

NBER WORKING PAPER SERIES

THE FIRM AS A DEDICATED HIERARCHY:
A THEORY OF THE ORIGIN AND GROWTH OF FIRMS

Raghuram G. Rajan
Luigi Zingales

Working Paper 7546
<http://www.nber.org/papers/w7546>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2000

We are grateful for support from the Brookings Institute and the National Science Foundation (grant #SBR-9423645). We benefited from the comments of Philippe Aghion, George Baker, Bruno Biais, Leonardo Felli, Edward Glaeser, Denis Gromb, Henry Hansmann, Oliver Hart, Thomas Hellman, John Moore, Canice Prendergast, Jean Tirole, and an anonymous referee. Heitor Almeida and Keiichiro Kobayashi provided excellent research assistance. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2000 by Raghuram G. Rajan and Luigi Zingales. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Firm as a Dedicated Hierarchy:
A Theory of the Origin and Growth of Firms
Raghuram G. Rajan and Luigi Zingales
NBER Working Paper No. 7546
February 2000
JEL No. D23, I 22

ABSTRACT

A fundamental problem entrepreneurs face in the formative stages of their businesses is how to provide incentives for employees to protect, rather than steal, the source of organizational rents. We study how the entrepreneur's response to this problem will determine the organization's internal structure, growth, and its eventual size. In particular, our model suggests large, steep hierarchies will predominate in physical capital intensive industries, and these will typically have seniority-based promotion policies. By contrast, flat hierarchies will be seen in human capital intensive industries. These will have up-or-out promotion systems, where experienced managers either become owners or are fired. Furthermore, flat hierarchies will have more distinctive technologies or cultures than steep hierarchies. The model points to some essential differences between organized hierarchies and markets.

Raghuram G. Rajan
University of Chicago
1101 E 58th Street
Chicago, IL 60657
and NBER
Rajan@gsbvax.uchicago.edu

Luigi Zingales
University of Chicago
1101 E 58th Street
Chicago, IL 60657
and NBER, and CEPR
luigi@gsblgz.uchicago.edu
<http://gsblgz.uchicago.edu>

At the root of most enterprises generating economic surplus is an entrepreneur with an unique critical resource such as an idea, good customer relationships, a new tool, or superior management technique. A fundamental problem of entrepreneurship is how to enlist the co-operation of the many agents necessary for production without ceding to them too much of the surplus generated by the enterprise. The risk of being expropriated is, however, always inherent in production. In particular, the entrepreneur has to give her employees (whom we call managers) close proximity or *access* to the critical resource for them to learn to produce effectively. For example, a manager has to understand the idea, be in contact with the key customers and suppliers, or even learn the entrepreneur's unique managerial techniques, in order to work effectively. But access also gives the manager the opportunity to expropriate this critical resource and compete against the entrepreneur. Managers may steal the idea, walk away with clients, or mimic the entrepreneur's management style, and start up a rival concern. The greater the access a manager has, the more he can appropriate, and the more effectively he can compete. This paper explores the role of organizational design in dealing with the problem of expropriation, and derives implications for the shape, size, and growth of organizations.¹

Although it is intended to serve as a metaphor for all those activities that allow managers to appropriate part of the value of the firm, the particular phenomenon we have in mind is not unimportant. Intel, the microprocessor manufacturer, was started, not in a garage or basement as many other Silicon Valley start-ups, but when Robert Noyce, the General Manager of Fairchild Semiconductor, and Gordon Moore, its head of Research and Development walked out of Fairchild and set up their own firm, Integrated Electronics. Shortly before their departure, a scientist in Moore's department had discovered the "silicon-gate" technique to produce semi-conductor memory devices. This became an important part of Intel's proposed product line. As a former employee of both companies put it "Intel was founded to steal the silicon gate process from Fairchild." [Jackson

1997, p26-27]. Clearly, of all Fairchild's employees, Noyce and Moore had the greatest access to Fairchild's inventions, and at the very least, took a lot of knowledge and, equally important, employees with them to the start-up. Thus Intel hit the ground running, and is now one of the most profitable firms while Fairchild Semiconductor is virtually a footnote in business history. The circumstances of Intel's founding are not an exception. Bhidé [2000, p.94] reports that 71% of the firms included in the Inc 500 (a list of young, fast growing firms) were founded by people who replicated or modified an idea encountered in their previous employment. The phenomenon has also been important in the past. Many during the Industrial Revolution, including Arkwright who "invented" the water frame, appropriated rather than discovered the technological advances their names are associated with (Hall [1998, p. 310-347]).

