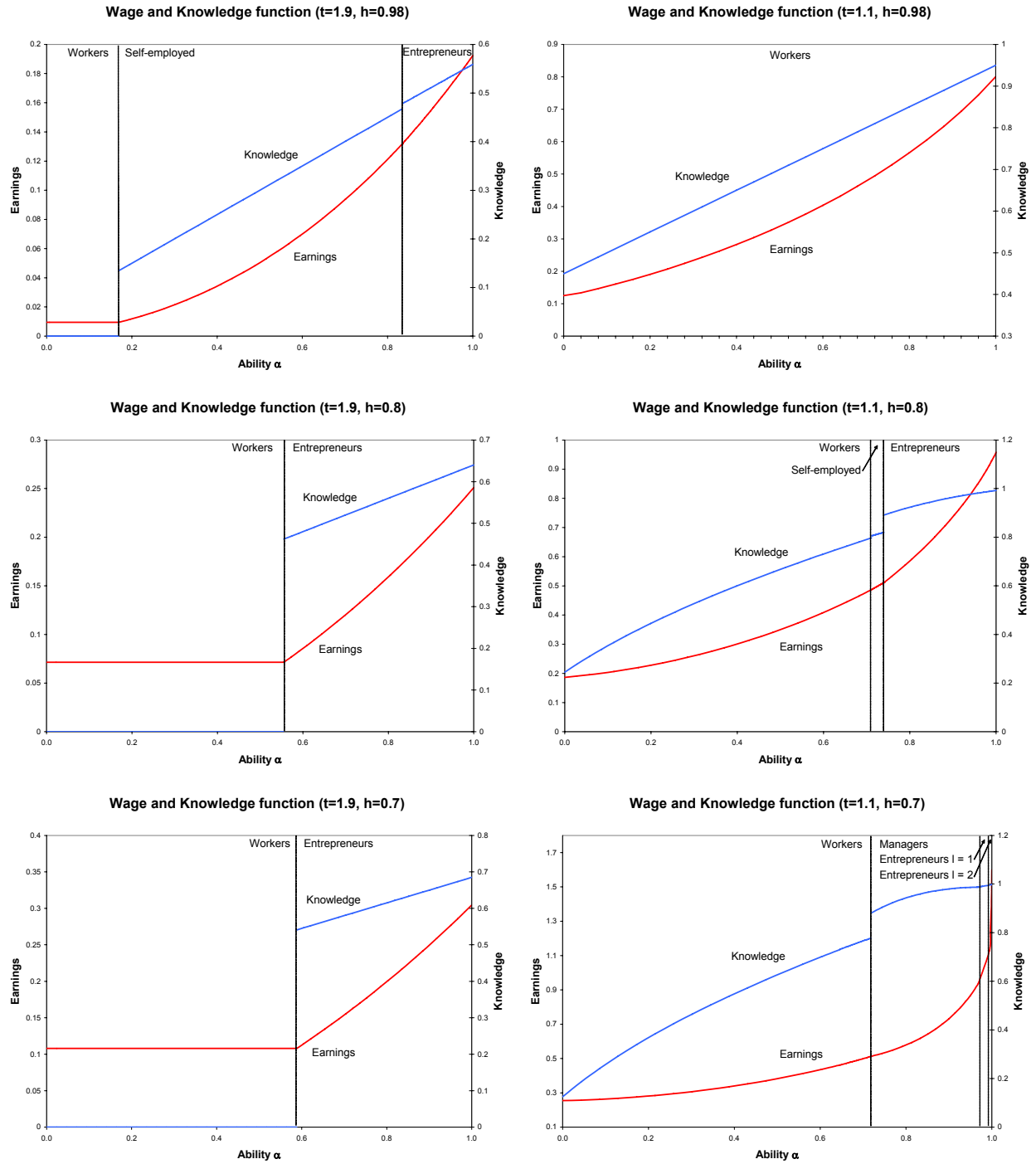


Figure 3: Numerical examples

The six graphs presented in Figure 3 show the equilibrium wage and knowledge of all

agents for different parameter values. The figures in the first row present the earnings and knowledge function for $h = 0.98$, and $t = 1.9$ and 1.1 respectively. The second row presents the same graphs when we lower the value of h to 0.8 for the same values of t , and the third row when we lower h further to $h = 0.7$. The parameter values have been chosen to maximize visibility. Table 1 may also be useful to understand the numerical results presented in Figure 3 and described in the next two subsections. Note that when $t = 1.9$ workers do not acquire knowledge since it is cheaper to ask about all problems than to learn. Therefore, in this case all workers earn the same in equilibrium.

There are several features of the equilibrium we have described above that are apparent in all the simulations presented. Wages are increasing and convex, higher ability agents learn more, there is positive sorting, and the set of entrepreneurs is connected. All of these are general results in our model and so should be present in all simulations. We now turn to the description of the effect of changes in communication and information technology.

Table 1: Numerical examples

Total wage inequality, SD			Worker/Self-employed wage inequality, SD			Entrepreneur/Manager wage inequality, SD		
$h \backslash t$	1.9	1.1	$h \backslash t$	1.9	1.1	$h \backslash t$	1.9	1.1
0.98	0.068	0.202	0.98	0.038	0.202	0.98	0.018	N/A
0.8	0.057	0.206	0.8	0	0.119	0.8	0.055	0.132
0.7	0.063	0.300	0.7	0	0.091	0.7	0.060	0.243
Span of Control ($\alpha = 1$)			Wage for $\alpha = 0$			Q of highest ability worker		
$h \backslash t$	1.9	1.1	$h \backslash t$	1.9	1.1	$h \backslash t$	1.9	1.1
0.98	1.02	0	0.98	0.01	0.12	0.98	0	0.95
0.8	1.25	6.16	0.8	0.07	0.19	0.8	0	0.80
0.7	1.43	109.89	0.7	0.11	0.26	0.7	0	0.78

4.1 Reduction in Communication Cost

Organization: A reduction in communication cost (h) increases the value of organization – the cost of asking others for help decreases relative to the cost of acquiring knowledge. As a result, as we move down the panels in Figure 3 from the highest possible cost of communication (recall that organization requires $h < 1$, thus we start from .98), the number of layers of management increases and the proportion of workers who are self-employed decreases. The panels present two cases. In the left hand figures, the cost of acquiring knowledge is high (high t), so production workers acquire no knowledge and always ask higher level workers for help; in the right hand set of figures (low t), the organizational pattern is similar, but workers always acquire some knowledge.

Knowledge: Increasing reliance on organization implies that the maximum knowledge acquired by production workers decreases with h (Proposition 3) – they ask more questions and use their own knowledge less. In contrast, the knowledge acquired by entrepreneurs increases – as communication costs goes down, spans of control increase and managers leverage more their knowledge, raising the marginal value of learning. Overall, lower communication costs imply more centralization – more problems are solved at the top of the hierarchy relative to the bottom.

Inequality: Lower communication costs serve as an equalizer within production workers. Since workers acquire less knowledge the differences in skill translate (by equation (6)) in smaller differences in wages. On the other hand, inequality between layers and within entrepreneurs increases. Since entrepreneurs and managers acquire more knowledge, differences in skill among them translate into larger differences in wages – organization amplifies differences in skill among them. These two effects are specially clear in Table 1 for the case of $t = 1.1$.²¹ Hence, decreases in h reduce within worker wage inequality and increase within manager wage inequality. Looking at overall wage inequality, the first effect dominates for high levels of h and the second for low levels – total wage inequality, measured by the standard deviation of wages, first decreases and then increases as we decrease h . Note that the wage of the lowest ability agent increases as communication costs fall, since the higher span of control results in higher demand for workers.

4.2 Reduction in the cost of acquiring knowledge

Organization: A reduction in the cost of learning to solve problems makes learning less costly relative to communication, and thus makes organization less attractive relative to self-employment. As a result, the proportion of self-employed agents goes up, as we move from the right to the left column of Figure 3. If communication costs are high these changes also imply an increase in spans of control and a decrease in the number of layers as agents learn more and rely less on communication.

