Organizing Growth*

Luis Garicano
University of Chicago

Esteban Rossi-Hansberg
Princeton University

August 6, 2007

Abstract

We present a model of endogenous growth in which agents acquire knowledge as organizations develop and allocate labor more efficiently. Building up organizations and acquiring the relevant pieces of complementary knowledge, takes time so organizations develop slowly. As the technology is better known and the majority of problems faced in production are well understood, incremental knowledge is less and less useful: productivity increases at a decreasing rate. But the depth of expertise accumulated facilitates the appearance of new, radical innovations. Eventually incremental growth slows down sufficiently that agents choose to switch to a new technology, making existing organizations obsolete. We show that better communication technology increases the long term growth rate of the economy as larger organizations can organize labor and knowledge more efficiently over longer periods. Our model can rationalize the cross-sectional relation between organization and development. It can also help explain the absence of returns to scale in research and the existence, which we document, of large long term technological cycles.

*The impulse for this paper grew out of several conversations with Philip Aghion, for which we are grateful. We thank Lorenzo Caliendo for excellent research assistance.

1. INTRODUCTION

Economic development is linked to the development of organizations that coordinate the role individuals play in production and the knowledge they acquire.[1] Modern economies are characterized by a complex network of organizations that in some cases span the whole world. In contrast, underdeveloped economies today and in the past are characterized by simple organizations, like the ones required in traditional agriculture. The more complex and sophisticated these organizations the more efficiently can current technologies be exploited, as organizations allocate the effort of individuals to more productive uses and give them incentives to acquire more knowledge. This knowledge accumulation allows agents to develop new radical innovations that underlie economic growth.

Indeed, the process of innovation is inextricably linked to the emergence and growth of new organizational hierarchies. Radical technological change, the one that takes place when a truly new technology is introduced, usually occurs outside of existing hierarchies. For example, associated with the arrival of the electricity at the end of the XIX century were notably Edison General Electric (now GE) and Westinghouse; associated with the development of the automobile a few years later were Ford and General Motors; with the development of film, Kodak; with the arrival of the computer and microprocessors are first IBM and then Intel; finally, with the development of the World Wide Web, Google, Yahoo, Amazon and E-Bay. As the new firms grow in complexity and in size, technology is refined and deepened, in a process of incremental change that results initially in large gains in productivity and eventually exhausts the improvements available. Thus larger firms with deeper hierarchies are not just the consequence, as Adam Smith argued, of larger market size making more specialization optimal; but also the cause of deeper, more intensive utilization of existing technologies, and of the development of radical innovations.[2]

We present a theory of development through organization in which the rate of innovation, the extent of knowledge accumulation, and the amount of organization

[1] We refer broadly to organizations as encompassing more than firms. Specifically, the knowledge transfers we study may be mediated through consulting markets, referral markets or indeed firms.

[2] Allyn Young (1928) first observed that the division of labor in turns brings about new knowledge and thus results in new growth, an insight on which we build.

in the economy are jointly determined. When a new innovation is discovered, the technology is known badly and entrepreneurs work on their own. Depending on the level of communication and coordination costs, the entrepreneur may set up an organization in which specialized problem solvers (e.g. consultants or managers) deal with the less common and harder problems. As more layers of problem solvers are added, organizations become more complex and more knowledge is optimally accumulated, which increases output per worker. However, as the technology is better known and the majority of problems are understood, incremental knowledge is less and less useful; thus productivity increases at a decreasing rate. However, the depth of accumulated expertise facilitates the appearance of new, radical innovations. That is, while knowledge improves as agents invest in trying to solve the problems posed by the current technology, radical innovation emerges as a by-product of this improvement.

The dynamics in our theory are the result of the time required to build organizations. We model economic organization as a collection of markets for expert services (a referral market) where only one of these markets can develop in any given period. These referral markets could be equivalently seen as consulting market arrangements or inside-the-firm hierarchies, as we have shown elsewhere.[3] Many factors prevent the instantaneous appearance of an organization that can exploit all the potential value of the existing technology. Two are most important. First, it is impossible to know what are the problems that will prove important in the next cycle of innovation. Second, agents have to be trained in the basic knowledge of the current technology before others can be trained in the more advanced knowledge; in fact, learning how to deal with the rare and advanced problems may not be useful if there are no agents specialized in simple tasks who can actually ask the right questions. For example, experts in internet marketing or sophisticated wireless networks became available only after the internet was developed and there was a demand for their services.

Progress thus takes place in leaps and bounds. A new, radical, innovation takes place, and then all the effort is placed on that innovation, as first the more productive pieces of knowledge and then the more esoteric ones are attained. Radical innovation will not take place again until the current innovation has been exploited to a certain degree. The appearance of a radical innovation is not exogenous; instead, agents can

[3] See Garicano and Rossi-Hansberg (2006).

choose between developing existing technologies and trying out for a new technology; as long as the value of continuing on an existing innovation is sufficiently high, the switch to the new technological generation does not take place. Adopting the new technology makes all the existing knowledge acquired about the previous technology obsolete, and thus requires agents to start accumulating new knowledge and start building new organizations.

In our theory, positive transitory technology shocks have permanent effects in output. A good technological shock leads to the immediate adoption of the new technology, who then becomes the technology in use and is exploited by expanding the number of layers in the hierarchy. When, instead, a bad technological shock takes place, the economy continues on its previous path of incremental innovation, as the economy expands existing organizations to exploit the current technology further, rather than use a bad new technology. In this sense, society can extract an option value from the radical innovation process. It can use new great ideas but it can discard bad ones by exploiting the current technology further. Thus an increase in volatility in the quality of new innovations increases the growth rate of the economy.

Our model captures three important aspects of knowledge. First, a technology can be replicated by everyone and new technologies are a by-product of the new knowledge acquired to solve the problems posed by the current technology. Hence, the production of new technologies implies an externality, as a new idea raises the output of all agents. As in Lucas (1988) individuals cannot capture individually the gains from all of their knowledge. Second, production within a technology exhibits decreasing return in knowledge, as knowledge is costly to communicate and so large organizations need to be built to exploit it. To solve a problem agents must communicate their knowledge to other agents. But each agent's ability to communicate is limited by his available time. So acquiring pieces of knowledge that are increasingly arcane requires larger and more complex organizations for such pieces of knowledge to be efficiently used. The gains from this knowledge are appropriable, given the time constraints of individuals (which prevents free replication) and the need to individually communicate solutions. Hence, investment in knowledge to solve problems is socially optimal, conditional on the organizational size. The diffusion of the gains obtained from this additional knowledge takes place through the increase in the prices charged by all agents to communicate production projects or problems. Third, inherently in new knowledge is a process of creative destruction (Schumpeter, 1942) whereby adopting a radical innovation makes the existing hierarchy obsolete.[4]

Our theory has four main empirical implications. First, exogenous changes in communication and coordination costs, and in the thickness of the tail of the distribution of problems, affect the rate of growth through economic organization. Consider differences in communication costs. Compare the result of the appearance of an innovation in a country with low population density (high communication costs) with the appearance of the same innovation in a country with high population density. In the low density country, organization and specialization is costly. The entrepreneur develops the innovation in as much as its efficient on his own, and the process stops there. In the high density one, organizations emerge to take advantage of the innovation, which allows for efficient use of the technology as the organization deepens the knowledge of this innovation. This allows in turn for longer, more creative leaps in the new technology. Thus, communication and coordination costs are critical both to the level of growth and to the observed organizational structure. Second, scale effects are absent. Empirically, as Jones (1995) first observed, adding more agents to the research side of the economy (in our case the allocation of more agents as problem solvers) does not necessarily increase the rate of growth. In our model, there are no scale effects on average. In fact, we actually obtain larger shares of agents in the 'knowledge' sector in the final periods of an existing technology, where the value of the extra innovations is the lowest.[5] Third, productivity moves in cycles, involving large gains at the initial stages of the organization of a radical innovation, and decelerating for a long period of time as the decreasing returns involved in increasingly deep exploitation of existing technologies take place. Fourth, in the absence of institutions that support the formation of the relevant markets in which agents can exchange their services, development stagnates as agents do not have incentives to acquire new knowledge.

[4] Previous models of creative destruction, following on the pioneering work of Aghion and Howitt (1992) and Grossman and Helpman (1991) do not take organizations into account– new products substitute the old, and monopoly rents disappear in those models, but organizations play no role.

[5] The main previous existing explanations of this puzzle are Kortum (1997), Young (1998) and Howitt (1999). The first one's explanation is that the first ideas are low hanging fruit, and as these ideas are exploited, future innovations become increasingly costly to find. The other two papers focus on the increase in the amount of varieties as innovation increases, which leads to an increase in the innovation cost, as workers must improve a larger number of products.

Our work has two main precedents on top of the seminal endogenous growth theories of Lucas (1988), Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992).[6] Becker and Murphy (1992) first studied the connection between coordination costs and growth through economic organization. Unlike in their model, we specifically take into account the knowledge accumulation process and the occupational distribution and organization that results. We also differ in differentiating between the normal, incremental process of innovation, which suffers from decreasing returns, and the exceptional radical steps. Second, Jovanovic and Rob (1990) develop a theory in which growth is generated by small innovations within a technology and large innovations across technologies. In their framework, alternative technologies are random and are not affected by the choices made within the current technology. In this sense, our theory endogenizes the quality of alternative technologies and adds organization as a source of growth. Jovanovic and Rob (1989) present a model in which communication technology also affects growth through the search process, not, again, through organizations.

One way to understand our contribution is as a theory of endogenous growth in labour-augmenting technology. Take a standard production function of the form $AF(K, HL)$ where A denotes TFP, K capital, H labor-augmenting technology (human and organizational capital) and L labor. In Sections 2.1 we present a theory of the evolution of H, where H increases as deeper organizations are formed and the technology A is exploited more efficiently. The evolution of H exhibits decreasing returns and so permanent growth can never be the result of more complex and efficient organizations. Even though H exhibits decreasing growth rates, the amount of new knowledge acquired by some agents every period converges to a constant. Hence, even though the returns to knowledge decrease, the range of new knowledge does not (although less agents acquire this new knowledge). But this knowledge also generates radical innovations or improvements in A. So sporadically a new technology is invented and the evolution of H starts all over again. This is the problem that we analyze in Section 2.2. Section 3 adds random shocks either to the technology A or to the

[6] Earlier precedents are Kondratrieff, who first observed the presence of long (40 years) technology induced cycles and Schumpeter (1939) who suggested that the phenomenon was linked to the bunching of incremental and radical innovation. His explanation for this bunching has to do with changes in the distribution of entrepreneurial ability; this is unsatisfactory, as Kuznets' book review (1940) first pointed out.

quality of random innovations. In Section 5 we incorporate our model of the evolution of labour augmenting technology or organizational capital in a standard neoclassical framework with capital accumulation where the production function is exactly of the $AF(K, HL)$ form and discuss the evolution of capital. Section 4 presents some empirical evidence on the long-term technology cycles implied by our theory and the relationship between organization and development.

2. THE MODEL

The economy is populated by a mass of size 1 of ex-ante identical agents that live for one period. Every period an identical set of agents is born. Agents have linear preferences so they maximize their income or the consumption of the unique good produced in the economy.

At the start of the period agents choose an occupation and a level of knowledge to perform their job. Agents can either work in organizations that use the current prevalent technology, or they can decide to invent a new technology. The quality of the new technology will depend on the level of knowledge the economy has acquired in the past. If they decide to work in an organization using the current technology, they need to decide what role to perform.

A technology is a method to produce goods using labor and knowledge. One unit of labor generates a project or problem. To produce, agents need to have the knowledge to solve the problem. If they do, they solve the problem and output is produced. If the worker does not know the solution to the problem, she has the possibility to transfer or sell the problem or project to another agent that may have the knowledge to solve it. Organizations are hierarchical, they have one layer of workers and potentially many layers of problem solvers (as in Garicano and Rossi-Hansberg, 2006). Problem solvers have more advanced knowledge than workers and so are able to solve more advanced problems, but they need to 'buy' these problems from workers or lower layer problem solvers since they do not spend time producing but finding out about existing problems.

A technology is used more intensely the more layers in the organization. In the first period a technology is in use, agents learn basic knowledge to develop it and they work as production workers. Since higher layers of management have not been developed, the problems they cannot solve go to waste. In the next period, agents observe that in the last period some valuable problems were thrown away and some of them decide to work as first layer problem solvers. These problem solvers, in turn, throw away some valuable problems that they cannot solve. This induces the entry of second layer managers in the next period. This process goes on making the hierarchy taller as time proceeds and the use of the prevalent technology more efficient through a better allocation of workers. Of course, the knowledge acquired by agents in all layers will depend on the number of layers in the organization as well as the fees or prices for transferring problems. The price at which an agent with a particular level of knowledge can sell a problem is a measure of the efficiency of the organizational structure in exploiting a technology. As we will see, the more organizational layers, the higher the price and so the more efficient is the organization in allocating labor and knowledge.

As emphasized in Garicano and Rossi-Hansberg (2006) there are many equivalent ways of decentralizing these organizations. First, as here, there can be market for problems and agents sell and buy problems for each other at a market price. Alternatively, there can also be firms that optimally organize as hierarchies and hire workers and managers for particular positions at a wage given their knowledge level. Finally, organizations can also be interpreted as consulting markets in which workers hire knowledgeable agents as consultants to solve problems for them for a fee. All of these interpretations are equivalent and can exists at the same time. In all of them agents obtain the same earning and perform the same roles. In what follows we emphasize the case when there are markets for problems and problem solvers buy them, but we may as well talk about firms and managers.

We now turn to the description of the formation of organization and the use of a technology. We then study the decisions of agents to drop the technology currently in use and make a radical innovation instead of going deeper in the development of the current technology (add a new layer).

2.1. Organization within a given Technology

Suppose a new technology $A \geq 1$ is put in place at time $t = 0$. The evolution of this technology will be our main concern in the next section. For now we just keep it fixed. Obtaining A units of output from this technology requires a unit of time and a random level of knowledge. An agent specialized in production uses his unit of time to generate one problem, which is a draw from the probability distribution $f(z)$. We assume that $f(z)$ is continuous and decreasing, $f'(z) < 0$, with cumulative distribution function $F(z)$. The assumption that $f'(z) < 0$ guarantees that agents will always start by learning how to solve the most basic and common problems.[7] In order to produce, the problem drawn must be within the workers' knowledge set, if it is not, then no output is generated. Knowledge can be acquired at a constant cost $\widetilde{c} > 0$, so that acquiring knowledge about problems in $[0, z]$ costs $\widetilde{c}z$. Denote the wage in period $t \in \{0, 1, ...\}$ of an agent working in layer $\ell \in \{0, 1, ...\}$ of an organization with highest layer L by $w^{\ell}_{L,t}$. Then, the earnings of a production worker (layer 0) working on a new idea (so the highest layer in the organization is $L = 0$) at time 0 are:

$$w^0_{0,0} = \max_z AF(z) - \widetilde{c}z,$$

where $AF(z)$ is total output by workers with ability z (they solve a fraction $F(z)$ of problems each of which produces A units of output) and $\widetilde{c}z$ is the cost of acquiring knowledge z. Denote by $z^0_{0,0}$ the level of knowledge that solves the problem above (where the notation is analogous to the one for wages). Note that an organization with only workers of layer zero will leave unsolved a fraction of problems $1 - F\left(z^0_{0,0}\right)$. These problems, if solved, would produce output $A(1 - F\left(z^0_{0,0}\right))$. But this simple organization, where workers only work by themselves, chooses optimally to discard them.