A simple example will help clarify the basic trade-off. Consider a master watchmaker in a small town who has a more productive technology to make watches. The watchmaker would like to produce with as many managers as possible. For now, there are two candidates for employment, Ram and Pietro. The watchmaker knows that Ram or Pietro may expropriate her technology and set up manufacture on their own. Also, the town is small so that there is room for only one watchmaker, and since there are scale economies in advertising, the watchmaker who can produce the most will prevail in any competition. Two factors will determine how much a watchmaker can produce - how much of the technology she knows, and how many managers she has working for her.

The watchmaker will certainly perform the tasks that are at the core of her new technology herself so that the others do not directly learn how they are done. But she has to interact with the managers. She has two ways of offering *access* to herself and the technology. She can have Ram and Pietro interact only with her (see Figure I). We call this the horizontal hierarchy because Ram and Pietro are equidistant from the watchmaker and her technology, and the watchmaker

mediates all transactions within the organization. An alternative way is for the watchmaker to interact only with Pietro (or Ram), and ask Ram to report to Pietro (see Figure II). This is the vertical hierarchy where there is a chain of employees with the watchmaker at the top, and only Pietro sees the watchmaker and her technology.

Once hired, managers can either attempt to expropriate the technology and compete (a combination of actions we will term “compete”), or they can learn to perform the tasks their immediate superior has assigned them (i.e., specialize). Because superiors assign subordinates to tasks that complement their own skills, managers, once specialized, are useless without their superior and can be productive only in a team that contains the superior. Managers who neither specialize nor compete are useless in production.

When managers interact directly with the watchmaker, they can observe her technology at close quarters, though not perfectly. While the watchmaker can get a marginal product of 1 from each manager in her team (and from herself), a manager operating with the expropriated technology can produce only 0.75. Therefore, 75 percent is (for now) the degree of expropriability of the technology and is a measure of the difficulty of enforcing property rights. Managers who do not interact directly with the watchmaker cannot expropriate the technology and will specialize if they get enough from doing so.

Finally, we need to specify how output is divided among members of a team. Let us assume that output in a production team is shared through bargaining from bottom up, with each manager getting half of what he produces and half of what his subordinates send up - unless he happens to be senior most in which case he gets everything that is sent up (we will describe the bargaining game more formally later).

Now let us consider managerial incentives in each organizational form. In the horizontal hierarchy, each manager produces 1 after specializing, and keeps $\frac{1}{2}$ of it. A manager could also compete.

However, since the managers interact indirectly with each other via the watchmaker, they cannot produce with each other as a team if the watchmaker does not join. So a manager who decides to compete will have to produce with the stolen technology on his own. He can produce 0.75, but to be able to generate this output he will first have to survive the competition. Since the watchmaker can produce 1 on her own, she will prevail in any competition. As a result, the manager will anticipate getting nothing if he competes, and instead, will prefer specializing.

By forcing all interactions to go through the watchmaker, the horizontal hierarchy divides the managers, in such a way that they will be easily defeated by the watchmaker even if they do expropriate the technology. This allows the watchmaker to rule effectively. Landes [1986] argues that organizations were, in fact, structured thus: “..patents were not always the best way to protect knowledge. Instead, inventors preferred to try and keep devices and techniques secret, sometimes by so dividing the process that no one worker could penetrate the technique. This is what the great watchmaker Abraham-Louis Breguet proposed to do when he planned the mass production of watches... the aim was ..security”.