Knowledge: Both workers and managers acquire more knowledge since learning costs are lower. Thus the ‘knowledge-intensity’ of production jobs increases, and a larger proportion of problems get solved at lower hierarchical levels – decentralization increases.

²¹For $t = 1.9$ workers do not learn, so there is no within group wage inequality. In fact, the decrease in worker wage inequality, combined with the increase in manager/entrepreneur wage inequality may result in some agents earning less after the improvement in communication costs. This is the case in Figure 4 when $t = 1.1$ and communication costs go from 0.8 to 0.7. The example shows how there may be some losers as technology improves!

Inequality: As Figures 3 and 4 and Table 1 show, the general equilibrium impact of the change is a substantial increase in wage inequality within groups. More knowledge acquired means larger wage differentials, by equation (6). The increase in knowledge acquired is higher at higher levels of the hierarchy, given the increase in leverage allowed by more knowledgeable workers, and thus so is the increase in wage inequality. As a result, overall wage inequality increases for all three levels of communication costs (Figure 4).

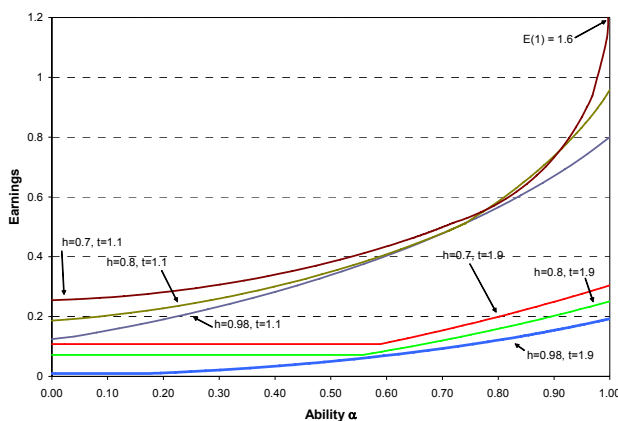


Figure 4: Equilibrium Wage Function

Overall, one of the interesting features of the model is the effect on wages of increases in managerial span. If a firm has many layers so that the highest layer entrepreneur manages a firm with a large number of workers, she can leverage her knowledge immensely which results in very high earnings. This effect can be appreciated in the numerical exercise for $h = 0.7$ and $t = 1.1$. This is the example in which we assume the best information and communication technology. The result is an economy with three layers. There are very few entrepreneurs of layer three. As can be observed in Figure 3, these entrepreneurs do not learn much more than the managers or entrepreneurs of layer two, however, they earn much more. The extent of this increase in wages can be appreciated in Figure 4, where we have plotted the equilibrium wage functions for all the examples presented above.²²

²² As is evident in Figure 4, the wage of agents with ability $\alpha = 0$, when $t = 1.9$, increases as we increase h . These workers cannot solve any problems, the increase in their wage reflects the fact that a higher fraction of problems is solved and so the expected value of production increases. Since layer 0 workers are necessary to produce, the increase in the expected value of production is reflected in their wage. In reality, it is not clear that the gains from a higher expected value of production are assigned to unskilled workers. This is why we focus the analysis on the structure, and not the level, of the earnings function.

5 US Empirical Patterns

The last 30 years have been characterized by large reductions in the cost of accessing and communicating information. First, since the late 60's, we have observed large decreases in the cost of processing information, as observed for example in the steep decline of the price of a transistor (down by several orders of magnitude, as captured by Moore's law). Second, more recently, large reductions in communication costs, particularly due to the introduction and widespread adoption during the late 90's of e-mail, cellular phones, and wireless technology. Our model suggests that two different patterns in the evolution of wages and organization should follow from such exogenous technological changes. A reduction in the cost of processing information (t), such as the one resulting from Moore's law, leads to an increase in the knowledge-content of all jobs and an increase in decentralization (as more problems are solved by those closer to the production floor), a reduction in organization (as self-employment increases), and an increase in wage inequality within worker and manager categories. A reduction in the cost of communication (h), such as the one resulting from wireless, e-mail and cellular phones, should reduce the knowledge content of production jobs. This in turn should decrease wage inequality among workers but, by increasing the importance of organization and the number of layers of management, increase inequality among managers and between managers and workers, as well as the size of hierarchies.

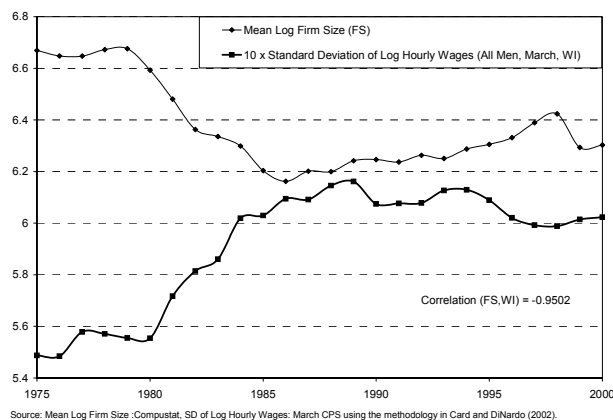


Figure 5: Firm Size and Wage Inequality

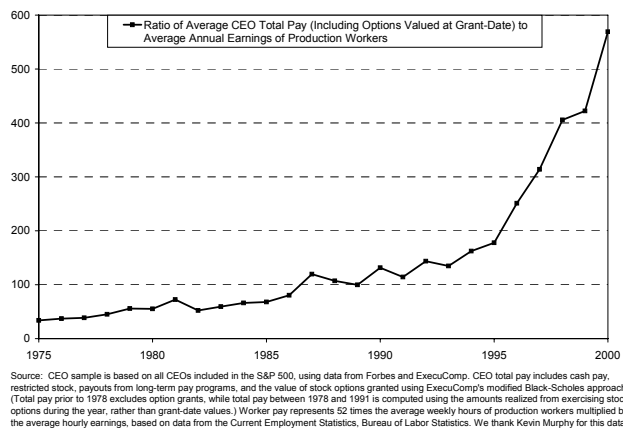


Figure 6: Worker-Manager Wage Inequality

We do indeed observe two differentiated patterns of changes in the evolution of wage inequality in recent data. During the 80's and early 90's, there was a large increase in wage inequality, mostly due to an increase in the demand for skill, and, more specifically, for cognitive skills, and which appears to be correlated with the increasing use of information

technology.²³ Instead, since the second half of the 90's, we observe a slow-down in the generalized increase in wage inequality (see Figure 5), which is replaced by two differentiated trends. While the inequality at the top as measured for example by the 90/50 wage gap, continues to grow, inequality at the bottom (as given by the 50/10 wage gap) has stopped growing or even declined (Murphy and Welch, 2001). A particularly striking version of this pattern of increased inequality at the top is observed in the CEO/worker wage gap, which has continued to increase substantially (see Figure 6).

In contrast, the existence of such differentiated patterns in the organization of work has not been documented. The empirical literature has only documented changes in the internal structure of hierarchy in the 80's and early 90's which are consistent with the comparative statics that result from decreases in t : an increase in the autonomy and responsibility of workers, a reduction of the number of layers of management, and an increase in the managerial span of control. Again, these changes have been found to be associated with the use of information technology.²⁴ Our theory suggests that the impact of communication technology during the last years of the 90's and early 2000's should lead organizations towards increasing centralization, an equalization of routine jobs, and a reduction of self-employment as hierarchies get larger through larger spans of control and more layers. There are no systematic studies of the reorganization of the workplace for the late 90's to which we could contrast the predictions of our theory. Therefore, we offer them as guidance for future empirical work. However, Figure 5 provides some support for the hypothesis that the trends in wage inequality are reflected in trends in organization. It presents the standard

²³Katz and Murphy (1992) are the first paper to show that increases in inequality are consistent with skill biased technological change. Later, Autor, Katz, and Krueger (1998) and Dunne, Haltiwanger, and Troske (1997) have shown that the composition of the labor force within industry and establishments continues to shift towards the more educated workers and more skilled occupations, in spite of raises in returns to skills. Evidence that a similar pattern is seen in other countries is provided by Berman, Bound, and Machin (1998) (see also the survey in Hornstein, Krusell, and Violante (2005)). Murnane, Willet and Levy, (1995) find a higher correlation between earnings and test scores for a more recent panel of graduates than for an earlier one. This finding is particularly clear for high test scores in math. It was Krueger (1993) who first documented a substantial premium associated with computer use, of up to half of the growth in the education premium since the eighties. This finding was confirmed later by Autor, Katz, and Krueger (1997). Lehr and Lichtenberg (1999) have found that larger computer purchases and skill are complementary.