In order to take advantage of the discarded problems next period, $t = 1$, some agents will decide to buy the discarded problems from workers as long as they can then solve some of them and obtain higher earnings. The assumption is that these agents need to first see that valuable problems are discarded to enter next period and take advantage of them. Agents can communicate the problems they did not solve in exchange for a fee or price. If communication is cheaper than drawing new problems, then some agents may find it in their interest to specialize in learning about unsolved

[7] That is, $f'(z) < 0$ will be chosen by agents if they can sequence the knowledge acquired optimally.

problems; they pay a price for these problems, but in exchange they can solve many of them as they do not need to spend time generating the problems, only communicating with the seller. Organization makes it, potentially, optimal to learn unusual problems, as agents can amortize this knowledge over a larger set of problems.

Thus at time $t = 1$ agents have a choice between becoming production workers or specialized problem solvers. If they become production workers they earn

$$w_{1,1}^0 = \max_z AF(z) + (1 - F(z))r_{1,1}^0 - \widetilde{c}z \tag{1}$$

where $r_{1,1}^0$ is the equilibrium price at which workers in layer 0 sell their problems. As problem solvers they need to spend their time communicating with workers to find out about the problems they are buying. The number of problems a manager can buy is given by the communication technology. Let h be the time a problem solver needs to communicate with a worker about a problem. Then, a problem solver has time to find out, and therefore buy, $1/h$ problems. Clearly h is a key parameter of the model that determines the quality of communication technology. The manager knows that workers only sell problems that they cannot solve, so he knows that all problems sold by workers will require knowledge $z > z_{1,1}^0$ (where $z_{1,1}^0$ solves the problem above). Hence, the manager acquires knowledge about the more frequent problems above $z_{1,1}^0$. The wage of the layer one problem solver is then given by

$$w_{1,1}^1 = \max_z \frac{1}{h}\left(A\frac{F(z_{1,1}^0 + z) - F(z_{1,1}^0)}{1 - F(z_{1,1}^0)} - r_{1,1}^0\right) - \widetilde{c}z.$$

Namely, they buy $1/h$ problems at price r_0 and solve a fraction $\left(F(z_{1,1}^0 + z_{1,1}^1) - F(z_{1,1}^0)\right) / \left(1 - F(z_{1,1}^0)\right)$ of them, each of which produces A units of output. On top of this, they pay the cost of learning the problems in $\left[z_{1,1}^0, z_{1,1}^0 + z_{1,1}^1\right]$. As long as $r_{1,1}^0 > 0$, the value of the problems that were being thrown out was positive, and so $w_{0,0}^0 < w_{1,1}^0 = w_{1,1}^1$, where the last equality follows from all agents being identical ex-ante. Hence, if $r_{1,1}^0 > 0$ adding the first layer of problem solvers is optimal at time $t = 1$. We will show below that in equilibrium under some assumptions on F, $r_{1,1}^0$ is in fact positive. Note also that agents in layer 0 will choose to acquire less knowledge as we add a layer of problem solvers: It is not worth it to learn as much since unsolved problems can now be sold at a positive price.

Next period, $t = 2$, agents observe that some valuable problems were thrown away last period. Namely, a fraction $1 - F(z_{1,1}^1)$ of problems. Hence, some agents enter as managers of layer 2 to buy these problems from problem solvers of layer 1. This process continues, adding more layers each period, as long as some valuable problems are thrown away and agents can acquire enough knowledge to solve them and earn higher wages. Hence, each period this economy potentially adds another layer of problem solvers. More unusual and specialized problems are solved and society acquires a larger and larger range of knowledge.

To avoid repetition, we write the problem for period $t = L$ when the hierarchy has a maximum layer L. As described above, production workers earn

$$w_{L,L}^0 = \max_z AF(z) + (1 - F(z))r_{L,L}^0 - \widetilde{c}z.$$

Call $Z_{L,t}^\ell$ the cumulative knowledge of agents up to layer ℓ, in a period t where the maximum number of layers is L : $Z_{L,t}^\ell = \sum_{i<\ell} z_{Lt}^i$. A problem solver of layer ℓ where $0 < \ell < L$ earns

$$w_{L,L}^\ell = \max_z \frac{1}{h}\left(\frac{A\left(F(Z_{L,L}^{\ell-1} + z) - F(Z_{L,L}^{\ell-1})\right) + \left(1 - F(Z_{L,L}^{\ell-1} + z)\right) r_{L,L}^\ell}{\left(1 - F(Z_{L,L}^{\ell-1})\right)} - r_{L,L}^{\ell-1}\right) - \widetilde{c}z,$$

where $r_{L,t}^\ell$ is the price of a problem sold by an agent in layer ℓ in an economy with organizations of $L + 1$ layers at time t. Note that intermediate problem solvers both sell and buy problems. They buy $1/h$ problems at price $r_{L,L}^{\ell-1}$ and sell the problems they could not solve (a fraction $\left(1 - F(Z_{L,L}^{\ell-1} + z)\right) / \left(1 - F(Z_{L,L}^{\ell-1})\right)$ at price $r_{L,L}^\ell$. Problem solvers in the highest layer L cannot sell their problems as there are no buyers, so their earnings are just given by

$$w_{L,L}^L = \max_z \frac{1}{h}\left(A\frac{F(Z_{L,L}^{L-1} + z) - F(Z_{L,L}^{L-1})}{1 - F(Z_{L,L}^{L-1})} - r_{L,L}^{L-1}\right) - \widetilde{c}z.$$

In what follows we will use an exponential distribution of problems. This will allow us to simplify the problem above substantially and will guarantee that the prices of problems at all layers are positive. Hence, absent a new technology, as time goes to infinity the number of layers also goes to infinity. In the next section we will introduce radical innovations that will prevent this from happening. For the moment, however, we continue with our technology A.

Let $F(z) = 1 - e^{-\lambda z}$. Then the earnings of agents in the different layers can be simplified to

$$\begin{aligned} w_{L,L}^0 &= \max_z \left(A - e^{-\lambda z}\left(A - r_{L,L}^0\right)\right) - \widetilde{c}z, \\ w_{L,L}^\ell &= \max_z \frac{1}{h}\left(\left(A - r_{L,L}^{\ell-1}\right) - e^{-\lambda z}\left(A - r_{L,L}^\ell\right)\right) - \widetilde{c}z \text{ for } 0 < \ell < L, \\ w_{L,L}^L &= \max_z \frac{1}{h}\left(\left(A - r_{L,L}^{L-1}\right) - e^{-\lambda z}A\right) - \widetilde{c}z. \end{aligned}$$

Thus, in a period where there are organizations with layer L as their highest layer (or organizations with $L+1$ layers), given prices, agents choose knowledge so as to maximize their earnings as stated above. The first order conditions from this problems imply that

$$\begin{aligned} e^{-\lambda z_{L,L}^0} &= \frac{\widetilde{c}}{\lambda\left(A - r_{L,L}^0\right)}, \\ e^{-\lambda z_{L,L}^\ell} &= \frac{\widetilde{c}h}{\lambda\left(A - r_{L,L}^\ell\right)} \text{ for } 0 < \ell < L, \\ e^{-\lambda z_{L,L}^L} &= \frac{\widetilde{c}h}{\lambda A}, \end{aligned} \tag{2}$$

or

$$\begin{aligned} z_{L,L}^0 &= -\frac{1}{\lambda}\ln\frac{\widetilde{c}}{\lambda\left(A - r_{L,L}^0\right)}, \\ z_{L,L}^\ell &= -\frac{1}{\lambda}\ln\frac{\widetilde{c}h}{\lambda\left(A - r_{L,L}^\ell\right)} \text{ for } 0 < \ell < L, \\ z_{L,L}^L &= -\frac{1}{\lambda}\ln\frac{\widetilde{c}h}{\lambda A}. \end{aligned} \tag{3}$$

and so earnings in the economy are given by

$$\begin{aligned} w_{L,L}^0 &= A - \frac{\widetilde{c}}{\lambda} - \widetilde{c}z_{L,L}^0 = A - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}}{\lambda\left(A - r_{L,L}^0\right)}\right), \\ w_{L,L}^\ell &= \frac{A - r_{L,L}^{\ell-1}}{h} - \frac{\widetilde{c}}{\lambda} - \widetilde{c}z_{L,L}^\ell = \frac{A - r_{L,L}^{\ell-1}}{h} - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}h}{\lambda\left(A - r_{L,L}^\ell\right)}\right) \text{ for } 0 < \ell < L, \\ w_{L,L}^L &= \frac{A - r_{L,L}^{L-1}}{h} - \frac{\widetilde{c}}{\lambda} - \widetilde{c}z_{L,L}^L = \frac{A - r_{L,L}^{L-1}}{h} - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}h}{\lambda A}\right). \end{aligned} \tag{4}$$

Note that the knowledge acquired is increasing in A and decreasing in $\widetilde{c}$, h (for problem solvers) and the price obtained for selling problems. The intuition for the effect of A and $\widetilde{c}$ is immediate. For h, remember that a higher h implies a worse communication technology. So a higher h implies that problem solvers can buy fewer problems and so they can span their knowledge over less problems. Knowledge becomes less useful. As the price at which agents sell problems increases, agents have an incentive to sell their problems instead of learning more to squeeze all their value, which creates incentives to learn less.

At any point in time t an economy with technology A and organizations with $L+1$ layers is in equilibrium if the knowledge levels of agents solve Equations (3) and

$$w_{L,t}^{\ell} = w_{L,t}^{\ell+1} \equiv w_t \text{ for all } \ell = 0, ..., L-1. \tag{5}$$

This condition is equivalent to an equilibrium condition requiring that the supply and demand of problems at every layer equalize at the equilibrium prizes $\left\{r_{L,t}^{\ell}\right\}_{\ell=0}^{L-1}$. The reason is that when wages are equalized, agents are indifferent as to their role in the organization, and thus they are willing to supply and demand positive amounts of the problems in all layers. Equilibrium in the markets for problems given L then implies that there are a number

$$n_{L,t}^{\ell} = h\left(1 - F(Z_{L,t}^{\ell-1})\right) n_{L,t}^{0} = he^{-\lambda Z_{L,t}^{\ell-1}} n_{L,t}^{0}$$

of agents working in layer ℓ. Since the economy is populated by a unit mass of agents, the number of workers is given by

$$n_{L,t}^{0} = \frac{1}{1 + h\sum_{\ell=1}^{L}\left(1 - F(Z_{L,t}^{\ell-1})\right)}.$$

So given t, A, and L an equilibrium for one generation of agents is a collection of L prices $\left\{r_{L,t}^{\ell}\right\}_{\ell=0}^{L-1}$ and $L+1$ knowledge levels $\left\{z_{L,t}^{\ell}\right\}_{\ell=0}^{L}$ that solve the $2L+1$ equations in (3) and (5). Before we move on to characterize the solution to this system of equations consider the solutions of the system as $L \to \infty$. In this case, since there is no final layer, the system has a very simple solution. Guess that $r_{\infty}^{\ell} = r_{\infty}$ for all ℓ.

Then, the first order conditions in (3) imply that

$$z_{\infty}^{0} = -\frac{1}{\lambda}\ln\frac{\widetilde{c}}{\lambda\left(A - r_{\infty}\right)},$$
$$z_{\infty}^{\ell} = -\frac{1}{\lambda}\ln\frac{\widetilde{c}h}{\lambda\left(A - r_{\infty}\right)} \text{ for all } \ell > 0.$$

Note that, since $h < 1$, $z_{\infty}^{0} < z_{\infty}^{\ell}$ for $\ell > 1$. That is, in the limit as the number of layers goes to infinity workers learn less than all other agents in the economy. Wages are then given by,

$$w_{\infty}^{0} = A - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}}{\lambda\left(A - r_{\infty}\right)}\right),$$
$$w_{\infty}^{\ell} = \frac{A-r}{h} - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}h}{\lambda\left(A - r_{\infty}\right)}\right) \text{ for all } \ell > 0.$$

Since r_{∞} is not a function of ℓ, earnings of problem solvers are identical as is the amount of knowledge they learn. This verifies our guess if we can find an r such that $w_{\infty} \equiv w_{\infty}^{0} = w_{\infty}^{\ell}$. It is easy to see that

$$r_{\infty} = A\left(1-h\right) + \frac{\widetilde{c}h}{\lambda}\ln h$$

solves this equation. Hence, earnings as $L \to \infty$ are given by

$$w_{\infty} = A - \frac{\widetilde{c}}{\lambda}\left(1 + \ln\left(\frac{A\lambda h}{\widetilde{c}} - h\ln h\right)\right),$$

and the knowledge acquired by agents is given by

$$z_{\infty}^{0} = \frac{1}{\lambda}\ln h\left(\frac{A\lambda}{\widetilde{c}} - \ln h\right),$$
$$z_{\infty}^{\ell} = \frac{1}{\lambda}\ln\left(\frac{A\lambda}{\widetilde{c}} - \ln h\right) \text{ for all } \ell > 0.$$

The case of $L \to \infty$ is helpful since it is evident that the economy will converge to it as the number of layers increases. Furthermore, when $L \to \infty$ no valuable problems are thrown away. Thus w_{∞} bounds the level of earnings agents can achieve with technology A. We now turn to the characterization of an equilibrium given t, A and L finite. The next proposition shows that an equilibrium given t, A and L finite exists, is unique, $r_{L,t}^{\ell}$ is decreasing in ℓ, and $z_{L,t}^{\ell}$ is increasing in ℓ. The logic is straightforward.