Matters are different in the vertical hierarchy. Ram cannot expropriate since Pietro is between him and the watchmaker. Ram will specialize to Pietro if specialization is relatively costless. But Pietro is in a very different position. If he specializes, he will keep half of the \$1 he produces plus half of the $\frac{1}{2}$ Ram gives up for a total of $\frac{3}{4}$. Pietro gets more than Ram, we will see, because the vertical hierarchy gives him *positional power* – Ram is productive for the watchmaker only if Pietro also joins them. This gives Pietro some bargaining power. Despite his positional power, however, Pietro will compete because Ram will follow him out (since Ram is useless in any production team without Pietro). Together with Ram, Pietro can produce a total of 1.5 if he competes. Since this is greater than what the watchmaker can produce on her own, Pietro’s team will win the competition. Moreover, Pietro only has to give Ram half of what Ram produces ($=0.5*0.75$) as compensation,

leaving 1.125 for himself.

We immediately see that the incentive for managers to compete is higher in the vertical hierarchy because superior managers – and not the watchmaker – command the loyalty of subordinates. The strength of their bench allows them to offset the technological advantage of the watchmaker, and win the competition. Note, however, that if the degree of expropriability of the technology were lower - less than 50 percent – Pietro would prefer to specialize.

Thus far, the horizontal hierarchy seems to dominate because the tactic of divide and conquer is so effective. But the horizontal hierarchy “works” because managers have no positional power whatsoever. While this is of little consequence when specialization costs managers little, it is more problematic when costs are high. To see why, let the one-time personal cost to managers of specialization be 1.01, and let each manager be able to work for two periods before retiring. Finally, let the resource have a lower degree of expropriability, say 0.49.

Now managers in the horizontal hierarchy will not specialize (though they will not expropriate either). This is because each period, their position gives them a payoff of only $\frac{1}{2}$, for a total of 1 over two periods, which is less than the cost of specialization. While, the relative lack of managerial power in a horizontal hierarchy works well to prevent expropriation, it provides poor incentives for the manager to specialize.

Even a vertical hierarchy does not seem possible since higher costs of specialization make competing still more attractive than specialization. In particular, assuming Ram specializes, Pietro does not want to specialize even though expropriability is at a level that ensured specialization when the costs of specialization were zero. By specializing, Pietro gets $2 * \frac{3}{4} - 1.01 = 0.49$ over two periods, while he gets $2 * 0.49 * 1.5 = 1.47$ by competing.²

There is a way, however, for the entrepreneur to use both Ram and Pietro in the vertical organization, though not in the first period. In the example thus far, Pietro will compete because

he is already in a position of power, with Ram below him, when he is required to specialize. Instead, the watchmaker should employ only Pietro in the first period. Pietro alone cannot overcome the watchmaker in competition. So he can either do nothing, or specialize and get $\frac{1}{2}$ immediately. In addition, Pietro knows that once he specializes, he will lose the incentive to compete in the second period (we know from earlier that if expropriability is less than 50 percent, and if Pietro does not have to incur the cost of specialization, he is loyal even with Ram as a subordinate). The watchmaker will then be willing to trust him in the second period and will bring Ram in below him as in Figure II. This structure will give Pietro positional power and enable him to get $\frac{3}{4}$. Anticipating a total payoff of $\frac{1}{2} + \frac{3}{4}$ over two periods, Pietro will specialize at the outset. Moreover, since Pietro can work only two periods, Ram will anticipate his turn at power in period 3, and also specialize and so on...³

Therefore, by restricting access initially to the technology to Pietro alone, the watchmaker can obtain Pietro's specialization. Pietro can then be entrusted with more positional power even though he would have chosen to compete if placed in that position when unspecialized. A specialized manager is more loyal, and a watchmaker with specialized managers can produce more than one starting anew. Specialized employees become a critical resource in their own right that grow slowly in number over time, even though nothing in the technology prevents infinite employees from being hired and specializing at one go. The hierarchy evolves through specialization into something more than a collection of people and resources.

Finally, there is a way to give managers in the horizontal hierarchy the incentive to specialize when specialization is costly, despite their impotence. It is to promise them potential participation in the rents from ownership of the critical resource if they specialize. The watchmaker will ensure an adequate rent to managers by limiting the number who have access and thus can bid for ownership. Moreover, when the critical resource is highly expropriable, the watchmaker's promise

to sell is credible because it is very difficult for her to run the hierarchy as an inactive owner once her two periods of active life are over. Thus the model not only suggests why organizations like legal and consulting partnerships – where assets like client relationships are hard to protect - are flat, but also indicates what limits their size.