²⁴Bresnahan, Brynjolfsson and Hitt (2002) find, using firm-level data, that greater use of computers is associated with the employment of more-educated workers, greater investments in training, broader job responsibilities for line workers, and more decentralized decision-making. Caroli and Reenen (2001) find evidence of organizational change complementary with increases in demand for skills. In particular, they find evidence of decentralization of authority and a widening of the range of tasks performed by workers. Rajan and Wulf (2003), in a recent paper, present evidence that from 1986 to 1995 firms have become flatter, with less layers of management and that managerial span of control has increased. The evidence from 1995 to 1999 is weaker but suggests, as in our theory, that the flattening stopped and managerial span kept increasing.

deviation of hourly wages together with the mean log firm size from 1976 to 2000 in the US. The correlation between the two series is -0.95 .²⁵ This remarkable negative correlation supports our view that there is a common source of variation driving changes in both of them. Our interpretation is that the underlying source of variation is the costs of acquiring and communicating information: improvements in the cost of accessing knowledge, mostly concentrated in the 80's and early 90's but present throughout the period (e.g. Moore's Law), followed by improvements in communication technology in the late 90's.²⁶

The evidence above suggests that our theory may be capturing some of the effects of information technology on the labor market and the organization of work. Whether it can capture the effects of technological improvements in other periods and countries is an open question that we leave for future research. A promising episode is, however, the turn of the century in the US. This is another period of important improvements in communication technology, in particular, the telegraph, railroads, and the telephone. The evidence in Chandler (1977) seems to suggest that these technological improvements lead to the creation of large enterprises with many layers of management: The modern American firm. Evidence in Goldin and Katz (1995) seems to suggest that this was accompanied by decreases in within worker wage inequality.

6 Conclusion

Paraphrasing Hayek, 'the economic problem of society is to use optimally the available knowledge.' In this paper, we have taken a step towards understanding how economic organization provides incentives for the acquisition and communication of knowledge, how the rewards to cognitive skill depend on the organization of knowledge, and how these rewards are affected by changes in information technology. To do this, we have presented an equilibrium theory of organization and earnings when production requires labor and knowledge and workers are heterogeneous in their cognitive skill. The equilibrium obtained is formed by a universe of knowledge-based hierarchies competing for workers and managers. Among the salient features of the equilibrium are positive sorting, the stratification of agent skill into ranks, and the accentuation by organization of innate ability differentials between

²⁵The correlation of first differences is -0.48 .

²⁶The data in Figure 5 is on firms, and not on hierarchies; recall that only the boundary of the hierarchy, and not of the firm, is determined in our theory. As long as the determinants of the boundaries of the firm are not affected significantly by changes in information technology, changes in the size of hierarchies result in corresponding changes in the size of firms.

the best and worse agents.

Our theory allows us to study the distinct impact of access to information and communication technology. A decrease in the cost of access to stored knowledge leads teams to want to increase worker's knowledge about tasks and, as a result, allows for an increase in managerial span and a reduction in managerial intervention in problem solving. If communication costs are high, these effects lead to a decrease in the number of layers which further enhances the effect on decentralization and reduces firm size. Wage inequality within workers and within managers increases, as the marginal product of skill is proportional to the knowledge acquired. In contrast, as communication technology improves, workers learn less since it is cheaper to ask their managers. This reduction in worker's knowledge results in small differences in worker wages and therefore less wage inequality among them. Managers, in contrast, can deal with more subordinates therefore increasing their spans of control, earnings, and inequality among them. Combined with taller hierarchies, these changes result in larger hierarchies or firms.

The distinction between these two technological changes is relevant to interpret recent qualitative trends in organization and inequality. More broadly, we show that to understand the determinants of wage inequality it is necessary to understand the internal structure of teams. Conversely, our analysis also shows that to understand changes in the organization of production it is necessary to incorporate hierarchies to an equilibrium framework.

The theory has some limitations that should be noted here and, hopefully, will lead to future work. First, our work focuses on heterogeneity and technological changes that take place within a single labor market. In fact, recent technological and institutional changes have allowed for the formation of cross-country teams. We have studied this problem in recent work (Antràs, Garicano, and Rossi-Hansberg, 2005), and refer the reader there. Second, hierarchies do more than acquire and communicate knowledge, as the literature discussed in the introduction suggests. Future work embedding other models of hierarchy in an equilibrium framework is required in order to gauge the performance of our theory. Third, we have focused on the impact of heterogeneity in cognitive ability among agents; future work should study how heterogeneity in communication costs affects equilibrium organization and wages. Finally, we believe that the type of equilibrium model developed in this paper is a first step towards incorporating a richer organizational structure into dynamic equilibrium models. These models could then, for example, be used to study the dynamic relationship between firm size and wage inequality without relying, as in this paper, on comparative statics exercises.

7 References

- Acemoglu, Daron, “Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality.” *Quarterly Journal of Economics*, Vol. 113(4), (November 1998): 1055–1089.
- Antràs, Pol, Luis Garicano and Esteban Rossi-Hansberg, “Offshoring in a Knowledge Economy.” Forthcoming in the *Quarterly Journal of Economics* (2005).
- Autor, David H., Lawrence Katz, and Alan B. Krueger, “Computing Inequality: Have Computers Changed the Labor Market?.” *Quarterly Journal of Economics*, Vol. 113(4), (November 1998): 1169-1213.
- Becker, Gary, *Treatise on the Family*. Cambridge: Harvard University Press, 1981.
- Berman, Eli, John Bound and Stephen Machin. “Implications Of Skill-Biased Technological Change: International Evidence,” *Quarterly Journal of Economics*, Vol. 113(4) (November 1998): 1245-1279.
- Bolton, Patrick, and Mathias Dewatripont, “The Firm As a Communication Network.” *Quarterly Journal of Economics* Vol. 109(4), (November 1994): 809-839.
- Bresnahan Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, “Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence.” *Quarterly Journal of Economics*, Vol. 117(1), (February 2002): 339-376.
- Card, D. and J. E. DiNardo, “Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles.” *NBER Working Paper* No. 8769, (2002).
- Caroli, Eve and John Van Reenen, “Skill Biased Organizational Change? Evidence from a Panel of British and French Establishments.” *Quarterly Journal of Economics*, Vol. 116(4), (November 2001): 1449-1492.
- Chandler, Alfred, *The Visible Hand: The Managerial Revolution in American Business*, Harvard University Press (1977).
- Dunne, Timothy, John Haltiwanger, and Kenneth R. Troske, “Technology and Jobs: Secular Changes and Cyclical Dynamics.” *Carnegie Rochester Series in Public Policy*, (1997): 107-178.
- Fernández, Raquel and Jordi Galí, “To Each According to...? Markets, Tournaments, and the Matching Problem with Borrowing Constraints.” *Review of Economic Studies*, Vol. 66, (1999): 799-824.
- Garicano, Luis, “Hierarchies and the Organization of Knowledge in Production.” *Journal of Political Economy*, Vol. 108(5), (October 2000): 874-904.
- Garicano, Luis and Tano Santos, “Referrals.” *American Economic Review*, Vol. 94(3), (June 2004): 499-525.