Start with layer L. These problem solvers cannot resell the problems to a higher layer. Hence, relative to agents one layer below, who can resell their problems, agent in L are willing to pay less for them than agents in layer $L-1$ are willing to pay for the problems they buy. Similarly, agents in layer $L-1$ are willing to pay less for the problems they buy than agents in layer $L-2$ as they can sell them for a low price to agent in layer L. This logic goes through for all layers. The more layers on top of an agent the more valuable the problem, as it can potentially be sold to all the layers above, up to L. Now consider the amount of knowledge acquired by agents. Agents in layer L cannot sell their problems and so they have an incentive to learn as much as possible to extract as much value as possible from each problem. In contrast, agents in layer $L-1$ are less willing to learn as they can sell their problems to agents in layer L. Agents in layer $L-2$ get a higher price for their unsolved problems so their incentives to learn are smaller than the agents above them. Again, this logic applies to all layers in the hierarchy, including layer 0 where the fall in knowledge is even larger since worker can span their knowledge over only one problem instead of $1/h$ of them (since they use their time to produce). Of course, as $L \to \infty$ this logic does not apply and all prices and knowledge levels of problem solvers are constant, since there is no final layer in which prices are equal to zero.

To prove the next proposition we will use the following parameter restriction which is necessary and sufficient for $z_{L,t}^{\ell} > 0$ for all ℓ and L.

Condition 1 *$A \geq 1$, $h < 1$ and $A, \lambda, \widetilde{c}$ and h satisfy*

$$\frac{A\lambda}{\widetilde{c}} > \frac{1}{h} + \ln h.$$

Proposition 2 *Under Condition 1, for any time t, A, and L finite, there exists a unique equilibrium determined by a set of prices $\left\{r_{L,t}^{\ell}\right\}_{\ell=0}^{L-1}$ and a set of knowledge levels $\left\{z_{L,t}^{\ell}\right\}_{\ell=0}^{L}$ such that $r_{L,t}^{\ell} > 0$ is strictly decreasing in ℓ and $z_{L,t}^{\ell} > 0$ is strictly increasing in ℓ.*

Proof. Use (3) to obtain the knowledge of each agent as a function of the price the agent receives for a problem passed. Letting $\alpha \equiv \frac{\widetilde{c}h}{\lambda}$ and $\beta \equiv A - \frac{\widetilde{c}h}{\lambda}$ and using (4) we obtain the following recursion for the set of prices:

$$r_{L,t}^{L-1} = \beta - hw_{L,t}^{L} + \alpha \ln \frac{\alpha}{A}$$
$$r_{L,t}^{\ell-1} = \beta - hw_{L,t}^{\ell} + \alpha \ln \frac{\alpha}{\left(A - r_{L,t}^{\ell}\right)} \text{ for } 0 < \ell < L.$$

Imposing (5) for $\ell = 1, ..., L-1$ we obtain that

$$r_{L,t}^{L-1} = \beta - hw_t + \alpha \ln \frac{\alpha}{A} \quad (6)$$
$$r_{L,t}^{\ell-1} = \beta - hw_t + \alpha \ln \frac{\alpha}{\left(A - r_{L,L}^{\ell}\right)} \text{ for } 0 < \ell < L.$$

For a given w_t there exists at most one $r_{L,t}^{0} > 0$ such that the whole system holds. Specifically, note that given w_t we can determine $r_{L,t}^{L-1}$. So choose some $w_t > 0$ such that the resulting price $r_{L,t}^{L-1} > 0$ (and denote by $r_L^{\ell}(w_t)$ the solution of the system above given w_t). It is easy to see that, since $r_{L,t}^{L-1} > 0$, $r_L^{L-2}(w_t) > r_L^{L-1}(w_t)$. Repeating this argument we can conclude that $\left\{r_L^{\ell}(w_t)\right\}_{\ell=0}^{L-1}$ is decreasing in ℓ. It is also immediate from (3) that the higher the price the lower the corresponding knowledge level, so $\left\{z_L^{\ell}(w_t)\right\}_{\ell=0}^{L}$ is increasing in ℓ (note that for $z_{L,L}^{0}$ there is an extra effect coming from the fact that workers cannot span their knowledge over many problems, a missing h in (3)). Condition 1 guarantees that the resulting values $\left\{z_L^{\ell}(w_t)\right\}_{\ell=0}^{L}$ are positive, as $r_L^{\ell}(w_t) < r_{\infty}$ since when $L \to \infty$ all prices are positive (as opposed to zero in layer L) and, as can be readily observed in the system of equations above, prices in layer $\ell - 1$ are increasing in prices in layer ℓ. Note also that as the price at which agents in layer L can sell problems is equal to zero, the prices for all other layers are strictly positive.

Note that $r_L^0(w_t)$ is decreasing in w_t as

$$\frac{dr_L^0(w_t)}{dw_t} = -h + \frac{\alpha}{A - r_L^1(w_t)} \frac{dr_L^1(w_t)}{dw_t}$$

and

$$\frac{dr_L^{L-1}(w_t)}{dw_t} = -h,$$

so

$$\frac{dr_L^0(w_t)}{dw_t} = -h\left(1 + \sum_{\ell=1}^{L-1} \prod_{k=1}^{\ell} \frac{\alpha}{A - r_L^k(w_t)}\right) < 0,$$

and we can therefore invert it to obtain $w_t^s\left(r_{L,t}^0\right)$ which is also a continuous and strictly decreasing function.

Now consider the equation determining the wages of production workers and define

$$w_t^p\left(r_{L,t}^0\right) = A - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}}{\lambda\left(A - r_{L,t}^0\right)}\right) \tag{7}$$

which is a continuous and strictly increasing in $r_{L,t}^0$.

The last equilibrium condition is given by (5) for $\ell = 0$, and so $w_t^s\left(r_{L,t}^0\right) = w_t^p\left(r_{L,t}^0\right)$ for the equilibrium $r_{L,t}^0$. Since w_t^s is strictly increasing and w_t^p is strictly decreasing, if a crossing exists it is unique. But note that at $r_{L,t}^0 = A - \widetilde{c}/\lambda$, $w_t^p\left(A - \widetilde{c}/\lambda\right) = A - \widetilde{c}/\lambda$ and

$$\begin{aligned} w_t^s\left(A\right) &= \frac{\widetilde{c}}{\lambda h} - \frac{\widetilde{c}}{\lambda} + \frac{\widetilde{c}}{\lambda}\ln\frac{\widetilde{c}h}{\lambda\left(A - r_{L,L}^{\ell}\right)} \\ &< \frac{\widetilde{c}}{\lambda}\left(\frac{1}{h} + \ln h - 1\right) \\ &< A - \widetilde{c}/\lambda \end{aligned}$$

by Condition 1 and $r_{L,t}^0 > r_{L,t}^1$. Hence, $w_t^p\left(A - \widetilde{c}/\lambda\right) > w_t^s\left(A - \widetilde{c}/\lambda\right)$.

Now let $r_{L,t}^0 = 0$. Then

$$w_t^p(0) = A - \frac{\widetilde{c}}{\lambda} + \frac{\widetilde{c}}{\lambda}\ln\frac{\widetilde{c}}{\lambda A}$$

and note that

$$\begin{aligned} w_t^s\left(0\right) &= \frac{A}{h} - \frac{\widetilde{c}}{\lambda} + \frac{\widetilde{c}}{\lambda}\ln\frac{\widetilde{c}h}{\lambda\left(A - r_{L,L}^1\right)} \\ &> \frac{A}{h} - \frac{\widetilde{c}}{\lambda} + \frac{\widetilde{c}}{\lambda}\ln\frac{\widetilde{c}}{\lambda A} \end{aligned}$$

since $h < 1$ and $r_{L,L}^1 \geq 0$. Thus, $w_t^p\left(0\right) < w_t^s\left(0\right)$. The Intermediate Value Theorem then guarantees that there exists a unique value $r_{L,t}^0$ such that $w_t^s\left(r_{L,t}^0\right) = w_t^p\left(r_{L,t}^0\right)$ and so a unique equilibrium exists. ■

We now turn to the properties of this economy as we change the highest layer L. Note that for now, without radical innovations, changes in L happen as time evolves and so studying the properties of our economy as we change the number of layers is equivalent to studying the properties of our economy as time evolves. This equivalence will change in the next section once we introduce radical innovations as we will have organizations evolving for different technologies across time. The next proposition shows that as the number of layers increases so do wages (or output per capita if knowledge cost are considered forgone output). Furthermore since wages are bounded by w_∞, there are eventual decreasing returns in the number of organizational layers. This is just the result of higher layers dealing with less problems as they are more rare. So adding an extra layer contributes to output per capita (since more problems are solved) but it contributes less the higher the layer since there are fewer and fewer problems that require such specialized knowledge.

The proposition also shows that as time evolves and the number of layers increases, $r^\ell_{L,t}$ increases and $z^\ell_{L,t}$ decreases for all ℓ. The first result is a direct consequence of the logic used in the previous proposition. As time elapses and the number of layers increases the number of layers above a given ℓ increases, which implies that $r^\ell_{L,t}$ increases, since the problems can be resold further if not solved. In turn, higher prices in turn imply less knowledge acquisition as the opportunity to resell problems is a substitute for solving them.

Proposition 3 *Under Condition 1, for any technology A, as time t and the number of layers L increase, w_t increases and $\lim_{t\to\infty} w_t = w_\infty$. Furthermore, as time t and the number of layers L increase, prices $r^\ell_{L,t}$ increase for all $\ell = 0, ..., L-1$ and knowledge levels $z^\ell_{L,t}$ decrease for all $\ell = 0, ..., L$. As $t \to \infty$ and $L \to \infty$, $r^\ell_{L,t} \to r_\infty$ for all $\ell = 0, ..., L-1$ and $z^\ell_{L,t} \to z^0_\infty$ all $\ell = 0, ..., L$.*

Proof. Consider the individual incentives of an agent in period t to form layer $L+1$ given that the economy's highest layer is L. Such an agent can use the problems thrown away by the agents in layer L. The wage such an agent in layer $L+1$ would command is given by

$$\frac{A}{h} - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}h}{\lambda A}\right)$$

which is always greater than the equilibrium wage in the economy given by

$$w_t = \frac{A - r^{L-1}_{L,t}}{h} - \frac{\widetilde{c}}{\lambda}\left(1 - \ln\frac{\widetilde{c}h}{\lambda A}\right),$$

since as shown in the previous proposition $r^{L-1}_{L,L} > 0$. Therefore, in the next period such an agent has incentives to enter and form layer $L+1$. Of course, once he enters, agents in layer L will demand a positive price for their problems and so some of the surplus will be distributed to other agents in the economy. However, the economy as a whole will produce more output as the higher price is only a redistribution of wealth between agents. Agents will also re-optimize and choose different levels of knowledge $\left\{z^{\ell}_{L,t}\right\}^{L}_{\ell=0}$ which will increase the surplus, as they have the option to choose the same level of knowledge they chose before. Hence, $w_{t+1} > w_t$ for all t.

This result can be formally proven as follows. Consider $r^0_L(w)$ defined in Proposition 2. As $r^L_{L,t} = 0$ (the last layer throws problems away) but $r^L_{L+1,t+1} > 0$ and since for a given w, by Equations (6), $r^{\ell-1}_{L+1,t+1}$ is increasing in $r^{\ell}_{L+1,t+1}$, we obtain that $r^0_L(w) < r^0_{L+1}(w)$. Now define the function r^p using Equation (7), the the price of problems sold by workers, as

$$r_p(w) \equiv A - \frac{\widetilde{c}}{\lambda} e^{\frac{\lambda}{c}(A-w)-1}.$$

In an equilibrium with L layers we know that $r_p(w) = r^0_L(w)$ and in an equilibrium with $L+1$ layers $r_p(w) = r^0_{L+1}(w)$. Since $r^0_L(w) < r^0_{L+1}(w)$ and $r'_p(w) > 0$ and $r^{0\prime}_L(w) < 0$, this implies that $w_{L+1} > w_L$ and that $r^0_{L+1,t} > r^0_{L,t}$. By (6) this in turn implies that $r^{\ell}_{L+1,t} > r^{\ell}_{L,t}$ for all $\ell < L-1$. Note also that by (3) this implies that $z^{\ell}_{L+1,t} < z^{\ell}_{L,t}$ for all $\ell < L-1$.

Note that as we have shown in Proposition 2, $r^{\ell}_{L,t} < r_{\infty}$ for all ℓ and L finite. Hence, since $\left\{r^{\ell}_{L,t}\right\}^{\infty}_{L=0}$ is a strictly increasing and bounded sequence it has to converge for all ℓ. Since the equilibrium is unique as shown in Proposition 2 the limit is r_{∞}. Hence, as $t \to \theta$, $\left\{r^{\ell}_{L,t}\right\}^{\infty}_{L=0}$ approaches r_{∞} from below. Equations (3) then imply that $\left\{z^{\ell}_{L,t}\right\}^{\infty}_{L=0}$ converges to z^{ℓ}_{∞} from above. ∎

The previous proposition shows that our economy will grow. But it also shows that the level of wages is bounded. Hence, growth in wages (or per capita output) will converge to zero. That is, the economy does not exhibit permanent growth. We now turn to embed this evolution over time of organizations with a given technology A in a growth model in which agents will have a choice to switch to better technologies as they learn. This will yield a long-run growth model that will exhibit permanent growth and where this growth will be driven by the ability of agents to organize.

Before we end this section it is important to make one remark about the evolution of the distribution of gross wages (without subtracting learning costs). Overall wage inequality, as measured by the ratio of the gross wages of the highest level problem solvers to the gross wages of workers, increases over time as a technology is more efficiently organized. To see this note that everyone gains the same net of learning costs, and knowledge levels of workers decrease with the number of layers, while knowledge levels of entrepreneurs at the highest layer are constant. In contrast, the distribution of gross wages among problem solvers becomes less dispersed. The reason is that more layers are added and knowledge levels of intermediate problem solvers converge to z_∞. Thus, as organizations develop over time, inequality between workers and managers increases while inequality within problem solvers decreases.

2.2. Long-run Growth through Radical Innovations

While knowledge has aspects that are fully appropriable, it also involves important externalities. Specifically, as agents strive to improve existing technologies, they add to the store of existing knowledge, which serves to improve the chances that radical innovations will take place. Radical innovations in this view share two characteristics. First, while their development benefits from the growth in the knowledge stock, the problems they involve are different than the ones posed by the previous technologies. As a result, they make previous organizations obsolete (e.g., digital photography largely makes chemical film obsolete). Second, they are not appropriable, as they raise the level of technology available to all agents in the economy. Thus the process of creating breakthroughs and developing radically new ideas involves an externality that benefits everyone.

In the previous section we studied how an economy organizes given a technological level A. In our economy, as organizations grow and become more complex, society learns how to solve a wider set of problems faced when using this technology. This knowledge is fully appropriable and society invests optimally, conditionally on A, on the development of this problem solving knowledge. In this section we link this growth in problem solving knowledge to the general state of technology. We now proceed to introducing these radical, non-appropriable, innovations. Our key assumption is that these innovations take place as a by-product of the knowledge accumulation that results from the process of learning (in a fully priced and optimal way) how to solve the problems presented by the current technology. Essentially, each generation leaves behind a large store of knowledge that helps the new generation improve the existing technology.