In sum, our model suggests that in the formative stages of a business, an entrepreneur uses control over access to the resource and specialized employees, as well as the allocation of ownership over the resource to design the right balance of power between herself and her managers. Too much power to managers will destabilize her own position, too little will give them little interest in the well being of the firm. In the vertical hierarchy, the entrepreneur controls access to the critical resource so as draw forth specialization, and then uses specialized employees strategically to control the actions of new employees. In the horizontal hierarchy, the entrepreneur limits access not just to limit the current power of employees, but also to give them confidence that if they specialize, they will be among a select few that will have a chance to own the organization. The relative abilities of these two distinctive structures to approach the technological production possibility frontier depends on the parameters of the model, and results in a variety of empirical implications.

The rest of the paper is as follows. In section I we lay out the framework. In section II we analyze vertical hierarchies in a simplified three-period model. The steady-state solution is derived in section III, while in section IV, we analyze the path to the steady state. In section V, we study horizontal hierarchies and compare them with vertical ones. In section VI, we discuss the implications of our model for a theory of the firm. Conclusions follow.

I. Framework

We consider an overlapping generations world where each agent has two periods of physically active life, a period when they are “young” and one when they are “old”. After these two periods, they retire and can, at best, exert control rights (which we shall define momentarily) for the rest

of his (or her) life. After retirement, an agent has a constant probability γ of surviving the next period. Everyone is risk neutral.

An agent called the entrepreneur, possesses a valuable resource that she wants to exploit. She needs “managers” to produce on a large scale. Many managers are born every period.

A. Technology

Since our focus is not on the technological limits to organizational size, we assume the production function is linear with each active agent contributing a unit of output. The total production of a team composed of an active entrepreneur and n managers equals $n + 1$.⁴

B. Sequence of events each period

The sequence of events is described in Figure III. At the beginning of the period, the agent who controls the resource determines the hierarchy of access to it.⁵ Then, the managers who are granted access choose whether to specialize or compete. If any manager decides to compete, the original hierarchy breaks up into competing teams. Only the team that can produce the most survives. Finally, all the members of the successful team bargain over future surplus and sign sharing contracts. At the end of the period, production takes place and the payoff to each contract is distributed.

C. Access

Following Rajan and Zingales [1998], we define access as the ability to use, or work with, the critical resource. Only a team of managers that has direct or indirect access to the resource can produce. Future access is not contractible.⁶

We will assume that one active agent (and at most one agent), has to have direct access to the resource. Call this agent the *head*. Other managers can have indirect access by having

access to the head, or access to a manager who has access to the head and so on... The more layers of management between a manager and the resource, the less access he has. We start by examining *vertical* hierarchies of access where each manager has access to at most one member of the production team who is closer to the resource (his superior) and one manager who is further (his subordinate). If we define the head to be tier 1 then a manager's tier, t , is uniquely defined as one plus the tier of his superior. Figure IV depicts a vertical hierarchy with the entrepreneur as the head and four additional managers.

D. Specialization

Managers can add value to a team only after they *specialize*, that is, after they familiarize themselves with the resource (which is why they need some access) and learn to work with their superior in the production team. A manager's cost of specializing is c^S . Since the cost of specialization (or training) for new recruits exceeds their value to the firm (see, for example, the evidence in Harhoff and Kane [1993] and the literature cited in Prendergast [1999]), and managers produce at most 1, we assume:

Assumption 1: $c^S > 1$

A manager who specialized to a superior in the past does not need to specialize again if he has the same superior this period, since he already knows the superior and is familiar with the resource. The head does not have a superior, and does not have to specialize if he did so in the past.⁷

E. Competition

Access also provides a manager with the opportunity to appropriate the critical resource. A manager could grab the resource and compete with the entrepreneur instead of specializing. His subordinates have no choice but to follow him since they are specialized to him.⁸

Suppose a manager in tier k decides to compete. Together with his $n - k$ subordinates, he can produce $\Lambda_k(n - k + 1)$ where Λ_k is the degree of expropriability in tier k . The further the manager is from the resource, the less he can expropriate, so $\Lambda_k < \Lambda_j < 1$ for all $k > j > 0$, and Λ for the entrepreneur is 1. In order that competition be a serious threat, we will assume that:

Assumption 2: $\Lambda_k > \frac{1}{2}$ for all $k > 0$.