- Galor, Oded and Omer Moav, "Ability-Biased Technological Transition, Wage Inequality, and Economic Growth." *Quarterly Journal of Economics*, Vol. 115(2), (May 2000): 469-326
- Galor, Oded and Daniel Tsiddon, "Technological Progress, Mobility and Economic Growth." *American Economic Review*, Vol. 87(3), (June, 1997): 363-382
- Geanakoplos, John and Paul Milgrom, "A Theory of Hierarchies Based on Limited Managerial Attention." *Journal of the Japanese and International Economy*, Vol. 5(3), (September 1991): 205-525.
- Goldin, Claudia and Lawrence Katz, "The Decline of Non-Competing Groups: Changes in the Premium to Education, 1890 to 1940." *NBER Working Paper* No. 5202, (1995).
- Grossman, Sanford and Oliver Hart, "The Cost and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, Vol. 94(4), (August 1986): 691-719.
- Hayek, Friedrich A., "The Use of Knowledge in Society." *American Economic Review*, Vol. 35 (4), (September 1945): 519-530.
- Hornstein, Andreas, Per Krusell, and Gianluca Violante, "The Effects of Technical Change on Labor Market Inequalities." Forthcoming in *The Handbook of Economic Growth* (P. Aghion and S. Durlauf, eds.), Amsterdam, The Netherlands: Elsevier, (2005).
- Katz, Lawrence and Kevin M. Murphy, "Changes in Relative Wages, 1963-87: Supply and Demand Factors." *Quarterly Journal of Economics*, Vol. 107(1), (February 1992): 35-78.
- Kremer, Michael, "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics*, Vol. 108(3), (August 1993): 551-575.
- and Eric Maskin, "Wage Inequality and Segregation by Skill." *NBER Working Paper* No. 5718, August 1996.
- Krueger, Alan B., "How Computers Changed the Wage Structure: Evidence from Micro Data." *Quarterly Journal of Economics*, Vol. 108(1), (February 1993): 33-60.
- Krusell, Per, Lee Ohanian, José-Victor Ríos-Rull, and Gianluca Violante, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica*, Vol. 68(5), (September 2000): 1029-1053.
- Legros, Patrick and Andrew Newman, "Monotone Matching in Perfect and Imperfect Worlds." *Review of Economic Studies*, Vol. 69(4) (October 2002): 925-42.
- Levy, Frank, and Richard Murnane, *Teaching the New Basic Skills*. New York: Free Press. 1996b.
- Lichtenberg, Frank R. and Bill Lehr, "Information Technology and Its Impact on Productivity: Firm-level Evidence from Government and Private Data Sources, 1977-1993." *Canadian Journal of Economics*, Vol. 32(2), (April 1999): 335-362.
- Lucas, Robert E., Jr., "On the Size Distribution of Business Firms." *Bell Journal of Economics*, Vol 9, No. 2, (Autumn 1978): 508-523.

- Moebius, Markus M, "The Evolution of Work", Working Paper, MIT, (1999).
- Murphy, Kevin M. and Welch, Finis, "Wage Differentials in the 1990s: Is the Glass Half-Full or Half-Empty?," F. Welch, *The causes and consequences of increasing inequality*. Chicago and London: University of Chicago Press, (2001): 341-364.
- Murnane, Richard J., John B. Willett and Frank Levy, "The Growing Importance of Cognitive Skills in Wage Determination.", *NBER Working Paper* No. 5076, (1995).
- Qian, Yingyi, "Incentives and Loss of Control in an Optimal Hierarchy." *Review of Economic Studies*, Vol. 61(3), (July 1994): 527-44.
- Rajan, Raghuram G. and Julie Wulf, "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies.", *NBER Working Paper*, No. 9633 (2003).
- Rosen, Sherwin, "The Economics of Superstars." *American Economic Review*, Vol. 71(5), (December 1981): 845-858.
- , "Authority, Control and the Distribution of Earnings." *Bell Journal of Economics*, Vol. 13(2), (Autumn 1982): 311-323.
- Saint-Paul, Gilles,. "On the distribution of income and worker assignment under intrafirm spillovers, with an application to ideas and networks." *Journal of Political Economy*, Vol. 109(1) (February 2001): 1-37.
- Sattinger, Michael, "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature*, Vol. 31, (June 1993): 831-880.
- , "Comparative Advantage and the Distributions of Earnings and Abilities." *Econometrica*, Vol. 43, (May 1975): 455-468.
- Spurr, Stephen, "Referral Practices among Lawyers: A Theoretical and Empirical Analysis." *Law and Social Inquiry*, Vol. 12, (Winter 1988): 87-109.
- Teulings, Coen N., "The Wage Distribution in a Model of the Assignment of Skills to Jobs." *Journal of Political Economy*, Vol. 103(2), (April 1995): 280-315
- , "Comparative Advantage, Relative Wages, and the Accumulation of Human Capital." *Journal of Political Economy*, Vol. 113(2), (April 2005): 425-461.
- Timbergen, Jan, "On the Theory of Income Distribution." *Weltwirtschaftliches Archiv* Vol. 77(2), (1956):155 - 73.
- Van Zandt, Timothy. "Organizations that Process Information with an Endogenous Number of Agents" in *Organizations with Incomplete Information*. Ed. Mukul Majumdar. 239-305. Cambridge: Cambridge University Press, (1998).

8 Appendix

Proof of Proposition 1. We need to show that $da(\alpha)/d\alpha = d\alpha_{l+1}/d\alpha_l > 0$ all $l < L$. We have established in the text (see equation 6) that for $l \leq L$

$$w'(\alpha_l) = z(q_l).$$

Taking the total derivative of this equation with respect to α_l yields

$$\frac{dq_l}{d\alpha_l} = \frac{w''(\alpha_l)}{z'(q_l)}.$$

Doing the same for $l + 1$ we obtain that if $dq_l/d\alpha_l \neq 0$ (it may be the case that $dq_0/d\alpha_0 = 0$, then there are multiple assignments that result in the same allocation and we simply choose the one that exhibits positive sorting),

$$\frac{d\alpha_l}{d\alpha_{l+1}} = \frac{w''(\alpha_{l+1})}{w''(\alpha_l)} \frac{z'(q_l)}{z'(q_{l+1})} \frac{dq_l}{dq_{l+1}},$$

for $l = 0, \dots, L - 1$. From the second line in equation (14) we know that

$$p'(q_l; w) = 1 - h[w(\alpha_{l+1}) + c(\alpha_{l+1}; t)z(q_{l+1})]$$

and so

$$\frac{dq_l}{dq_{l+1}} = -\frac{hc(\alpha_{l+1}; t)z'(q_{l+1})}{p''(q_l; w)}$$

using equation (6). Note that $c'(\alpha_l) = -1$ and $p''(q_l; w) < 0$ in an equilibrium where firms maximize profits. Therefore,

$$\frac{d\alpha_l}{d\alpha_{l+1}} = -\frac{hc(\alpha_{l+1}; t)z'(q_{l+1})}{p''(q_l; w)} \frac{z'(q_l)w''(\alpha_{l+1})}{z'(q_{l+1})w''(\alpha_l)} > 0$$

if $w'(\cdot)$ is monotone. To show that $w'(\cdot)$ is monotone, suppose it is not. Then if $w''(\cdot)$ is continuous, by the Mean Value Theorem there exists a $\bar{\alpha}$ such that $w''(\bar{\alpha}) = 0$. If $w''(\cdot)$ is not continuous, then either there exists a $\bar{\alpha}$ such that $w''(\bar{\alpha}) = 0$, or there exists a $\bar{\alpha}$ such that $w''(\bar{\alpha} - \varepsilon)w''(\bar{\alpha} + \varepsilon) < 0$ for all $\varepsilon > 0$. If there exists a $\bar{\alpha}$ such that $w''(\bar{\alpha}) = 0$, then if $\bar{\alpha}$ is employed in layer l ,

$$\left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l = \bar{\alpha}} = \infty.$$