At every period t denote the best technology available, that may or may not be in use, by A_t. Every time a new technology is adopted new organizations have to be built to exploit it —once the preexisting technology is obsolete, the organizations that were developed to exploit that technology become obsolete as well. As in the previous section, building organizations takes time as only one layer can be added per period. The best technology may not be in use since agents may prefer to use a technology invented in the past, for which they have built a large organization, instead of starting from scratch on a new technology if it chooses to do so. .

The best technology available depends on the best technology that was available last period and the amount of new knowledge created by society $z_{L,t}^{L}$. We specify the evolution of the best technology available as

$$A_t = A_{t-1} e^{g z_{L,t}^{L}}, \tag{8}$$

where g is a parameter that governs how useful is deeper knowledge about old technologies to invent new ones. Note that the exponential specification will lead to a constant long term growth path, which is the reason we use it. We could also let the best technology available be a function of the total amount of knowledge in society in a given period, $\sum_{i<\ell} z_{L,t}^{i}$, as well as the level of technology in use. All of our conclusions would go trough for that case but at the cost of somewhat more complicated algebra.

The cost of learning new technologies, measured in terms of foregone income, increases with the level of the new technologies —the more productive the technology the higher the cost of spending time learning to solve problems. Thus we specify the learning cost of the new technology as $\widetilde{c} = cA$. Then the cost of learning how to solve problems in an interval of size z is equal to Acz where A is the technology currently in use (not necessarily the best one). All the analysis in the previous section remains unchanged apart from the parameter $\widetilde{c}$ now becoming Ac.[8]

[8] Note that if we do not scale c by the level of technology, as the economy grows the cost of

Switching to the new technology involves making the existing organization, and the knowledge that it acquired, obsolete; that is, agents must start from scratch. As agents live only for one period, their decision to adopt a new technology only depends on the current gains of using that technology versus their opportunity cost of adding a new layer. The wage w_t^R that agents in period t can obtain if they switch to the best available technology A_t (a radical innovation) is given by the wage of workers in an organization with zero layers using this new technology. Namely,

$$\begin{aligned} w_t^R &= \max_z A_t F(z) - A_t cz \\ &= \max_z A_t \left(1 - e^{-\lambda z}\right) - A_t cz \\ &= A_t \left(1 - \frac{c}{\lambda} + \frac{c}{\lambda} \ln \frac{c}{\lambda}\right), \end{aligned} \tag{9}$$

which is proportional to the level of the best available technology A_t.

Every period agents compare their wage when they undertake a radical innovation, w_t^R, with the wage they would command if they develop one more layer using the current technology. Note that, if the current hierarchy has $L+1$ layers, a new technology was put in place in period $t-L-1$. So the technology in use is A_{t-L-1}. Those wages, if the new layer is layer L, are given, as discussed above, by

$$\begin{aligned} w_{L,t} &= A_{t-L-1}\left(1 - \frac{c}{\lambda} + \frac{c}{\lambda} \ln \frac{c}{\lambda\left(1 - \tilde{r}_L^0\right)}\right), \\ &= A_{t-L-1}\left(\frac{1 - \tilde{r}_L^{\ell-1}}{h} - \frac{c}{\lambda} + \frac{c}{\lambda} \ln \frac{ch}{\lambda\left(1 - \tilde{r}_L^{\ell}\right)}\right) \text{ for } 0 < \ell < L, \\ &= A_{t-L-1}\left(\frac{1 - \tilde{r}_L^{L-1}}{h} - \frac{c}{\lambda} + \frac{c}{\lambda} \ln \frac{ch}{\lambda}\right) \end{aligned}$$

where $r_{L,t}^{\ell} = A_{t-L+1}\tilde{r}_L^{\ell-1}$. This normalization is possible since if $\left\{r_{L,t}^{\ell}\right\}_{\ell=0}^{L-1}$ solves Equations (5), then $\tilde{r}_L^{\ell-1}$ solves the same system with technology $A=1$ and the system does not depend on time. So if $w_t^R \geq w_{L,t}$ a radical innovation occurs in period t, otherwise, if $w_t^R < w_{L,t}$, the economy develops layer L. Agents make this choice every period. It may be the case that radical innovations occur every period. This is the case if agents learn enough in an organization of zero layers (when they

acquiring knowledge relative to output will converge to zero. This would introduce an obvious scale effect in the model.

work on their own) to improve their technology sufficiently so that switching to the next technology is better than building the first layer of problem solvers. That is, when

$$A_t\left(1-\frac{c}{\lambda}+\frac{c}{\lambda}\ln\frac{c}{\lambda}\right)\geq A_{t-1}\left(1-\frac{c}{\lambda}+\frac{c}{\lambda}\ln\frac{c}{\lambda\left(1-\tilde{r}_0^0\right)}\right)$$

or if

$$\frac{A_t}{A_{t-1}}=e^{gz_0^0}=e^{\frac{g}{\lambda}\ln\frac{\lambda}{c}}>\frac{1-\frac{c}{\lambda}+\frac{c}{\lambda}\ln\frac{c}{\lambda\left(1-\tilde{r}_0^0\right)}}{1-\frac{c}{\lambda}+\frac{c}{\lambda}\ln\frac{c}{\lambda}}.$$

Clearly, this comparison is independent on the level of technology.[9] Hence, whether a society organizes a technology or not depends only on the four primitive parameters of our economy $\{h,\lambda,c,g\}$.

Suppose $\{h,\lambda,c,g\}$ are such that

$$e^{\frac{g}{\lambda}\ln\frac{\lambda}{c}}<\frac{1-\frac{c}{\lambda}+\frac{c}{\lambda}\ln\frac{c}{\lambda\left(1-\tilde{r}_0^0\right)}}{1-\frac{c}{\lambda}+\frac{c}{\lambda}\ln\frac{c}{\lambda}}. \tag{10}$$

In this case agents will prefer to develop the first layer of an organization rather than switching to a new technology immediately.

The next question that arises concerns the frequency of radical innovations when Condition (10) holds. From Equation (3) we know that

$$z_{L,t}^L=\frac{1}{\lambda}\ln\frac{\lambda}{ch}$$

whenever $L\geq 1$, and

$$z_{0,t}^0=\frac{g}{\lambda}\ln\frac{\lambda}{c}.$$

Under Condition 1 it is easy to verify that $z_{L,t}^L>z_{0,t}^0>0$. Hence

$$\ln A_{t+1}-\ln A_t>\frac{g}{\lambda}\ln\frac{\lambda}{c}>0$$

which implies, since growth in w_t^R is driven by the evolution of technology, that

$$\ln w_{t+1}^R-\ln w_t^R=\frac{g}{\lambda}\ln\frac{\lambda}{c}>0.$$

In contrast, Proposition 3 guarantees that

$$\lim_{L,t\to\infty}w_{L,t}=w_\infty\left(A_{t-L-1}\right)=A_{t-L-1}\left(1-\frac{c}{\lambda}-\frac{c}{\lambda}\ln\left(\frac{\lambda h}{c}-h\ln h\right)\right)$$

[9]Our examples below show that it is indeed possible that the inequality holds in either direction.

and that for any finite L, $w_{L,t} < w_\infty$. Hence, growth in $w_{L,t}$ converges to zero, that is

$$\lim_{L,t\to\infty} (\ln w_{L,t} - \ln w_{L+1,t+1}) = 0.$$

Clearly this implies that there exists a finite τ such that every τ periods there is a radical innovation. Note also that since $A_t/A_{t-\tau}$ is independent of the level of the technology in use, $A_{t-\tau}$, τ is independent of the level of technology and is determined only by the parameters of our economy, $\{h, \lambda, c, g\}$. Hence, given these parameters, our economy grows as described in Proposition 3 for τ periods at which time there is a radical innovation and we start growing as an economy with no organization again. These are recurrent cycles in which a technology is invented, is exploited increasingly more efficiently for τ periods, and then another innovation occurs. During those τ periods, growth first accelerates as the first layers of the organization form and then eventually decreases as the number of layers adds less and less to the economy (which we know from the fact that wages when using this technology are bounded). Together with Proposition 2 the existence and uniqueness of τ implies that a unique dynamic equilibrium exists. We summarize this result in the following proposition.

Proposition 4 *Under Condition 1, there exists a unique dynamic competitive equilibrium that exhibits technological cycles. The length τ of these cycles is constant over time and depends only on the four parameters $\{h, \lambda, c, g\}$.*

We now turn to the calculation of the growth rate in this economy. First note that if innovations never occur, so Condition 10 does not hold, growth will be determined by how much workers learn. Absent organization we know that worker's knowledge rage is given by

$$z_{0,t}^0 = \frac{1}{\lambda} \ln \frac{\lambda}{c}$$

and so

$$\ln w_{t+1}^R - \ln w_t^R = g z_{0,t}^0 = \frac{g}{\lambda} \ln \frac{\lambda}{c}.$$

In this case no organizations form and so the growth rate of wages and output per capita is also trivially given by $\frac{g}{\lambda} \ln \frac{\lambda}{c}$.

Consider now the more interesting case in which Condition 10 does hold. In this case, we know that organizations will form and will grow and add layers each period until the highest layer is τ and a new technology is put in place. First note that the long term average growth rate of wages (or output per capita) will be determined by the growth rate of w_t^R since innovations keep happening every τ periods. Hence, the average growth rate of wages, $\tilde{g}$, over one technological cycle (which is the same as over many cycles) is given by

$$\begin{aligned}\tilde{g} &= \frac{\ln w_{t+\tau} - \ln w_t}{\tau} = gz_{0,t}^0 + (\tau - 1)\, gz_{L,t}^L \\ &= \frac{g}{\lambda} \ln \frac{\lambda}{c} + (\tau - 1) \frac{g}{\lambda} \ln \frac{\lambda}{hc} \\ &= \frac{g}{\lambda} \left(\ln \frac{\lambda}{c} + \frac{\tau - 1}{\tau} \ln \frac{1}{h} \right).\end{aligned} \tag{11}$$

As the technological cycles become longer, an increase in τ, the growth rate increases since $1/h > 1$. Note also that g has a direct positive effect on the growth rate as the new acquired knowledge becomes more important for inventing new technologies and therefore for growth. Of course, g will also affect τ and therefore the growth rate indirectly. We are particularly interested in the effect of communication technology on the average growth rate $\tilde{g}$. The direct effect is immediate, better information technology increases the knowledge acquired by the highest layer of problem solvers as they can span their knowledge over more problems. Hence, higher $1/h$ has a positive direct effect on $\tilde{g}$. But better information technology also affects the length of the technology cycle τ. Similarly, a decrease in c increases the amount of knowledge learned by agents and therefore the average long-run growth rate. The next proposition shows that the direct effects discussed here dominate (or go in the same direction of) the indirect effects through τ.

Proposition 5 *The average long-run growth rate $\tilde{g}$ increases with g and $1/h$ and decreases with c. Furthermore an increase in $1/h$ increases the set of parameter values for which the first layer of organization is created.*

Proof. We start with the proof that g and $\tilde{g}$ increases in g. Start with case in which Condition 10 does not hold before (and therefore after) the increase in g. Then there is no organization, $\tau = 1$, and $\tilde{g}$ increases with g by (11). Now suppose Condition 10 holds before but not after the increase in g in period t, then $\tau_{t-1} > \tau_{t+1} = 1$ since in period $t - 1$ the first layer formed, but in period $t + 1$ it did not. So the number of layers went down. Note however that given the level of technology in use earnings in the hierarchy are independent of g. Hence, in order for the number of layers to go down $\ln w_{t+1}^R - \ln w_t^R$ must have increased. Hence, $\tilde{g}$ increases. The same logic applies for increases in g when Condition 10 holds before and after. As $\ln w_{t+1}^R - \ln w_t^R$ increases since $\frac{g}{\lambda} \ln \frac{\lambda}{c}$ and $\frac{g}{\lambda} \ln \frac{\lambda}{hc}$ are increasing in g and wages in the hierarchy are independent of g, the crossing between w_t^R and $w_{L,t}$ happens for a smaller τ. But growth is given by the average growth of w_t^R which increased so $\tilde{g}$ increases with g. Note that the negative effect of τ in (11) can never dominate since τ is smaller only if $\tilde{g}$ increased in the first place.

Now consider the effects of changes in $1/h$. First note that

$$r_{L,L}^0 = A_{t-L-1} \left(1 - h \frac{w_{L,t}}{A_{t-L-1}} - \frac{hc}{\lambda} - c z_{L,t}^0 \right)$$

and so given $w_{L,t}$ and A_{t-L-1}, $r_{L,L}^0$ is increasing in $1/h$. Hence, $r_L^0(w)$ defined in Proposition 2 increases with $1/h$ (note that we do not need to take into account the effect of h on $z_{L,t}^0$ by the Envelope Theorem) and since $r_p(w)$ defined in Proposition 3 is not a function of h, an increase in $1/h$ increases $w_{L,t}$ for hierarchies of any number of layers $L+1$. The increase in $\tilde{g}$ with $1/h$ follows since $\ln w_{t+1}^R - \ln w_t^R = \frac{g}{\lambda} \ln \frac{\lambda}{c}$ which is independent of h or $\ln w_{t+1}^R - \ln w_t^R = \frac{g}{\lambda} \ln \frac{\lambda}{hc}$ which is increasing in $1/h$. Hence, $\tilde{g}$ increases with $1/h$.

Note that as $\tilde{r}_0^0$ (the price that workers get for their problems once we normalized by the level of technology) increases with $1/h$, the left-hand-side of Condition 10 increases and so the condition is more likely to hold. Hence an increase in $1/h$ increases the set of cases in which the first layer is organized.

A similar argument to the one we used for the case of $1/h$ hold for the case of c. The equation above shows that $r_L^0(w)$ is decreasing in c. Hence, w_t is decreasing in c. As $\ln w_{t+1}^R - \ln w_t^R$ is also decreasing in c, $\tilde{g}$ is decreasing in c. ∎

The proposition above shows that an improvement in communication technology increases the growth rate and the set of parameters for which organizations form. However, it does not show that a larger $1/h$ will lead to more layers in hierarchies (a higher τ). It is more complicated to show this for the second layer than for the first one because for the second layer (or any other layer except the first one) the growth rate of w_t^R also increases with $1/h$. Which effect dominates depends on parameter values through equilibrium prices of problems. In practice, however, one can verify numerically that for a large set of parameter values the first layers of organization imply much larger increases in wages than the following ones, and so the number of layers in a hierarchy increases with $1/h$. Note that this has to be the case for a low enough g since in that case the effect of a deeper more efficient hierarchical structure has to dominate the effect of faster potential radical innovation (which is proportional to g).