The personal cost to the manager of competing is c^R .⁹ If a manager k decides to compete, all costs in subsequent periods for his production team are also multiplied by Λ_k , indicating the stolen resource is proportionally as hard to specialize to, or to appropriate. This preserves the symmetry of the problem.

After the initial sub period when managers choose simultaneously to specialize, do nothing, or compete, there is no time to train additional managers before production takes place. So in Figure IV, if manager M_3 decides to compete while the rest specialize, there are two feasible production teams: the first is headed by the entrepreneur, E , with subordinate manager M_2 and it can produce 2, and the second is headed by manager M_3 with subordinates M_4 and M_5 and it can produce $3\Lambda_3$.

To simplify the dynamics, we assume extreme economies of scale in marketing, which result in a winner-take-all market where only the most productive team is successful and survives the competition. So, in the example above, the team headed by the entrepreneur will survive if $2 \geq 3\Lambda_3$, else the team headed by M_3 will survive. We break ties in favor of the incumbent head.

We want to study how organizations emerge even in the most primitive economy, where contracts are not easily enforced and sophisticated commitment mechanisms are not available. For this reason we follow the incomplete contract literature (e.g., Grossman and Hart [1986]) in assuming that the end-of-period payoff from production is not contractible until immediately before production begins. As a result, just before a successful team is ready to produce, the members will

have to bargain about their respective shares of output.

F. Bargaining

We assume the bargaining game to have the following extensive form. The manager at the bottom of the hierarchy bargains with a unified coalition of his direct and indirect superiors over the total surplus. The share the coalition gets from this first stage of bargaining is then further subdivided through bargaining, between the manager at the bottom of this coalition and a unified coalition of his superiors. This carries on until the highest member of the hierarchy bargains with the second highest and divides the residual.

In the bargaining between the coalition of superiors and the manager, the outside option of the coalition is to produce without the manager (and his subordinates). The outside option of the manager is to obtain a wage of $1 - \theta$ elsewhere.¹⁰ If a manager does not join the production team, none of his subordinates are of use to the team. This is because it is he who mediates their interaction with the rest of the hierarchy.

G. Outcome of Bargaining

Consider first a hierarchy with n managers where the entrepreneur has retired. Total production is also n since the entrepreneur contributes no labor. Now let us see how this surplus is split.

The lowest manager, M_n , with reservation value $(1 - \theta)$, bargains with the coalition formed by higher tiers, who have a collective reservation value equal to $n - 1$ (the value they produce on their own without this last manager). Splitting the incremental value he adds evenly, he gets¹¹

$$(1) \quad \pi_n = \frac{1}{2}[n - (n - 1)] + \frac{1}{2}[1 - \theta] = 1 - \frac{\theta}{2}.$$

Thus $\frac{\theta}{2}$ is the portion of the lowest manager's production the coalition of superiors gets because of

their control over the critical resource, and their total surplus is $(n - 1) + \frac{\theta}{2}$. The lowest manager of this coalition, M_{n-1} , then bargains with his superiors and he gets $\pi_{n-1} = \frac{1}{2}[(n - 1 + \frac{\theta}{2}) - (n - 2)] + \frac{1}{2}[1 - \theta] = 1 - \frac{\theta}{4}$.

Manager M_{n-1} gets more than manager M_n because only when M_{n-1} joins the production team will the hierarchy above him be able to use his subordinate, manager M_n . Thus manager M_{n-1} can extract half the surplus (i.e., half of $\theta/2$) that manager M_n sends up because he controls his subordinate's access.¹²

Manager M_k 's payoff in a hierarchy with n working managers is (see appendix):

$$(2) \quad \pi_k = 1 - \frac{\theta}{2^{n-k+1}}.$$

A manager who controls access to more subordinates, $n - k$, has more positional power, and thus more rents. Note that the manager's payoff does not depend at all on the number of his superiors. This simplifies the analysis considerably.¹³ Also, the average pay of employees increases in the size of the hierarchy, n , implying our model is consistent with the finding (e.g., Brown, Hamilton, Medoff [1990]) that larger firms pay higher wages on average.