This, however, contradicts equation (28) since $\phi(\alpha_l)/\phi(\alpha_{l+1})$ is finite and (28) needs to hold point by point. Now consider the case where $w''(\cdot)$ has a discontinuity and changes signs at $\bar{\alpha}$. In that case the above condition implies

$$\left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \uparrow \bar{\alpha}} < 0 \text{ and } \left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \downarrow \bar{\alpha}} > 0 \quad \text{or} \quad \left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \uparrow \bar{\alpha}} > 0 \text{ and } \left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \downarrow \bar{\alpha}} < 0.$$

This, however, cannot maximize firm's profits since at least for one of these choices problem (4) is convex in α_l . The result is a corner solution were several supervisors hire the same subordinate, that is

$$\left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \uparrow \bar{\alpha}} < 0 \text{ and } \left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \downarrow \bar{\alpha}} = 0 \quad \text{or} \quad \left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \uparrow \bar{\alpha}} = 0 \text{ and } \left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l \downarrow \bar{\alpha}} < 0.$$

This again implies a contradiction with (28). Hence, $w''(\cdot)$ is monotone, in fact the above argument shows that $w(\cdot)$ has to be convex, since if not the problem in (4) is convex in abilities which leads

to corner solutions, which are not allowed by the labor market equilibrium condition (28). Notice that this then implies that for all l ,

$$\frac{dq_l}{d\alpha_l} \geq 0.$$

To finish the proof we just need to show that in fact $p''(q_l; w) < 0$. Note however that we can use a similar argument to the one above to show that this has to be the case in equilibrium. First note that by (9) and $z'' > 0$ $p''(q_0; w) < 0$. For $l > 0$, towards a contradiction assume that $p''(q_l; w) \geq 0$, then the problem in (11) given by

$$\max_{q^l} [q_{l+1} - q_l - h(1 - q_l) [c(\alpha_{l+1}; t)z(q_{l+1}) + w(\alpha_{l+1})] + p(q_l; w)],$$

is not concave in q_l (which comes from the fact that the expression is concave in q_{l+1} since $z''(\cdot) > 0$ and the cross-derivative is positive). Thus, manager's profit maximization is solved at a corner. This implies that $dq_l/d\alpha_l = 0$, and therefore that $w''(\alpha_l) = 0$, for an interval with positive Lebesgue measure and $l > 0$. This again implies that

$$\left. \frac{d\alpha_l}{d\alpha_{l+1}} \right|_{\alpha_l = \bar{\alpha}} = \infty \text{ for } l > 0.$$

for an interval of positive Lebesgue measure, which contradicts (28). Therefore, $p''(q_l; w) < 0$ and so in any equilibrium of this economy the assignment exhibits positive sorting in all layers of the hierarchy. \square

Proof of Proposition 2. The proof of Proposition 1 shows that the wage function is convex. Since

$$w'(\alpha_l) = z(q_l) > 0,$$

we also know that the wage function has to be increasing within a layer. From the agent's occupational choice (3) we know that wages have to be continuous, since if not some agents would like to choose a job that is not designed for their ability, which is inconsistent with (6). Thus the wage function is increasing. Note that this result, together with the above equation implies that $q(\cdot)$ is increasing since $z'(\cdot) > 0$ and w is convex. Of course, in general, $q(\cdot)$ will not be continuous since the wage function does not have to be differentiable in the thresholds that divide layers. \square

Proof of Proposition 3. Consider the knowledge choice of a production worker (20),

$$w(\alpha_0) = \max_{q_0} [q_0 - p(q_0; \cdot) - c(\alpha_0; t)z(q_0)]. \quad (31)$$

while if self-employed the worker chooses: $\max_{q_0} [q_0 - c(\alpha_0; t)z(q_0)]$. The worker receives a positive transfer for the problems she does not know and passes on, since $-p(q_0; \cdot) > 0$. The marginal value of worker knowledge is lower with organization ($1 - p'(q_0; \cdot) < 1$) than in self-employment, and the marginal cost is the same. Thus the knowledge the worker acquires, q_0 , is lower than in autarchy, and so is the marginal value of her skill by (6).

Consider now the choice of q of the highest level problem solver (22):

$$w(\alpha_L) = \max_{q_L, q_{L-1}} \left[\frac{q_L - q_{L-1} + p(q_{L-1}; \cdot)}{h(1 - q_{L-1})} - c(\alpha_L; t)z(q_L) \right]. \quad (32)$$

If the worker is self-employed she solves $\max_{q_L} [q_L - c(\alpha_L; t)z(q_L)]$ instead. The marginal value of knowledge of an entrepreneur in (32) is 1 times the number of workers who ask her questions, $1/h(1 - q_{L-1})$ which is greater than 1 (the marginal value of knowledge under self-employment), and the marginal cost is the same as when she is working on her own; thus the knowledge acquired

by an entrepreneur is higher than what it would be if she was working on her own, and so is the marginal value of skill w' from (6). \square

Proof of Proposition 4. Take an equilibrium with a maximum number of L layers. Proposition (1) guarantees that such an equilibrium exhibits positive sorting. Thus, in order to show that the equilibrium exhibits occupational stratification we need to show that self-matching is not an option and that the set of entrepreneurs is a connected set. To show that self-matching is not optimal for an employer note that since by assumption the equilibrium has L layers, $w(\alpha_L) > 0$. If entrepreneur L were to self-match (hire an employee in layer $L - 1$ with ability α_L) she would have an income of

$$-c(\alpha_L; t)z(q_L)$$

which is negative, and where we are using $p(\alpha_L; \cdot) = 0$ since there are entrepreneurs with ability α_L . A parallel arguments goes through for intermediate layers since the term

$$q_l - q_{l-1} + p(q_{l-1}; \cdot) - p(q_l; \cdot)$$

in equation 18 is equal to zero with self-matching. Thus, self-matching never maximizes firm profits.

We now turn to the proof that the set of entrepreneurs is connected, so an equilibrium where the entrepreneur in a team has lower ability than the worker or intermediate manager in another team is not a possibility. Equation (14) implies that $p'(q(\alpha); \cdot) > 0$ for all intermediate managers and workers and by condition (21), $p(q(\alpha); \cdot) \geq 0$ for all $\alpha \geq a^{-1}(1)$. Suppose that the set of entrepreneurs is not connected, that is $\exists \hat{\alpha} \notin [a^{-1}(1), 1]$ such that $p(q(\hat{\alpha}); \cdot) \geq 0$. To economize on notation let $\hat{\alpha}$ be the largest such ability level. Then, since by equation (14), $p'(q(\alpha); \cdot) > 0$ for $\alpha \in B \subset (\hat{\alpha}, a^{-1}(1))$, $p(q(\bar{\alpha}); \cdot) > 0$ for some $\bar{\alpha} \in (\hat{\alpha}, a^{-1}(1))$. Furthermore, $\bar{\alpha}$ works as a worker or middle manager. The earnings of a person with ability $\bar{\alpha}$ are given by

$$w(\bar{\alpha}) = \frac{\bar{\alpha} - a^{-1}(\bar{\alpha}) + p(a^{-1}(\bar{\alpha}); \cdot) - p(q(\bar{\alpha}); \cdot)}{h(1 - a^{-1}(\bar{\alpha}))} - c(\bar{\alpha}; t)z(q(\bar{\alpha})).$$

However, if the same agent works as an entrepreneur, with exactly the same team of agents below her, she would earn,

$$\frac{\bar{\alpha} - a^{-1}(\bar{\alpha}) + p(a^{-1}(\bar{\alpha}); \cdot)}{h(1 - a^{-1}(\bar{\alpha}))} - c(\bar{\alpha}; t)z(q(\bar{\alpha})) > w(\bar{\alpha})$$

where the inequality comes from $p(q(\bar{\alpha}); \cdot) > 0$. A contradiction with the agent's maximization problem in (3). Therefore the set of entrepreneurs is connected and includes the agents with the highest ability. The last part of the proof involves showing that lower ability agents work for lower ability layers of teams. This, however, is immediate from the fact that by a parallel argument to the one used for self-matching, hiring subordinates with higher ability is never optimal since it yields negative earnings. \square