Figure 1 illustrates the equilibrium in our economy for different values of h. The straight line (or almost straight line) depicts the natural logarithm of the wage if the new alternative technology was used, $\ln w_t^R$ (a radical innovation). The lighter line with concave segments (the technology cycles) depicts the natural logarithm of the equilibrium wage, $\ln w_t$.

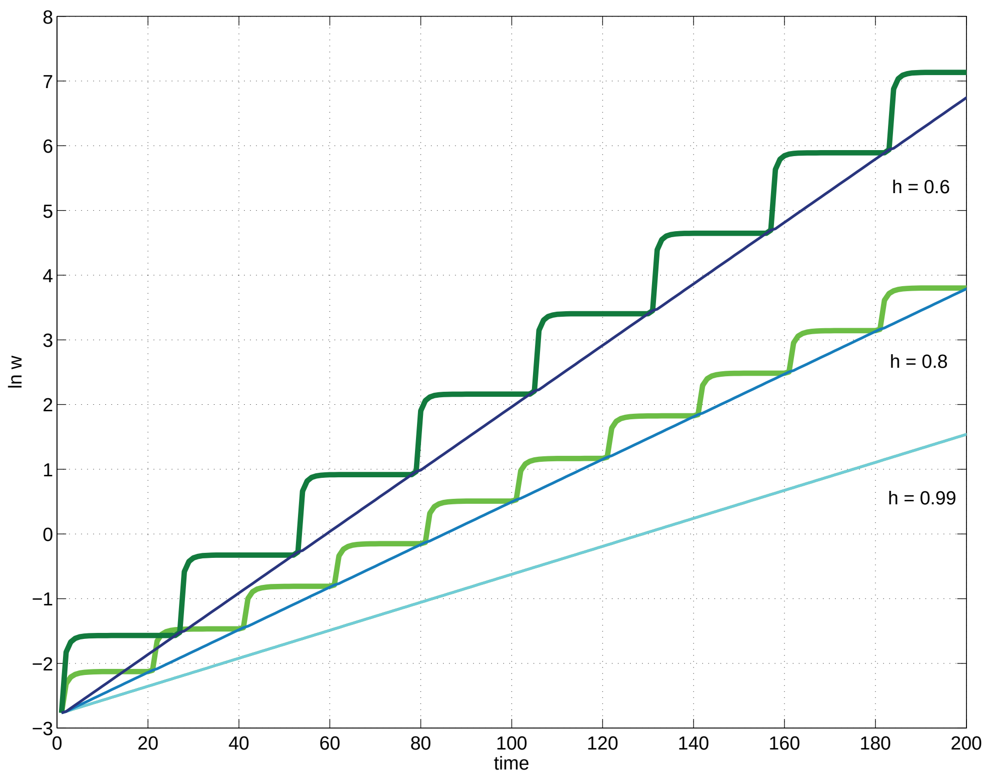

Figure 1: The effect of improvements in comunication technology

The figure is calculated for $\lambda = 1.5$, $c = 1$, $g = 0.08$ and $A_0 = 1$. However we would obtain a qualitatively similar picture from a wide range of parameters satisfying Condition 1. When $h = .99$ and so communication technology is as bad as possible (given that $h < 1$), there is no organization. Every period the economy changes to a new technology, so $\tau = 1$. This case also exhibits the lowest growth rate as implied by Proposition 5. As we improve communication technology by lowering h to 0.8, the growth rate in the economy increases. Forming organizations is now optimal for agents which gives them incentives to acquire more knowledge and therefore the economy grows faster. In this case organizations grow up to $\tau = 10$ layers. At that point agents switch to a new technology. If we improve communication technology further to $h = 0.6$, the growth rate increases and the technology cycle expands. Hierarchies grow to 26 layers before agents decide to switch to a new technology.

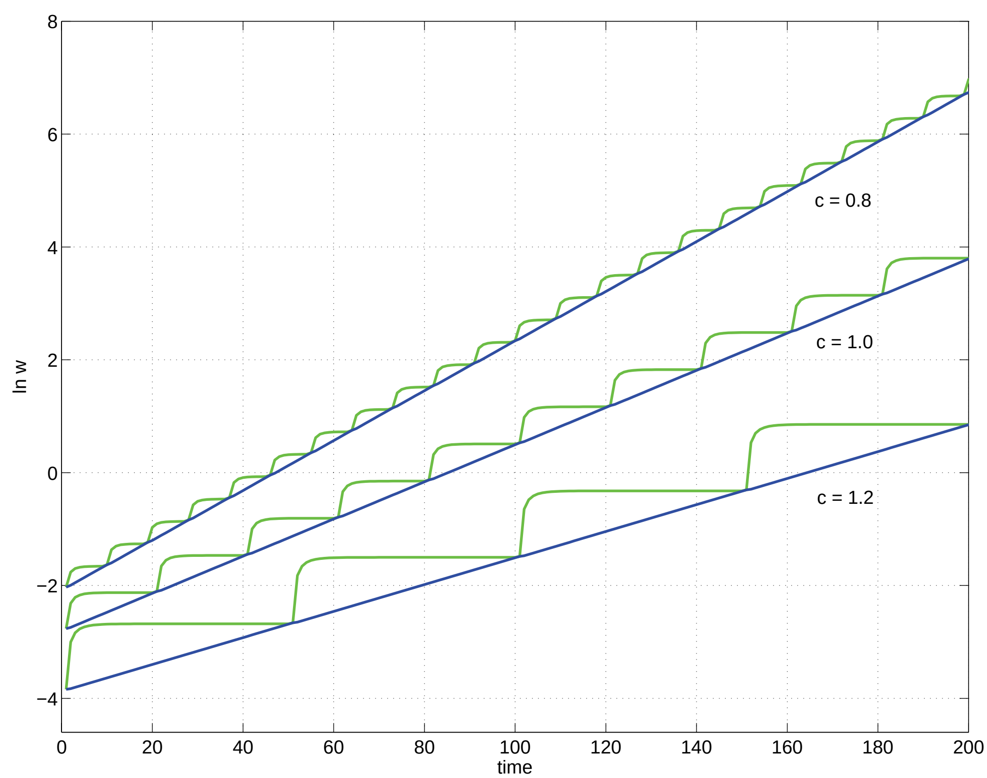

Figure 2: The effect of the cost of acquiring knowledge

Figure 2 present three exercises for different values of costs of knowledge acquisition c. All cases are computed for the case of $h = 0.8$. Decreases in the cost of acquiring information lead to an increase in the average growth rate as proven in Proposition 5, as agents learn more. Decreases in c also lead to a decrease in the length of a technology cycle, τ. On one hand the increase in the growth rate as c declines leads to smaller cycles. On the other hand decreases in c imply agents learn more and so benefit more from the hierarchical structure which leads to larger hierarchies and longer technology cycles.

Figure 2 illustrates that in general the first effect dominates and so a decline in c leads to shorter technology cycles. Note also that decreases in c lead to a positive level effect. As we lower c, workers producing on their own learn more ($z^0_{0,t}$ increases) and their wages increase (even though learning costs increase), as can be verified using Equation (9) and Condition 1.

We still need to discuss the effect of the other two parameters in the model: g and λ. The effect of g, the rate at which new knowledge is transformed into innovations, is straightforward from Proposition 5. Higher g increases the long-run average growth rate and makes technology cycles smaller since it does not affect the gains from hierarchical organization. The effect of λ is much more complicated and in many cases varies depending on the value of other parameter values, in particular c. Note that a lower λ implies a distribution of problems with more mass in the upper tail. Hence, we should expect a decline in λ to generate larger technology cycles as hierarchical organization becomes more important to solve a large range of problems. However, the effect of λ on the growth rate is ambiguous as is immediate from inspecting Equation (11). In particular, it depends on the derivative of $z^{\ell}_{L,t}$ with respect to λ. A sufficient condition for λ to increase the average long-run growth rate is $\ln \lambda / c < 1$. This condition is satisfied in Figure 3 where we show the equilibrium for three values of λ, $c = 1$ and $h = 0.8$. Clearly, as we make λ larger the effect on growth rates becomes smaller (and for even larger λ it is negative).

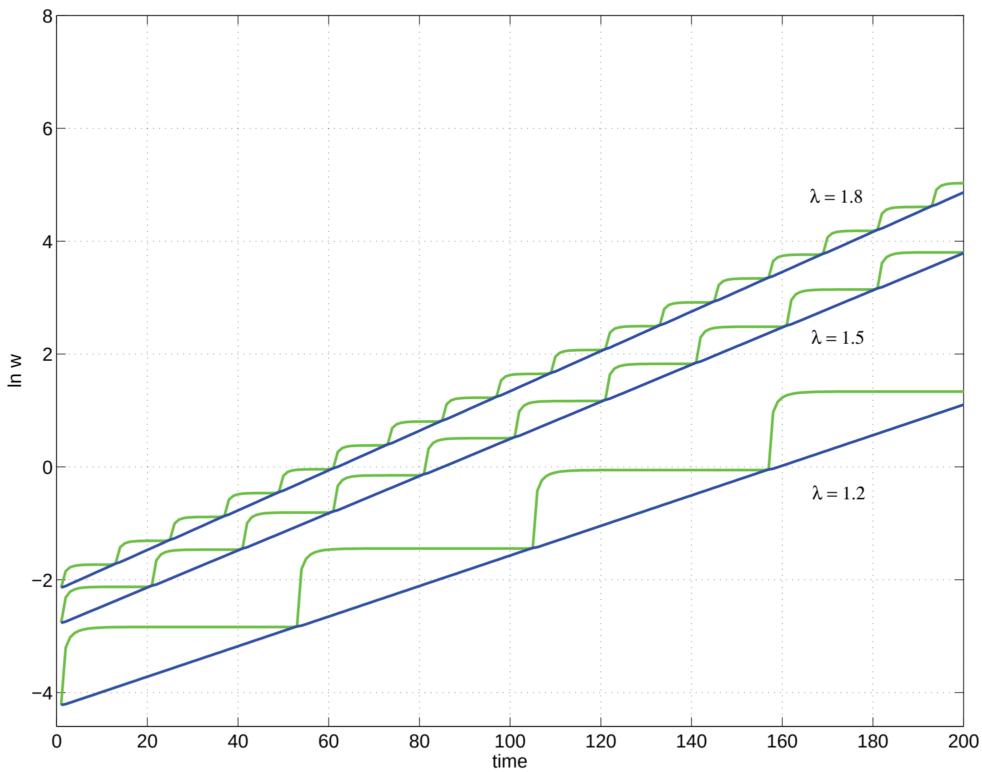

Figure 3: The effect of λ

2.3. The Price of Problems and the Formation of Markets of Solutions

The prices of problems are the main equilibrium prices in this economy. All other prices, like wages, can be constructed (as we did above) using them. If organizations are formed within firms we will observe only wages and not this prices directly. In contrast, if organizations are formed through consultant or referral markets we will observe them directly. The price of problems indicates the level of efficiency and sophistication of the available organization and so, as shown in Proposition 3, given the layer in which an agent works, they increase as organizations develop to exploit a technology. So these prices indicate the level of organizational capital in the economy.

The markets that determine these prices develop progressively as organizations add layers. That is, we are assuming that there is a coordination failure so agents do not sell expertise that is not demanded in the market. An example of this are all the specialized internet companies that have formed as the internet is used more. Some of these companies sell specialized knowledge that was not demanded in the early 90's when internet use became widespread. What prevented this markets for specialized knowledge from forming immediately as agents foresee demand for their services? In this paper we assumed that there is a coordination failure that requires these markets to form sequentially but do not explore these fundamental frictions directly.

Of course, one important difference between the growth experiences of countries could be the completeness and efficiency of these markets. Lower efficiency will, in general, have a similar effect to worse communication technology. More extremely, the inability to create these markets has the potential to eliminate growth as agents do not acquire new knowledge. In this respect, adverse selection is a particular source of concern. What is being traded here, knowledge, is unobservable. What is to prevent an agent from claiming to have knowledge about solving a certain type of problems which in fact he cannot solve? For example, how do we separate a chaman from a doctor? Or, in a referral formulation like the one we presented here, how do we know the level of difficulty of the problem being passed if the quality of the agent who passes it is not observed? Certification mechanisms in developed countries ensure that an agent who claims to be an expert at solving a certain type of problem is indeed such an expert. In their absence, the market mechanisms required to ensure the right match between problems and expertise may be impossible to sustain, and thus the returns to knowledge accumulation may be eliminated. So the new knowledge that enters Equation (8) would be zero and technology would never experience a radical innovation. Growth would collapse to zero.

In this respect contracting institutions in the sense of North (1981) play a critical role in the emergence of organization, specialization, and economic development. Certification mechanism, legal systems, the rule of law, etc. are important to the extent that they are required for the relevant markets for expertise to exists. These markets are the ones that drive the improvements in the efficiency of production and incentivate the acquisition of knowledge that leads to radical technology innovations. Beyond agreeing with and providing a rationale for the common finding about the important role of the rule of law, and more generally institutions, in promoting growth (see e.g. Barro, 1996), our approach suggests certain specific avenues for further research and for policy interventions. Namely, the key role of institutions that certify and enforce the claims of experts to their expertise through school systems, professional organizations, etc. and protect agents from non-experts and charlatans. [10]

3. TECHNOLOGY SHOCKS

It is easy to introduce technology shocks in the model presented in the previous two sections. There are several natural ways of introducing these shocks. First, we can introduce shocks to the best new alternative technology. In this case the value of radical innovations will fluctuate as a function of the shocks. An alternative is to add these shocks to the technology that is currently in use. Then, wages will fluctuate with shocks but the value of the alternative technology will not. Of course, we can also introduce both forms of shocks. Any of these forms of introducing randomness in technology will affect the length of the technology cycle and, therefore, the number of layers that are eventually developed to exploit these technologies. The effect of shocks on growth rates is more subtle and depends on the way we introduce the shocks.

Consider a shock $\varepsilon_t^1 \sim N(0, v)$. Throughout this section we will only consider shocks of mean zero since we do not want to introduce positive or negative level effects on technology that are not generated by the model. Consider introducing shocks in the evolution of the best alternative technology for a radical innovation. Specifically we add shocks to the evolution of the technology in Equation (9) and let

$$A_t = E_{t-1} A_{t-1} e^{g z_{L,t}^L} + \varepsilon_t^1$$

where E_t denotes the expectation conditional on the information in $t-1$.[11] We add the expectation in order to make the shock transitory. In this way if a shock

[10] Much of the theory and evidence in the past focuses on 'property right' type mechanisms that protect assets from expropriation by the government and elites, rather than 'contractual mechanisms' (Acemoglu and Robinson, 2005b). Our framework suggests that the past importance of 'property right' institutions may be less relevant in a future where natural resources are less important. See also Acemoglu and Robinson (2005a) for a review of the evidence.