The retired entrepreneur's payoff is the residual amount, which is easily shown to be

$$(3) \quad E(n) = n - \sum_{j=1}^n \pi_j = \theta(1 - \frac{1}{2^n}).$$

Thus, and quite naturally, the retired entrepreneur's rents increase as managers' reservation wage $(1 - \theta)$ decreases. The entrepreneur's payoff if she is active and heads $n - 1$ managers can be obtained by simply adding her rents from control of the resource to the payoff she would get as the top active manager, the rents to her human capital. Thus her payoff if she is active is $\theta(1 - \frac{1}{2^n}) + (1 - \frac{\theta}{2^n})$. Since $E(n) - E(n - 1) > 0$, the entrepreneur's payoff increases in n . This

implies that in the absence of any other consideration, the entrepreneur would set the length of the hierarchy to its first best value, i.e., infinity.

The two important features of this bargaining structure are: i) for a given size of the hierarchy a manager's rent increases as his position moves up the hierarchy; ii) for a given position, a manager's rent increases in the number of subordinates. These features are more general than the particular bargaining structure in which they are derived. They come from the hierarchy of access, which gives managers positional power: Since only continuous chains of specialized managers are productive, intermediate managers get some rent because they are indispensable links to lower managers.

H. The Entrepreneur's Problem

The entrepreneur's objective is to maximize the present value of the rents she will get. Long term compensation contracts cannot be credibly committed to, so managerial compensation will be determined by bargaining, which in turn depends on the skills acquired by the manager and their role in the hierarchy of access.¹⁴ The entrepreneur therefore can shape managerial incentives only through organizational structure. We have already examined one tool she can use, control over access, which she uses every period. Another tool entrepreneurs have when they start out is the extent to which their organization is differentiated.

The entrepreneur often can determine the extent to which the manager is mobile. For example, among equally productive technologies, the entrepreneur can choose one that is more unique, making it harder for an un-cooperative manager to find a job elsewhere. She can locate far away from comparable employment so that the manager has to relocate completely if he quits (e.g., Rodriguez, [1999]). She can make the organization's culture distinctive so that managers get disutility working elsewhere. This ability of the entrepreneur to "differentiate" the technology, location, or culture will be reflected in the specialized manager's reservation wage conditional on specializing. Specialization will take the manager's reservation wage from zero to $(1 - \theta)$, where θ

is the extent of differentiation chosen by the entrepreneur and $1 \geq \theta > \underline{\theta} \geq 0$. Since differentiation is based on factors like technology that are typically hard to alter, we assume the entrepreneur picks it once and for all at the outset.

Having determined the degree of differentiation at the outset, the entrepreneur will choose access each period – which includes setting both the number of managers who have access and also who reports to whom – so that the maximum number of managers will all specialize; more specialized managers this period implies more production, and since managers can always be let go next period, it poses no constraints on future production. Since the cost of specialization exceeds the maximum possible single period rent any position in the hierarchy can offer, unspecialized managers will be given access only when they are young. Furthermore, an old manager will be given access only if he specialized in the past, and he can be placed as the head this period or below his former superior (else his past specialization is wasted and he has no incentive to specialize again). Also, the entrepreneur will not want to allow access to any manager who is expected to break away and compete since that manager (and his subordinates) are useless to the entrepreneur’s production team, even assuming they are not successful in the competition.

Most issues of organizational design can be examined in a simple three period model. This is where we start.

II. The Three Period Model

Let the world last for three periods and assume that the hierarchy reaches its large but finite steady-state length N in the third period. The steady state length, which we will derive formally later, turns out to be the maximum size the hierarchy can reach. For now, it gives us a simple terminal condition. Let n_t be the number of managers who get access for the first time in period t and specialize.