Proof of Proposition 5. To go through this proof it may prove helpful to refer to Figure 2. Fix a maximum number of layers L , a vector of thresholds $\vec{\alpha}_{01}^L \equiv (\alpha_{01}, \alpha_{12}, \dots, \alpha_{L-2L-1}, \alpha_{L-1L-1}, \alpha_{L-1L}, \alpha_{LL+1} = 1)^{27}$ and a vector of wages $\vec{\omega}_0^L = (\omega_0, \omega_1, \dots, \omega_{L-1})$ at abilities $(0, \alpha_{01}, \alpha_{12}, \dots, \alpha_{L-2, L-1})$ correspondingly where

$$\begin{aligned} \omega_0 &\geq \max_{q_0 > 0} [q_0 - c(0; t)z(q_0)] \text{ and} \\ \omega_l &\geq \max_{\{q_i, n_i\}_{i=0}^l} \frac{q_l}{h(1 - q_{l-1})} - c(\alpha_{l-1l}; t)z(q_l) - \sum_{i=0}^{l-1} \frac{h(1 - q_{i-1})}{h(1 - q_{l-1})} [c(\alpha_{i-1i}; t)z(q_i) + \omega_i] \end{aligned}$$

²⁷We are using a notation in which the subindex of the vector of thresholds refers to the subindex of the first element of the vector. The superindex refers to the number of layers.

for $l = 1, \dots, L - 1$. Notice that by Proposition 4 we are focusing only on the relevant thresholds and that we are assuming that the initial wages are higher than the wages that agents could get if they lead the worst possible team (the one formed by the threshold agents) with the number of layers given by their position. Also by definition $\omega_0 < \omega_1 < \dots < \omega_{L-1}$. If wages are below these thresholds then everyone is self-employed or the equilibrium has less than L layers. Given this, we can solve the problems in (4) for all hierarchies with top entrepreneurs in $[\alpha_{L-1, L-1}, \alpha_{L-1, L}]$, and therefore $L - 1$ layers, and with entrepreneurs in $[\alpha_{L-1, L}, 1]$ and L layers. To do so, solve for the level of knowledge q at the thresholds, then use (6) to calculate wages at nearby locations and (29) to calculate assignments. For entrepreneurs use (19) and (21) with equality to calculate their wages and use (11) to calculate transfers. Let $q(\alpha; \vec{\omega}_0^L, \vec{\alpha}_{01}^L)$, $w(\alpha; \vec{\omega}_0^L, \vec{\alpha}_{01}^L)$, $a(\alpha; \vec{\omega}_0^L, \vec{\alpha}_{01}^L)$ and $p(\alpha; \vec{\omega}_0^L, \vec{\alpha}_{01}^L)$ denote the knowledge, wage, assignment and transfer functions obtained. Note that the Theorem of the Maximum ensures that all functions q , w , and a are continuous in the elements of $\vec{\omega}_0^L$, $\vec{\alpha}_{01}^L$, and α within layers. By definition

$$\begin{aligned} w(0; \vec{\omega}_0^L, \vec{\alpha}_{01}^L) &= \omega_0, \\ w(\alpha_{01}; \vec{\omega}_0^L, \vec{\alpha}_{01}^L) &= \omega_1, \\ &\vdots \\ w(\alpha_{L-2, L-1}; \vec{\omega}_0^L, \vec{\alpha}_{01}^L) &= \omega_{L-2}, \end{aligned}$$

and the assignment functions are invertible since they are singled valued and monotone as shown in Proposition (1). Note that in this allocation wages are not continuous and supply and demand for workers do not necessarily equalize at the thresholds (Equilibrium conditions i and iv are not satisfied). We now turn to construct an allocation that satisfies all of them.

Let

$$\alpha_{01}^*(\vec{\omega}_0^L, \vec{\alpha}_{12}^L) \equiv \min \left[\left\{ \alpha : w(\alpha; \vec{\omega}_0^L, \vec{\alpha}_{12}^L) = \omega_1 \right\}, \alpha_{12} \right].$$

Given our restrictions on $\vec{\omega}_0^L$ the function α_{01}^* is well defined (w is an increasing function of α by equation (6) and $\omega_0 < \omega_1$). Since the problem in (4) is continuous in w , by the Theorem of the Maximum it is also a continuous function of the elements of $\vec{\omega}_0^L$. Notice also that since w is continuous in the entries of $\vec{\alpha}_{01}^L$, α_{01}^* is continuous in the entries of $\vec{\alpha}_{12}^L$. Finally, note that $\alpha_{01}^*(\vec{\omega}_0^L, \vec{\alpha}_{12}^L)$ is a strictly decreasing function of ω_0 since w is increasing in α .

Define the excess supply of workers as a function of the vector of thresholds and the wage of the lowest ability worker by

$$\begin{aligned} ES(\vec{\omega}_0^L, \vec{\alpha}_{12}^L) &\equiv \int_0^{\alpha_{01}^*} \phi(\alpha) d\alpha \\ &\quad - \int_{\alpha_{01}^*}^{\alpha_{12}^L} \frac{1}{h \left(1 - q(a^{-1}(\alpha; \vec{\omega}_0^L, (\alpha_{01}^*, \vec{\alpha}_{12}^L)); \vec{\omega}_0^L, (\alpha_{01}^*, \vec{\alpha}_{12}^L)) \right)} \phi(\alpha) d\alpha, \end{aligned} \tag{33}$$

and let $\omega_0^*(\vec{\omega}_1^L, \vec{\alpha}_{12}^L)$ be the wage for agents with ability $\alpha = 0$ such that

$$\omega_0^*(\vec{\omega}_1^L, \vec{\alpha}_{12}^L) \equiv \left\{ \omega : ES((\omega, \vec{\omega}_1^L), \vec{\alpha}_{12}^L) = 0 \right\}.$$

We need to show that $\omega_0^*(\vec{\omega}_1^L, \vec{\alpha}_{12}^L)$ is well defined, continuous and singled valued. First notice that (33) implies that $ES((\omega, \vec{\omega}_0^L), \vec{\alpha}_{12}^L)$ is continuous in ω . Notice also that

$$ES((\omega_1, \vec{\omega}_1^L), \vec{\alpha}_{12}^L) = - \int_0^{\alpha_{12}^L} \frac{1}{h \left(1 - q(a^{-1}(\alpha; \vec{\omega}_0^L, (\alpha_{01}^*, \vec{\alpha}_{12}^L)); \vec{\omega}_0^L, (\alpha_{01}^*, \vec{\alpha}_{12}^L)) \right)} \phi(\alpha) d\alpha < 0$$

since in this case $\alpha_{01}^* = 0$, and

$$ES((\omega_0, \vec{\omega}_1^L), \vec{\alpha}_{12}^L) = \int_0^{\alpha_{12}^L} \phi(\alpha) d\alpha > 0,$$

since in this case workers $\alpha_{01}^* = \alpha_{12}$. Hence, by the Mean Value Theorem, ω_0^* exists.

The function ES is a strictly decreasing function of $\omega \in [\omega_0, \omega_1]$ and hence there is a unique value at which it is equal to zero. To show ES is decreasing in ω note that as ω increases $q(0)$ is not affected (since by (9) ω affects $p(0; \cdot)$ but not $p'(0; \cdot)$, and so the choice of $q(0)$ in (10) is not affected given ω_1). A parallel argument goes through for all $\alpha \in [0, \alpha_{01}^*]$. The result then follows from α_{01}^* being a strictly decreasing function of ω_0 , and ES an increasing function of α_{01}^* . With this result in hand, define ω_0^* as

$$\omega_0^*(\vec{\omega}_1^L, \vec{\alpha}_{11}^L) \equiv \left\{ \omega : ES((\omega, \vec{\omega}_1^L), \vec{\alpha}_{11}^L) = 0 \right\}.$$

The argument above implies that ω_0^* exists, is single valued and continuous in both arguments.