[11] Note that the way in which we introduce ε_t^1 in principle does not guarantee that A_t is positive. In the simulations below we use an initial technology $A_0 = 1$ and a variance v where this is not a problem Alternatively we could truncate the normal or use a multiplicative lognormal shock with mean one. What is important for us is that, on average, the shock does not change the level of technology. None of our conclusions depends on which of these three options we use.

has permanent effects on wages it will be because of the structure and forces in the model and not because of the way we introduce shocks. Note that since $E_t A_t = E_{t-1} A_{t-1} e^{g z_{L,t}^L}$ we know that when a hierarchy has developed for L layers

$$\begin{aligned} A_{t+L+1} &= A_t e^{g z_{0,t}^0 + L g z_{L,t}^L} + \varepsilon_t^1 \\ &= A_t e^{\frac{g}{\lambda}\left(\ln \frac{\lambda}{c} + \frac{L}{L-1} \ln \frac{1}{h}\right)} + \varepsilon_t^1. \end{aligned} \tag{12}$$

So these shocks will only change the average long-run growth rate through their effect on the length of the technology cycle τ. However, as the shocks are transitory they will not have an effect on the average length of the cycle and therefore will not have an effect on the long-run growth path of the economy. At any given time, the shocks will have level effects on the best alternative technology. A good shock will trigger a radical innovation earlier than average. A sequence of negative shocks will delay radical innovations.

There are two interesting effects that result from introducing shocks in this fashion. First, even though this shocks are transitory they have persistent effects. This is illustrated in Figure 4 where we plot the case of an economy with the same parameters used in Figure 1 and $h = 0.8$. We first generate a sequence of 200 normal random numbers with $v = 0.25$ and use it in all numerical exercises below. In each exercise we then multiply these number by the level of the technology in use so that the volatility does not become smaller as the economy grows. Note from the figure that the length of the product varies with the realization of the shocks and that, as expected, $\ln w_t^R$ is now random. To illustrate how the model creates persistence, we take the realization of shocks and increase the size of the shock in period 100 from its original value of .26 to $\varepsilon_{100}^1 = 3$. We leave the rest of the shocks unaltered. As can be observed in Figure 4 the effect of this large shock is to increase the level of $\ln w_t$ permanently. The reason is that the good shock triggers faster radical innovation. The new technology then becomes the technology in use and is exploited by expanding the number of layers in the hierarchy. Even though the good shock lasted only one period, the fact that a new technology was developed and exploited through the hierarchy implies that it has permanent effects on the level of output of the economy. Hence, the model we have introduced is a mechanism to increase the persistence of transitory shocks. The search for suitable persistence mechanism of shocks has been one of the main concerns of the business cycles literature (as discussed in Kydland and Prescott, 1982).

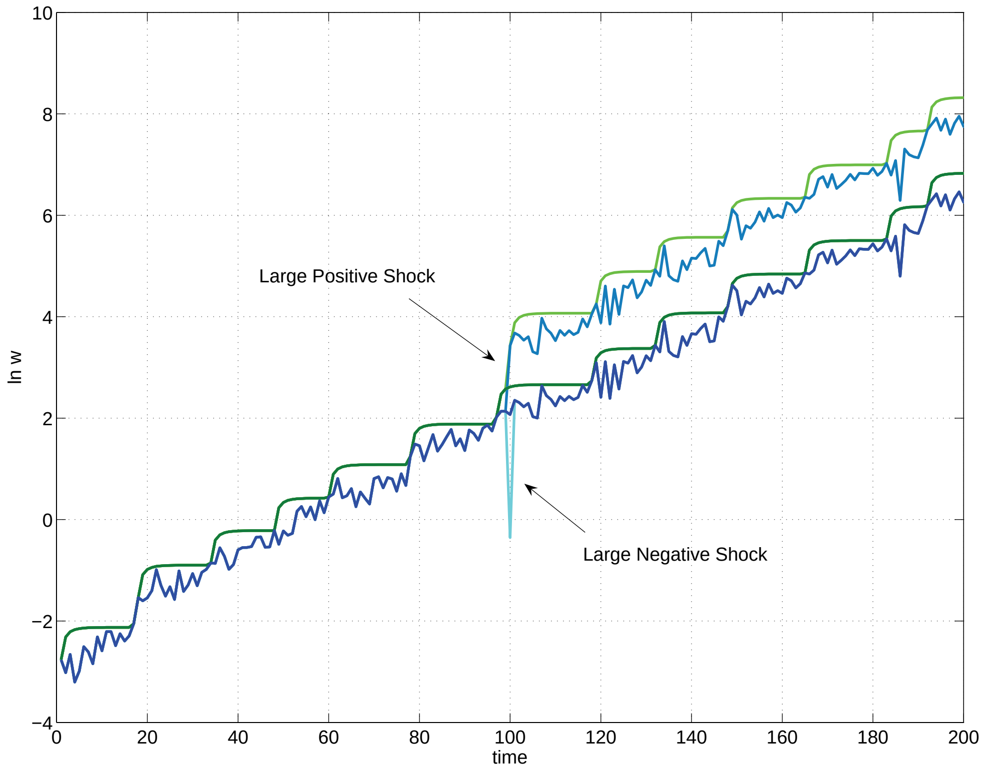

Figure 4: The effect of transitory shocks to the technology for radical innovations

As Figure 4 illustrates, the model generates persistence only of positive but not of negative shocks. The reason is that bad technologies will never lead to a radical innovation, the economy will expand current organizations to exploit the current technology further, rather than use a bad new technology. In this sense, society can extract an option value from the radical innovation process. It can use new great ideas but it can discard bad ones by exploiting the current technology further. Higher volatility, v, will therefore increase the level of output as society takes advantage of the good shocks and discards the bad ones. The average long run growth rate between innovations will not change, but the frequency and size of radical innovation will, which will lead to more and larger jumps in the level of wages.

An alternative way of introducing shocks is to let them affect the technology currently being developed through the organizations. That is, if a new technology was introduced in period t and it is still being developed through taller organizations, denote the technology in use by A_t^O where $A_t^O = A_t$ in this example. Let $\varepsilon_t^2 \sim N(0, v)$ be a second productivity shock which is distributed identically to the previous shock, ε_t^1. We want this shock to affect the technology currently in use so we let [12]

$$A_{t+1}^O = A_t^O + \varepsilon_t^2.$$

So A_t^O is a random walk[13]. Then, if the technology invented at time t is still in use in period $t + L + 1$ wages are given by

$$w_{L,t+L+1} = A_{t+L+1}^O \left(1 - \frac{c}{\lambda} + \frac{c}{\lambda} \ln \frac{c}{\lambda\left(1 - \tilde{r}_L^0\right)}\right).$$

Figure 5 illustrates the results when we introduce both types of shocks with $v_t = 0.15 A_t^O$ and the same parameters as in Figure 4. We present the equilibrium allocation for two different values of h. Clearly, the effect of increases in $1/h$ is, as before, to increase the average long-run growth rate in the economy and increase the maximum number of layers and the length of the average technology cycle. The average long-run growth rate goes from 2.8% to 3.8% as we decrease h from 0.8 to 0.6. Otherwise, we just add randomness in wages and to the length of the technological cycle. In Figure 5 the first 20 period experience good ε_t^2 shocks which leads to a very long technological cycle (between 70 and 80 periods depending on h). In contrast, in later periods some technology cycles have only two periods. Note that, since the variance of shocks depends on the state of the technology currently in use, a good realization of ε_t^2 can increase the variance of both series and, in particular, the variance of w_t^R as ε_t^1 is a transitory shock.

[12] As with ε_t^1 in practice we choose a level of v for which technology remains positive. Alternatively we could use a lognormal shock with mean 1.

[13] We could, as before, eliminate the persitence in A_t^O, but since A_t^O represents the actual technology in use, inovations are likly to have a more lasting effect.

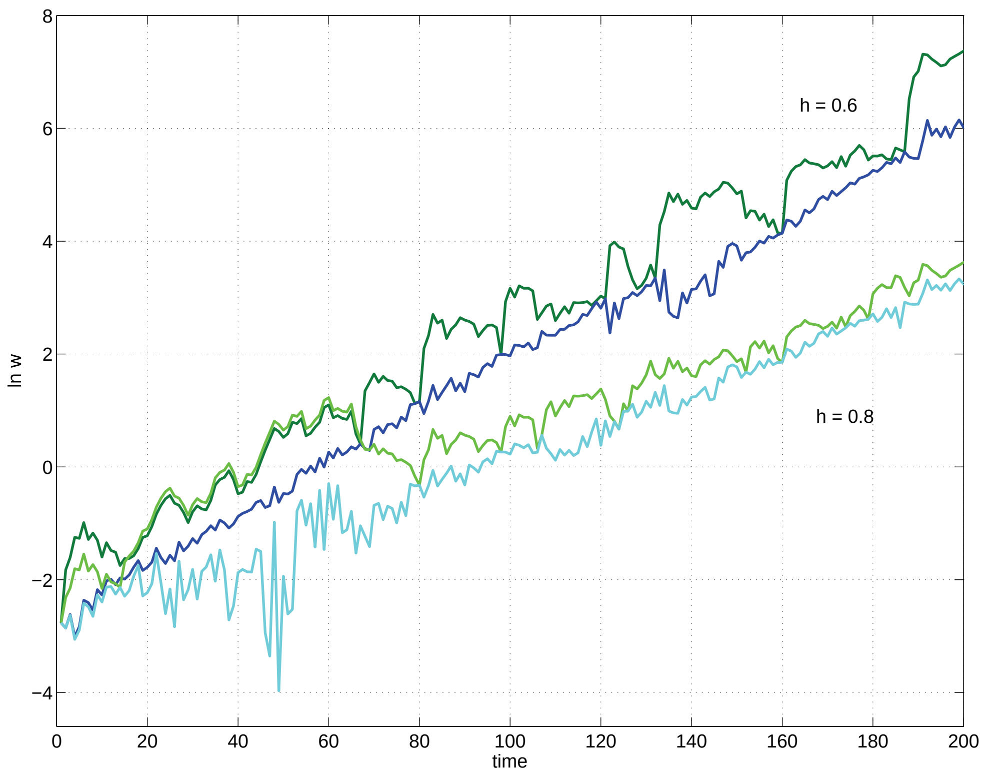

Figure 5: Two shocks

4. EMPIRICAL IMPLICATIONS AND EVIDENCE

Our theory leads to a variety of empirical implications. First, we have the implications of Proposition 5. Long-run average growth rates are determined by the quality of information technology and the cost of acquiring knowledge or information. These has both cross-sectional and time series implications. As communication technology within a country improves we should see larger growth rates. So the model implies that telephones, mobile technology, or the internet should have positive growth effects. Empirically contrasting these predictions with the data is, however, not a simple task. The reason is that we would need long time series of data, for several countries, since the model also predicts long technological cycles that may be hard to distinguish from long term growth rates. We would like to interpret communication technology

h and the cost of acquiring knowledge broadly c. When we talk about basic economic systems 1000 years ago, ability to communicate across long distances using roads, common languages or religions for large populations, or geographic concentration in large cities may be important determinants of h. In contrast, today the internet and information and communication technology is probably the main determinant of h, although geography may still be important.

Second, our model implies that we should observe long term technology cycles in which we see an initial acceleration of growth and a later decline in growth rates. Throughout the technology cycle, organizations should become more complex and knowledge acquisitions should be intense. The level of output increases at a decreasing rate but the rate at which new knowledge is acquired is weakly increasing. This is what makes new technologies eventually catch up with the used and old, but organized, ones.

It is important to caution the reader that ours is not a theory of the size of firms only of the size of organizations. In that sense it is hard to contrast the prediction that organizations become more complicated with firm size data. However, the model also predicts that new organizations need to be created to deal with the new technologies once a radical innovation is adopted. Some of these organizations are likely to be constituted as firms. Old organizations should be dismantled and new simple organizations have to be created. These new organizations will grow and become more complex with time. Causal observation is consistent with this implication. Wal-Mart, Google, Microsoft, Apple, E-Bay are all new and large organizations that started small, have grown, and have replaced the old large organizations. An expression of this phenomenon are the famous stories (such as Hewlett-Packard) where founders (in HP's case, Bill Hewlett and Dave Packard) developed the idea that gave raise to a large firm in their garage.

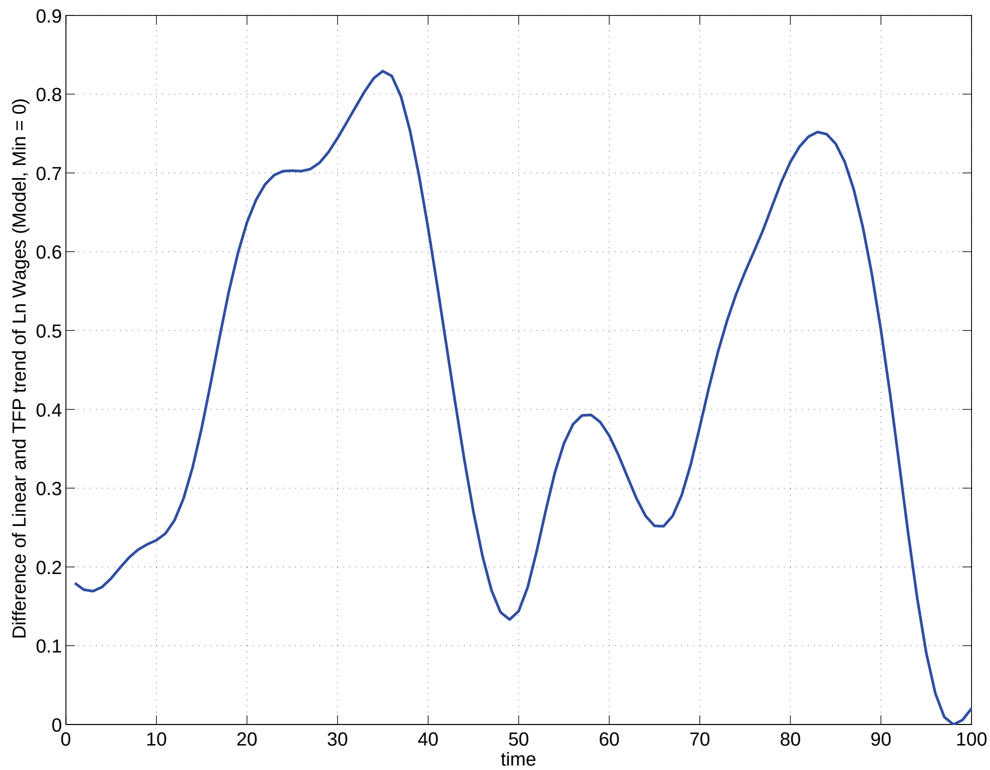

Figure 6: Long-Term Technology Cycles in the Model

Our model predicts that we should observe technological cycles in the GDP per hour data. To do this, we need to find a way of identifying these large cycles from other smaller fluctuations as the ones observed in Figure 5. We will use the following strategy. Take the series of wages generated by the model for a given set of parameter values. Then, we calculate the linear trend of log wages, as well as the Hodrick-Prescott filter of the series with smoothness parameter equal to 100 in order to capture the long term cycles. The difference between the linear and H-P trends captures the long term cycles (we normalize the minimum difference to zero). If we apply this methodology to an exercise similar to the one in Figure 5 but with $h = .7$ we obtain the technology cycles over 100 periods. We plot them in Figure 6. The graph identifies three technological cycles. The first one about 50 years long (that maybe in turn be broken in two), the next one about 20 years long, and the last one about 30 years long.