We can proceed sequentially defining functions so that

$$\alpha_{i+1}^*(\vec{\omega}_i^L, \vec{\alpha}_{i+1+i}^L) \equiv \min \left[\left\{ \alpha : w(\alpha; \vec{\omega}_i^L, \vec{\alpha}_{i+1+i}^L) = \omega_{i+1} \right\}, \alpha_{i+1+i}^L \right],$$

$$ES(\vec{\omega}_i^L, \vec{\alpha}_{i+1+i}^L) \equiv \int_{\alpha_{i-1}^*}^{\alpha_{i+1}^*} \phi(\alpha) d\alpha - \int_{\alpha_{i+1}^*}^{\alpha_{i+1+i}^L} \frac{1 - q((a)^{-1}(\alpha); (\omega_0^*, \dots, \vec{\omega}_i^L), (\alpha_{00}^*, \dots, \vec{\alpha}_{i+1+i}^L))}{1 - q(\alpha; (\omega_0^*, \dots, \vec{\omega}_i^L), (\alpha_{00}^*, \dots, \vec{\alpha}_{i+1+i}^L))} \phi(\alpha) d\alpha,$$

and

$$\omega_i^*(\vec{\omega}_{i+1}^L, \vec{\alpha}_{i+1, i+2}^L) = \left\{ \omega : ES((\omega, \vec{\omega}_{i+1}^L), \vec{\alpha}_{i+1, i+2}^L) = 0 \right\},$$

for all $i = 1, \dots, L-2$. Clearly, we are abusing notation and ignoring the dependence of all function on thresholds for which we have obtained an equilibrium function. Clearly all of these functions are well defined, singled valued, and continuous by arguments that parallel the arguments above.

The earnings of entrepreneurs at α_{L-1L-1}^L and α_{L-1L}^L are given by

$$\omega_{L-1L-1}(\alpha_{L-1L-1}^L, \alpha_{L-1L}^L) = \frac{q(\alpha_{L-1L-1}^L) - q(a^{-1}(\alpha_{L-1L-1}^L)) + p(q(a^{-1}(\alpha_{L-1L-1}^L)); \cdot)}{h(1 - q(a^{-1}(\alpha_{L-1L-1}^L)))} - c(\alpha_{L-1L-1}^L; t)z(q(\alpha_{L-1L-1}^L))$$

and

$$\omega_L(\alpha_{L-1L-1}^L, \alpha_{L-1L}^L) = \frac{q(\alpha_{L-1L}^L) - q(\alpha_{L-2L-1}^*) + p(q(\alpha_{L-2L-1}^*); \cdot)}{h(1 - q(\alpha_{L-2L-1}^*))} - c(\alpha_{L-1L-1}^L; t)z(q(\alpha_{L-1L-1}^L)),$$

where we are ignoring the dependence on α_{L-1L-1}^L and α_{L-1L}^L in the notation of all functions. Hence for layer $L-1$ we can do something similar and define

$$\alpha_{L-1L-1}^*(\omega_{L-1}, \alpha_{L-1L}^L) \equiv \min \left[\left\{ \alpha : w(\alpha; \alpha_{L-1L}^L) = \omega_{L-1L-1} \right\}, \alpha_{L-1L}^L \right],$$

$$ES(\omega_{L-1}, \alpha_{L-1L}^L) \equiv \int_{\alpha_{L-2L-1}^*}^{\alpha_{L-1L-1}^*} \phi(\alpha) d\alpha - \int_{\alpha_{L-1L}^L}^1 \frac{1 - q((a)^{-1}(\alpha); \alpha_{L-1L}^L)}{1 - q(\alpha; \alpha_{L-1L}^L)} \phi(\alpha) d\alpha,$$

and

$$\omega_{L-1}^* (\alpha_{L-1L-1}^L) = \{\omega : ES(\omega, \alpha_{L-1L-1}^L) = 0\},$$

where the only difference with the equations for intermediate layers is that entrepreneurs of layer $L-1$ do not form part of the supply of managers of layer $L-1$ and we are solving for a threshold that divides managers and entrepreneurs within a layer, α_{L-1L-1}^* .

We still need to determine the threshold α_{L-1L}^* . Define

$$\alpha_{L-1L}^* \equiv \min [\{\alpha : w(\alpha) = \omega_L(\alpha)\}, 1].$$

Since all other thresholds and wages are now endogenous, we have a wage function (given our abuse of notation noted above) that depends only on α . To show that this function is well defined and singled valued we can use similar arguments as the ones used above.

The construction of an allocation described above is identical for all layers except the last two. Layer $L-1$ is special because it is the only layer that contains a threshold that divides the set of managers and the set of entrepreneurs. Whether or not an equilibrium has L layers depends on whether the above algorithm implies that this threshold is in a corner or internal to layer $L-1$. If it is in a corner, wages are not continuous (apart from the case where in equilibrium the number of entrepreneurs of layer $L-1$ is exactly zero, a point in the parameter space), and the equilibrium number of layers is different than L . As we will show below, modifying the number of layers will eventually lead to an internal threshold. This then determines the equilibrium number of layers.

The construction above allows for two situations in which the allocation found is not an equilibrium. In any other cases the constructed allocation is an equilibrium of our model. First, the case when $\alpha_{L-1L-1}^* = \alpha_{L-1L}^*$ and $w(\alpha; \alpha_{L-1L}^L) \neq \omega_{L-1L-1}$. A case in which there are no entrepreneurs of layer $L-1$. Second the case when $\alpha_{L-1L-1}^* = \alpha_{L-2L-1}^*$ and $w(\alpha; \alpha_{L-1L}^L) \neq \omega_{L-1L-1}$ so $\alpha_{L-1L}^* = 1$. That is, the case where there are no entrepreneurs of layer L . The first case indicates that the equilibrium number of layers is larger than L , the second that the equilibrium number of layers is lower than L . Note that by construction $p(q(\alpha_{L-1L-1}^*)) = 0$ only if $w(\alpha; \alpha_{L-1L}^L) = \omega_{L-1L-1}$. In order to complete the proof we need to show that there exists an L for which $\alpha_{L-2L-1}^* < \alpha_{L-1L-1}^* < \alpha_{L-1L}^*$ and so $w(\alpha; \alpha_{L-1L}^L) = \omega_{L-1L-1}$ and $p(q(\alpha_{L-1L-1}^*)) = 0$.

Suppose we are in the first case so the number of layers is larger than L . Now consider the wage of the highest ability agent in a candidate equilibrium with L layers

$$w(1; L) = \frac{q(1; L) - q(a^{-1}(1; L); L) + p(q(a^{-1}(1; L); L); L)}{h(1 - q(a^{-1}(1; L); L))} - c(1; t)z(q(1; L)).$$

and note that

$$\lim_{L \rightarrow \infty} w(1; L) = \lim_{L \rightarrow \infty} \left[\frac{p(q(a^{-1}(1; L); L); L)}{h(1 - q(a^{-1}(1; L); L))} - c(1; t)z(q(1; L)) \right],$$

since

$$\lim_{L \rightarrow \infty} \frac{q(1; L) - q(a^{-1}(1; L); L)}{1 - q(a^{-1}(1; L); L)} = 0. \quad (34)$$

To prove this last statement, note that $\{q(\alpha; L)\}_{L=1}^{\infty}$ is a bounded and monotone sequence and so it converges. Hence, either (34) is satisfied or $\lim_{L \rightarrow \infty} q(a^{-1}(1; L); L) \rightarrow 1$ since

$$\lim_{L \rightarrow \infty} [q(1; L) - q(a^{-1}(1; L); L)] = 0.$$

Towards a contradiction assume that $\lim_{L \rightarrow \infty} q(a^{-1}(1; L); L) \rightarrow 1$. The total derivative of equation (12) implies that

$$\frac{dq(1; L)}{dq(a^{-1}(1; L); L)} = \frac{1}{c(1; t)z''(q(1; L))(1 - q(a^{-1}(1; L); L))}.$$

Then $\lim_{L \rightarrow \infty} q(a^{-1}(1; L); L) \rightarrow 1$ implies that $\lim_{L \rightarrow \infty} [q(1; L) - q(a^{-1}(1; L); L)] \neq 0$, a contradiction.²⁸ Hence (34) is satisfied and so $p(q(a^{-1}(1; L); L); L) > 0$ for sufficiently large L in order for $w(1; L) > 0$. Given that $\lim_{L \rightarrow \infty} [q(1; L) - q(a^{-1}(1; L); L)] = 0$, this implies that $p(q(1); L) > 0$, and so, since by the Theorem of the Maximum p is continuous and q is continuous between layers, that there exists an L^* such that $p(q(\alpha)) = 0$ for some $\alpha \in [\alpha_{L-2L-1}^*, \alpha_{L-1L}^*]$, and by construction $\alpha = \alpha_{L-1L-1}^*$. Hence, there exists an equilibrium allocation with L^* layers.

Suppose alternatively that we are in the second case so that the number of layers is lower than L . Consider the case with 1 layer. Then $p(q(\alpha)) = q(\alpha) - c(\alpha; t)z(q(\alpha)) - w(\alpha)$. If $p(q(\alpha)) \geq 0$ then $L = 0$ and we have found an equilibrium number of layers. If not then $p(q(\alpha)) < 0$ and $L > 0$. Again the reasoning above implies that there is an $1 < L^* < L$ such that $p(q(\alpha)) = 0$ for some $\alpha \in [\alpha_{L-2L-1}^*, \alpha_{L-1L}^*]$. Thus, there exists an equilibrium allocation with L^* layers.

We turn now to show that the constructed equilibrium allocation is unique. The arguments above shows how to construct an equilibrium allocation, given the results in Propositions 1, 2, and 4. Furthermore, the arguments used in the proof of Proposition 2 imply that the wage function is convex and such that the solution to the problem in (4) is unique. Hence, this construction yields a unique allocation since all thresholds, and wages at these thresholds, are uniquely determined. Therefore, we have found the unique allocation with the properties described in Propositions 1, 2, and 4, and since any equilibrium has to exhibit these properties, the equilibrium is unique. \square

Proof of Proposition 6. First, note that the First Welfare Theorem applies using the standard argument. That is, consider a competitive equilibrium with wages $w(\alpha)$. Suppose that there is an alternative feasible allocation where some workers are allocated to different firms and where output is higher. This, however, constitutes a contradiction of the definition of competitive equilibrium, since in equilibrium firms maximize profits given $w(\alpha)$. Intuitively, there exist complementarities, (see Section 3.1), but they are priced (in the firm case, they are internalized by the firm) and thus no improvement on the competitive equilibrium is possible.

Although the general argument holds, writing the Social Planner's problem is illustrative and provides an alternative way of proving the result. To do this, notice first that the existence of complementarities between the skills of agents at consecutive layers implies that the optimal allocation must exhibit positive sorting. Towards a contradiction, suppose it does not. Then there exist a pair of managers, where the less skilled one hires better subordinates. The planner would then prefer to switch the subordinates of these two managers and assign the talented subordinates to the talented manager, since complementarity in production implies that this reallocation increases total output. The optimization problem is then given by

$$q(\cdot), \left\{ \alpha_{il}^*, \alpha_{i+1}^* \right\}_{l=0, L}^L \int_{\cup_{l=0}^L A_{lE}} \frac{q(\alpha)}{h(1 - q(a^{-1}(\alpha)))} \phi(\alpha) d\alpha - \int_0^1 c(\alpha; t) z(q(\alpha)) \phi(\alpha) d\alpha$$

where $A_{0M} = [0, \alpha^{00}]$, $A_{lM} = [\alpha_{l-1}^*, \alpha_{il}^*]$, $A_{lE} = [\alpha_{il}^*, \alpha_{i+1}^*]$, $A_{LE} = [\alpha_{LL}^*, 1]$, and $A_M = \cup_{l=0}^L A_{lM}$ subject to

$$a'(\alpha) = \begin{cases} \frac{1 - q(\alpha)}{(1 - q(a^{-1}(\alpha)))} \frac{\phi(\alpha)}{\phi(a(\alpha))} & \text{for } \alpha \in A_M \setminus A_{0M} \\ h(1 - q(\alpha)) \frac{\phi(\alpha)}{\phi(a(\alpha))} & \text{for } \alpha \in A_{0M} \end{cases}$$

$$a_0^*(0) = \alpha_{01}^*, \quad a(\alpha_{il}^*) = \alpha_{i+1l+2}^*, \quad a(\alpha_{i+1}^*) = \alpha_{i+1l+2}^*, \quad \text{and } a(\alpha_{L-1L-1}^*) = 1 \quad \forall l = 0, \dots, L-2.$$

Note that the first term is just the integral over all entrepreneurs of $q_L n_0$ after substituting (5). Let $\lambda(\alpha)$ be the Lagrange multiplier associated with the first constraint (where we incorporate below

²⁸Note that this reasoning goes through only because knowledge is cumulative. If knowledge was not cumulative we would need to impose more restrictions on the support of f .

$\phi(\alpha)/\phi(a(\alpha))$ as part of the co-state). Then, the maximum principle's necessary conditions for this problem are given by

$$\frac{1}{h(1-q(a^{-1}(\alpha)))} - c(\alpha; t) z'(q(\alpha)) = 0 \text{ for } \alpha \in A_{IE}, \quad (35)$$

$$-c(\alpha; t) z'(q(\alpha)) (1 - q(a^{-1}(\alpha))) \phi(\alpha) + \lambda(\alpha) = 0 \text{ for } \alpha \in A_M \setminus A_{0M}, \quad (36)$$

$$-c(\alpha; t) z'(q(\alpha)) \phi(\alpha) + h\lambda(\alpha) = 0 \text{ for } \alpha \in A_{0M}. \quad (37)$$

The co-state variable $\lambda(\alpha)$ is a continuous function that satisfies

$$\frac{\partial \lambda(\alpha)}{\partial a(\alpha)} = \frac{q'(a(\alpha))}{h(1-q(\alpha))} \phi(\alpha) = c(a(\alpha); t) z'(q(a(\alpha))) q'(a(\alpha)) \phi(\alpha) \text{ for } a(\alpha) \in A_{IE}, \quad (38)$$

$$\frac{\partial \lambda(\alpha)}{\partial a(\alpha)} = \lambda(a(\alpha)) \frac{q'(a(\alpha))}{(1-q(\alpha))} \phi(\alpha) = c(a(\alpha); t) z'(q(a(\alpha))) q'(a(\alpha)) \phi(\alpha) \text{ for } \alpha \in A_M, \quad (39)$$

where the second equality in each equation follows from substituting the first order conditions. All other conditions are either pinned down by the constraints or by the continuity of $\lambda(\alpha)$ (for example p continuous and therefore that the set of entrepreneurs is connected).

Comparing (35) – (37) with (12) – (14), combined with (16) and (17) to obtain an Euler equation that is not a function of p , we obtain that in order for the equilibrium allocation to be efficient

$$\frac{\lambda(\alpha)}{\phi(\alpha)} = c(a(\alpha); t) z(q(a(\alpha))) + w(a(\alpha))$$

then

$$\frac{\partial \lambda(\alpha)}{\partial a(\alpha)} = [c'(a(\alpha); t) z(q(a(\alpha))) + c(a(\alpha); t) z'(q(a(\alpha))) q'(a(\alpha)) + w'(a(\alpha))] \phi(\alpha)$$

and so

$$\frac{\partial \lambda(\alpha)}{\partial a(\alpha)} = [c'(a(\alpha); t) z(q(a(\alpha))) + w'(a(\alpha)) + c(a(\alpha); t) z'(q(a(\alpha))) q'(a(\alpha))] \phi(\alpha)$$

The equations above combined with (38) and (39) then imply that

$$w'(\alpha) = c'(\alpha; t) z(q(\alpha)),$$

as in (6). Since all other optimality conditions are then equivalent to the equilibrium conditions this implies that the proof of Proposition 5 applies and so the equilibrium allocation is Pareto Optimal. \square