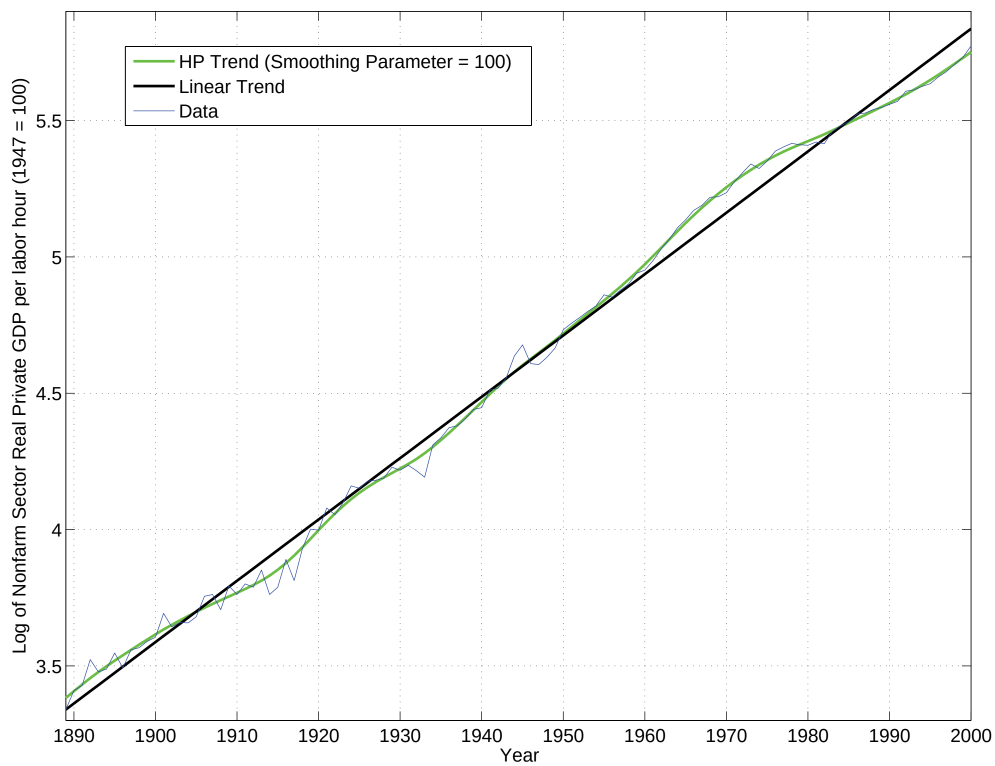

Figure 7: Data and Trends

We can do a similar exercise using US data. We use GDP per hour in the non-farm private sector from 1989 to 2000. The data comes from the Historical Statistics of the United States, Millennial Edition. Figure 7 presents the actual data in natural logarithms and the linear and the H-P trends with smoothing parameter equal to 100. The growth rate of the linear trend is equal to 2.2%. Figure 8 presents the difference between the linear and the H-P trends which should identify the long-term technology cycles in the data. This methodology identifies three long term technological cycles. A large one at the turn of the century. A relatively small one in the middle of the century and a very large one in the second half. The one in the second half is related to what has been called the productivity slowdown. In the last part of the 90's we seem to be moving to a new technological cycle.

What we can learn from these data that these large cycles are present and that they are a feature of the data that the theory we have presented can generate. In fact, Schumpeter (1939) had already analyzed them for a different time period. Of course, the above graphs cannot be compared directly as the theoretical exercise was not designed to match the data. In particular we are using random shocks and not shocks identified in the data. However, one of the main features of our theory is that these technological cycles should be observed and should be relatively large, which seems to be the case in the data where the deviations from the linear trend can be as large as 18%.

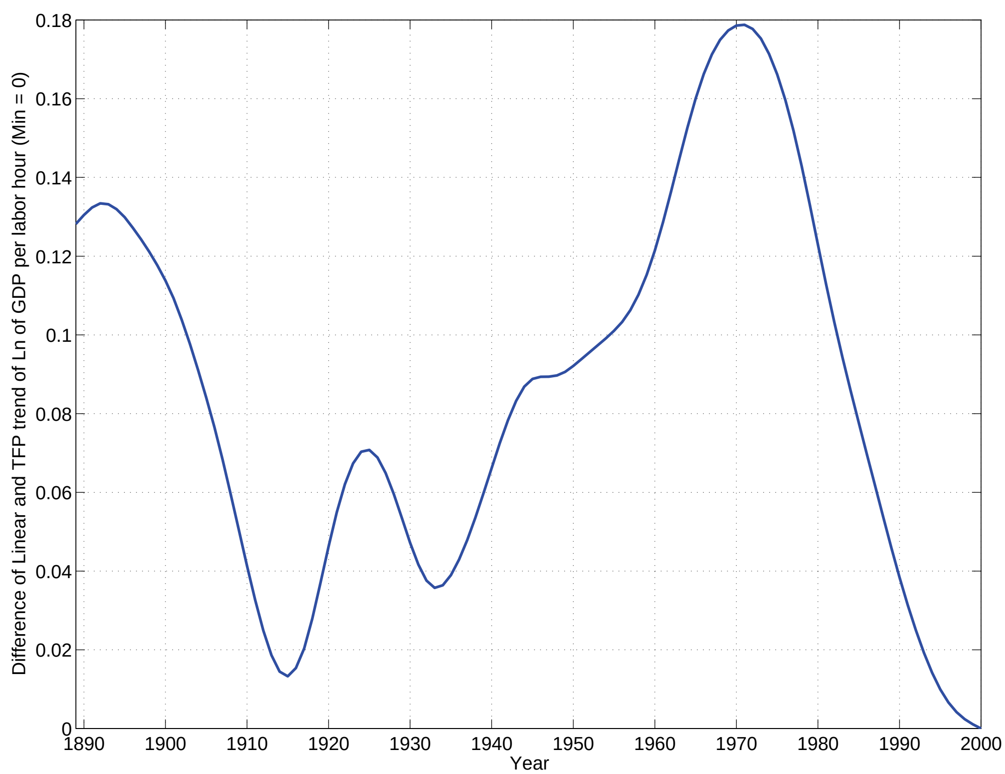

Figure 8: Technology Cycles in the Data

The key prediction of our theory is that development is the result of the ability of countries to organize production through economic organization. The extent to which different economies exploit available technologies by organizing in complex organizations is mediated by communication technology, the cost of acquiring knowledge, and the distribution of problems faced in production. Any change in these fundamental parameters will change the number of periods τ that a technology is used, the average (over time) size of these organizations, as well as the output level and growth rate in the economy. Organizations can take many forms as we have discussed above. However, it is hard to obtain information on all these different forms. Hence, in the rest of this section we assume that at least some of these organizations incorporate in firms and look for evidence in firm dynamics and firm sizes consistent with the predictions of our model of organization.

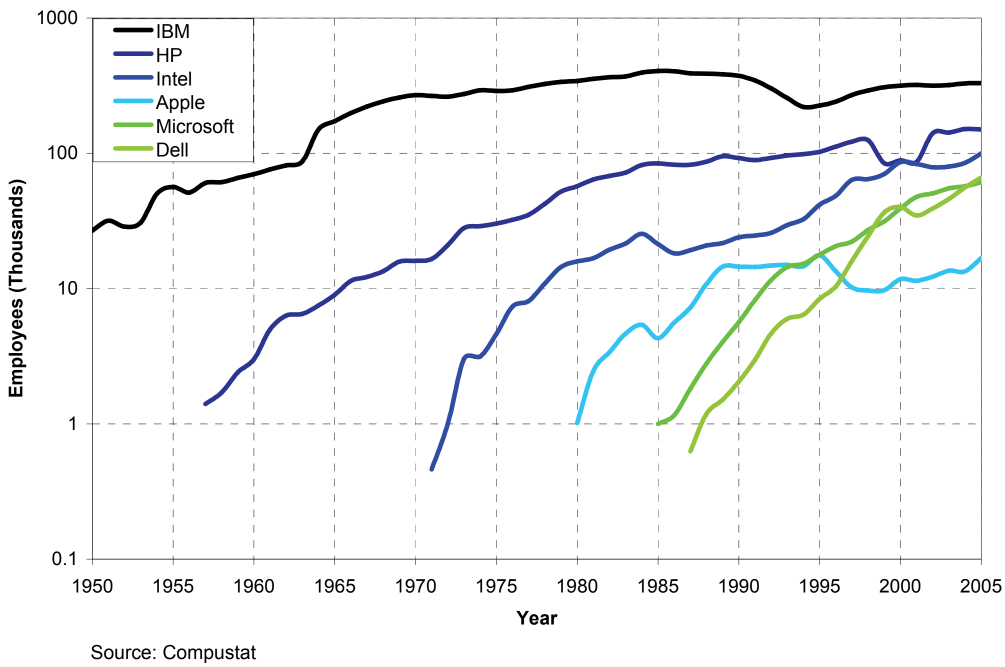

Figure 9: Evolution of the Numer of Employees of 6 Large High-Tech Firms

We present three pieces of evidence which are consistent with our model. First, our theory predicts that the size of these organizations should increase at a decreasing rate. In fact, we should observe the type of concave evolution observed in Figure 1. At least, when we focus our attention on the largest firms in the economy which are more likely to encompass the whole organization. This is the case no matter if we look at output per worker or at the total number of workers in the organization. Figure 9 presents evidence on firm sizes over time for a collection of Fortune 500 high-tech firms in the US (see also Luttmer, 1997). The data on the natural logarithm of the total number of employees in these firms comes from Compustat and includes all employees (domestic and foreign). The period depicted is governed by data availability and the initial period in which these firms became public. It is clear from the graph that the growth rate of these firms is decreasing over time as our model predicts.

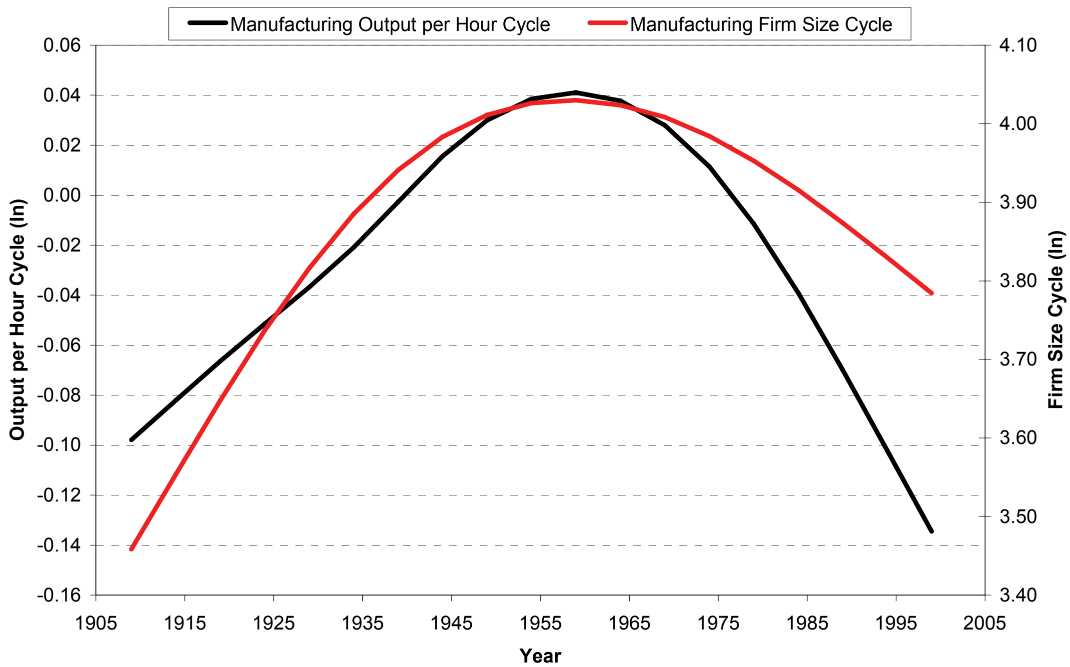

Figure 10: Manufacturing Firm Size and Output per Hour Cycles

The second piece of evidence relates to the long term cycles in output and how they relate to long term cycles in average firm size. The model predicts that we should observe these long term cycles in both average firm size and output. We use data from the Historical Statistics of the United States as above, but only for manufacturing since we can obtain a historical series of firm size only in this sector. We apply the same methodology used above to measure the long term cycle in average firm size and in output in the manufacturing sector to identify long term cycles. The only difference is that our model does not imply a linear trend underlying the cyclical fluctuations in average firm size so we do not subtract a linear trend from the HP filter. Unfortunately, the lack of long historical data series on firm size prevents us from capturing many of these long term cycles. However, the data is sufficient to capture one of them. Figure 10 shows the long term cycle in manufacturing production and average firm size in manufacturing. Clearly, both series move together during these period, as our theory predicts.

Finally our model also implies that growth rates and organization are negatively correlated over the short run and positively correlated over longer periods of time. Over the short run when new radical innovations are starting, large firms are being destroyed and small firms are just starting to grow. The larger the firm size the more the innovation has been exploited and thus the lower the growth rate. Over long periods of time, conversely, however, countries with larger organizations on average are those where growth rates are higher, as more knowledge is accumulated. We present here some preliminary evidence on this prediction. Table 1 presents short run (5 year) and long run (20 years) growth rates for 26 EU countries between the years 1984 and 2004 for which Eurostat has collected firm size data for European countries as well as 5 and 10 year average firm sizes in up to 20 sectors in those countries. The average GDP per capita (in constant local currency units) growth rates for 5, 10, and 20 years were constructed using data from the World Development Indicators (Poland, Lithuania, Czech Republic and Slovenia were excluded from the sample because they did not presented data on GDP per capita prior to 1994). The series of average firm sizes corresponds to averages for 5 (1999-2004) and for 10 (1995-2004) years and were constructed using the ratio of the series: Number of employees (code: v16130) and Number of enterprises (code: v11110) for each two digit sector (NACE classification) in each country using data from Eurostat.

We present results with and without sector fixed effects (which makes no difference for our results). The regression coefficients should not be read as causal, but as conditional and unconditional correlations. The table shows that indeed over shorter runs, smaller firm sizes are correlated with larger grow rates, while over longer periods the opposite is true. Although these last correlations are not significantly different than 0, we have found the positive sign to be robust to all specifications we have tried. This evidence suggests, although by no means allows us to conclude, that the relation between organization and growth may have a shape like the one predicted by our model. Obviously, more research is needed, particularly, since our data include the transition of many countries from communism to market economies.

	5 year Growth Rate			10 year growth rate			20 year growth rate		
Average Size	-0.059	-0.062	-0.077	-0.038	-0.041	-0.051	0.007	0.006	0.007
(over 5, 10 years*)	(0.013)	(0.012)	(0.013)	(0.009)	(0.009)	(0.01)	(0.005)	(0.005)	(0.005)
Ln GDPx100		-0.201	-0.202		-0.131	-0.132		-0.034	-0.033
		(0.023)	(0.023)		(0.017)	(0.017)		(0.009)	(0.009)
	N	N	Y	N	N	Y	N	N	Y
Intercept	0.034	0.158	0.159	0.033	0.114	0.115	0.02	0.041	0.041
	(0.001)	(0.014)	(0.014)	(0.001)	(0.011)	(0.011)	(0.001)	(0.006)	(0.006)
N Obs	399	399	399	399	399	399	399	399	399
RSq	0.0519	0.2044	0.2024	0.042	0.166	0.164	0.005	0.038	0.0379

Table 1: Firm Size and Output per Hour Growth Correlations
* 5 years average firm size (in hundreds of workers) in the first column; 10 years average in the other two.

5. INCORPORATING THE FRAMEWORK IN A NEOCLASSICAL GROWTH MODEL

So far we have developed a model of growth in which the only factor of production is labor. This has helped us highlight the main mechanism at work in generating permanent growth through organizations and radical innovation. It is easy to expand our framework to introduce capital accumulation. To show this we need to introduce agents that live for more than one period. Let agents live for two periods. In the first period they work and earn their wages as described in the previous sections. Instead of consuming all of what they earn in the first period, agents now save part of their labor income to consume in the second period when they retire. A new generation of size N is born every period without capital. So let the preferences of an agent born in period t be given by

$$U\left(c_t, c_{t+1}\right) = c_t^{\alpha} + \beta c_{t+1}^{\alpha}.$$

The problem of the agent is then to maximize $U\left(c_t, c_{t+1}\right)$ subject to the inter-temporal budget constraint

$$\omega_t = c_t + \frac{c_{t+1}}{\rho_t}$$

where ρ is the gross interest rate and ω_t is the level of wages in the economy. Consumption levels are then given by

$$c_t = \frac{\rho_t^{\frac{1+\alpha}{\alpha}} \beta^{\frac{1}{\alpha}} \omega_t}{\left(1 + \rho_t^{\frac{1+\alpha}{\alpha}} \beta^{\frac{1}{\alpha}}\right)},$$

$$c_{t+1} = \frac{\rho_t \omega_t}{\left(1 + \rho_t^{\frac{1+\alpha}{\alpha}} \beta^{\frac{1}{\alpha}}\right)},$$

and capital per capita is c_{t+1}/ρ_t.

We also need to modify the production function so as to incorporate capital. Towards this, define

$$\begin{aligned} H_t &\equiv w_t \\ &= A_{t-L_t-1} \left[F\left(Z_{L_t,t}^{L_t}\right) n_{L_t,t}^0 - c \sum_{\ell=0}^{L_t} z_{L_t,t}^{\ell} n_{L_t,t}^{\ell} \right] \end{aligned}$$

where L_t denotes the highest layer of organization in period t. Note that under our assumptions, H_t is determined exactly as described in the previous sections as it is independent of the savings decision. The reason is that agent's decisions of how much knowledge to accumulate maximize the term in brackets in the previous equation, and that decision is static since agents only work for one period. We also need to make sure that the switch to a new technology is independent of the level of capital, which is the case since agents are price takers and, again, only receive wages in one period. H_t has an intuitive interpretation. What used to be the wage now is the level of labor-augmenting technology which is determined endogenously in our model and varies with time as described above. That is, the growth model above can be interpreted as a model of the growth in labor-augmenting technology (or the level of efficiency units of labor). Through time, this labor augmenting technology evolves as labor is more efficiently used by building more sophisticated organizations, but also by doing radical innovations. This part of the model is independent of the capital accumulation process, and so it can be embedded in a standard capital accumulation model but with endogenous technology evolution and, through the evolution of this technology, permanent growth.

Let production in the economy be determined by a constant return to scale Cobb-Douglas production function $Y(K_t, NH_t)$, where K_t denotes aggregate capital saved and available for production in period t. So we assume that the capital used in production is the capital that agents saved in the same period (not last period). Note that

$$K_t = \frac{\omega_t}{\left(1+\rho_t^{\frac{1+\alpha}{\alpha}}\beta^{\frac{1}{\alpha}}\right)}N.$$

The problem of an organization is to hire capital and the labor composite so as to solve

$$\max Y(K_t, NH_t) - \rho_t K_t - \omega_t N,$$

where we are assuming that capital depreciates completely every period. If γ is the share of capital, then

$$\frac{\rho_t K_t}{Y(K_t, NH_t)} = \gamma$$

and

$$\frac{\omega_t N}{Y(K_t, NH_t)} = 1-\gamma,$$

which determines ρ_t and ω_t given H_t. Let

$$k_t = \frac{K_t}{NH_t} = \frac{\frac{\omega_t}{H_t}}{\left(1+\rho_t^{\frac{1+\alpha}{\alpha}}\beta^{\frac{1}{\alpha}}\right)}.$$

Then

$$\rho_t = \gamma\frac{y(k_t)}{k_t},$$
$$\frac{\omega_t}{H_t} = (1-\gamma)\,y(k_{t-1}),$$

where $y\,(k_t) = Y(k_t, 1)$.

In the balanced growth path we will have $y^* = y(k_t) = y\,(k_t)$ and so $k^* = k_{t-1} = k_t$ which implies that $\frac{\omega_t}{H_t}$ and ρ_t are constant so

$$\rho^* = \gamma\frac{y(k^*)}{k^*} = \gamma k^{*\gamma-1},$$
$$\omega^* \equiv \frac{\omega_t}{H_t} = (1-\gamma)\,y(k^*) = (1-\gamma)\,k^{*\gamma}.$$

Note that wages are therefore given by $\omega H_t = \omega w_t$, where w_t evolves exactly as described in the previous sections. So in the balanced growth path k^* is implicitly given by

$$1+\left(\gamma k^{*\gamma-1}\right)^{\frac{1+\alpha}{\alpha}}\beta^{\frac{1}{\alpha}} = (1-\gamma)\,k^{*\gamma-1}.$$

This equation can we used to show that, as expected, k^* increases with β and γ.

In sum, in the balanced growth path, variables normalized by H_t are constant. In the previous section we have described the evolution of H_t. We have argued that it grows fast at the beginning of a technological cycle. Growth slows down at the end of the cycle, until a new radical innovation is implemented which starts a new technological cycle. This will then be the evolution of output, output per capita, capital, and wages in this model. In contrast, the interest rate will remain constant in the balance growth path. Note that the average long-run growth rate of output, output per capita, capital and wages will be the same a the one for H_t, which is the same as the average long-run growth rates derived in the previous sections, namely, $\tilde{g}$.

The objective of this section was to illustrate how our model of knowledge accumulation and organizational change could be incorporated in a standard neoclassical growth model with capital. This highlights our contribution, which is to develop a micro-founded model of the evolution of labor augmenting technology and therefore of endogenous economic growth. In order to do this, we took two shortcuts. First, we have assumed that the depreciation rate is equal to one. This served us only to simplify the notation. Adding a capital depreciation rate in this setup is immediate. Second, we have made two assumptions to simplify the decision of when to switch technologies: that agents only work for one period and retire in the next and that capital is saved and used in the same period. Removing these two assumptions would not affect the dynamics within a technology, but it would affect agents' calculations about when a radical innovation should be developed. Agents would need to take into account the whole future path of wages when deciding to switch to a new technology, and this in turn changes the evolution of H_t. While the main qualitative insights of our analysis would be unlikely to change, specifically the 'bumpy' pattern of growth through technological innovations and the ebb and flow of organization, such a development would substantially complicate the model. Thus we leave that extension for future work.

6. CONCLUSIONS

Change is a fundamental aspect of growth. Growth is not a smooth continuous process of accumulation. As the 'new growth literature' has recognized, it involves the creation and destruction of products and, as we underscore, of organizations. Each new idea requires a particular type of knowledge, and new organizations —with experts in the relevant sets of knowledge— gradually emerge to exploit this knowledge. When technology is revolutionized, knowledge becomes obsolete and so do the organizations that have developed to allocate it and exploit the existing technology. We have presented here a model that captures this process. Incremental efficiency gains take place within existing organizations. As organizations grow by adding more hierarchical layers, they steadily improve the efficiency of the allocation of agents in production. Radical innovation makes existing knowledge, and organizations, obsolete. Decreases in communication costs that allow for the establishment of deeper hierarchies permit more exploitation of existing knowledge. Better communication thus increases the growth rate through a more efficient exploitation of technology.

Our model has provided us with some insights on the interplay between growth and organization. First, we rationalize and document that developed economies have larger and deeper organizations.[14] Second, increases in the share of problem solvers in the economy will not translate, on average, in larger productivity growth. In fact, the periods where the proportion of knowledge workers is highest are those where technology has reached the diminishing returns and thus those with the lowest productivity growth. Third, our analysis produced some clues on the source of the long term persistence of temporary productivity shocks and the importance of institutions in generating growth.

While our analysis allows us to study the impact of organizational variables (notably the cost of communication) on both growth and organization, our approach has some obvious limitations. Most importantly, since we have a one good economy, when a radical innovation is introduced, the existing knowledge, and the existing organization, is made obsolete. Clearly, this is a limitation that leads to an extreme conclusion:

[14] As technology cycles become longer due to better communication technology the economy will exhibit lower frequency long term fluctuations. This maybe misinterpreted as a reduction in volatility in the data as observed in the last part of the twentieth century: the 'Great Moderation'.

after a radical innovation, existing organizations are wiped out. While it is quite reasonable that the development of the automobiles wiped out the stagecoach industry, it is clearly not the case, in a multi-good economy, that all existing firms disappear. A generalization of the model to a world with differentiated commodities would yield a smoother prediction, while preserving the key empirical implication: the correlation between productivity bursts, entry of new organizations, and exit of existing ones. Especially, if some radical innovations, like new general purpose technologies, affect several industries concurrently.

We view our analysis as the start of an effort to understand, at a deeper microeconomic level, the use of the labor input usually introduced in aggregate production functions. What matters for development is not how many units of labor are used, but how these units are organized, and how this changes over time. The dynamics in our theory are due to the difficulty of building up organizations and of acquiring the relevant pieces of complementary knowledge. Or, in other words, the dynamics are the result of the difficulty of forming markets so that agents can sell their specialize knowledge and buy the knowledge of others. We believe that, in a world where the sources of growth are the creativity and the ideas of individuals, rather than raw materials and capital, understanding the way individuals organize to produce is fundamental to our understanding of the observed income differences across countries.

REFERENCES

[1] Acemoglu, Daron and Robinson, James, 2005a, "Institutions as the Fundamental Cause of Long Run Growth,", *Handbook of Economic Growth.* North Holland.

[2] Acemoglu, Daron and Robinson, James, 2005b. "Unbundling Institutions," *Journal of Political Economy*, 113: 5, 949-995.

[3] Aghion, Philippe and Howitt, Peter, 1992, "A Model of Growth Through Creative Destruction," *Econometrica*, 60:2, 23-351, March.

[4] Barro, Robert, 1996, "Democracy and Growth,", *Journal of Economic Growth*, 1:1, 1-27, March.

[5] Garicano, Luis and Rossi-Hansberg, Esteban, 2006, "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 121:4, 1383-1435, November.

[6] Grossman, Gene M. and Helpman, Elhanan, 1991, "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 59:1, 43-61, January.

[7] Howitt, Peter, 1999, "Steady Endogenous Growth with Population and R & D Inputs Growing," *Journal of Political Economy*, 107:4, 715-730, August.

[8] Jones, Charles I., 1995, "R & D-Based Models of Economic Growth," *Journal of Political Economy*, 103:4, 759-84, August.

[9] Jones, Charles I., 1999. "Growth: With or Without Scale Effects?," *American Economic Review*, 89:2, 139-144, May.

[10] Jovanovic, Boyan & Rob, Rafael, 1990. "Long Waves and Short Waves: Growth through Intensive and Extensive Search," *Econometrica*, 58:6, 1391-1409, November.

[11] Jovanovic, Boyan & Rob, Rafael, 1989. "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, 56:4, 569-82, October.

[12] Kortum, Samuel S.l 1997. “Research, Patenting, and Technological Change.” *Econometrica*, 65:6, 1389-1419, November.

[13] Kuznets, Simon., 1940. “Schumpeter’s Business Cycles,” *American Economic Review*, 30:2, 257-271, June.

[14] Kydland, Finn E. and Prescott, Edward C. 1982. “Time to Build and Aggregate Fluctuations,” *Econometrica*, 50:6, 1345-70, November.

[15] Lucas, Robert Jr., 1988, “On the Mechanics of Economic Development,” *Journal of Monetary Economics*, 22:1, 3-42, July.

[16] Luttmer, Erzo, G. J., 2007, “New Goods and the Size Distribution of Firms,” mimeo, University of Minnesota.

[17] North, Douglass C., 1981, *Structure and Change in Economic History*, New York; W.W. Norton & Co.

[18] Romer, Paul M, 1990. “Endogenous Technological Change," *Journal of Political Economy*, 98:5, S71-102, October.

[19] Schumpeter, Joseph A., 1939, *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*, Vols. I and 11. New York: McGraw-Hill.

[20] Schumpeter, Joseph A., 1942, *Capitalism, Socialism and Democracy.* New York: Harper and Brothers.

[21] Young, Allyn A., 1928, “Increasing Returns and Economic Progress,” *The Economic Journal*, 38, 527-542, December.

[22] Young, Alwyn, 1998, “Growth Without Scale Effects”, *Journal of Political Economy,* 106:1, 41-63, February.