A. Managerial Incentives

A manager will not choose to compete if he expects his team to lose the winner-take-all competition against the entrepreneur and her followers. Call this the *competition constraint* (CC). Even conditional on expecting to win, he will not compete if he expects less surplus from doing so than from specializing. Call this the *conditional incentive compatibility* (IC) constraint. Only one of these constraints has to hold for the manager not to compete. Also, in his initial period the manager should prefer to specialize and get a rent rather than do nothing and get nothing (IR constraint). The entrepreneur will use differences in these constraints across managers and over time to build the hierarchy.

B. Incentives at date 2

Let us proceed by backward induction, and examine the incentives at date 2 first. The n_1 managers who specialized at date 1 are now old but still active. They are useful only if they are placed immediately below the still active entrepreneur in the same order that they were in last period – if any of these managers ends up with a different superior than in the past, he will have to specialize again. Since the one-period payoff does not compensate for the cost, he will not specialize and will render everyone below him useless to the hierarchy this period.

Of course, some or all the old managers could be fired at date 2 starting from the managers at the bottom. Let $n'_1 \leq n_1$ old managers be retained at date 2, and placed in their date-1 order immediately below the active entrepreneur. The n_2 young unspecialized managers who have a cost c^S of specialization will be placed after this.

We want to find conditions under which all managers specialize. If all specialize, the length of the productive team will be $1 + n'_1 + n_2$ because the entrepreneur is still active. Assuming all others specialize, a manager in tier k knows if he competes, he and his $1 + n'_1 + n_2 - k$ subordinates

will be defeated by the $k - 1$ member team consisting of his superiors and the entrepreneur if

$$(CC_{S2}) \quad \Lambda_k (2 + n'_1 + n_2 - k) < k - 1.$$

The fear of being defeated is least for the highest manager in the hierarchy ($k=2$); he has the most access, carries the most subordinates out with him if he competes, and only has to defeat the solitary entrepreneur. In fact, our assumptions ensure that the highest old manager will defeat the entrepreneur if he competes.¹⁵

The only reason then that he will specialize is if he gets more doing so than competing. For an old specialized manager in tier k this requires

$$(IC_{S2}) \left[1 - \frac{\theta}{2^{2+n'_1+n_2-k}}\right] \geq \Lambda_k \left[\theta \left(1 - \frac{1}{2^{2+n'_1+n_2-k}}\right) + \left(1 - \frac{\theta}{2^{2+n'_1+n_2-k}}\right)\right] + \gamma \Lambda_k \left[\theta \left(1 - \frac{1}{2^N}\right)\right] - c^R.$$

The left-hand side is the manager's rent from working for the hierarchy for the last period of his active life. The right hand side is the expected payoff from competing, where the first term is the break away manager's rent this period from control of the appropriated resource, the second is his rent from his own human capital, the third is his rent in the third period when he will be retired, but will still control the resource.

From assumption 2, it is easily seen that the right hand side of (IC_{S2}) will increase faster with the number of the manager's subordinates, $1 + n'_1 + n_2 - k$ than the left hand side.¹⁶ Therefore, the old specialized manager with the greatest incentive to breakaway and compete is again the one with the greatest number of subordinates, i.e., the one just below the active entrepreneur in tier $k = 2$. It turns out, however, that so long as the hierarchy's size is less than the steady state size, N , this constraint will not bind. The reason is that in steady state at date 3, the old head will have more subordinates and be closer to the resource (because the entrepreneur will have retired)

than any old manager at date 2. Since in steady state the head will not compete, it must be that no old manager wants to compete at date 2.

This implies that at date 2 the only limit to the size of the hierarchy is imposed by the incentives of the young, unspecialized managers. For them to specialize rather than compete, either they should not expect to survive the winner-take-all competition or they should expect to get less from competing even if they survive. Consider the latter condition first. From arguments similar to ones above, the young manager with the greatest incentive to compete is the one who is placed highest in his incoming cohort of n_2 managers, i.e., in tier $2 + n'_1$. Since the entrepreneur, and all managers who entered at date 1, retire at date 3, this manager will become the head at date 3 if he specializes. He specializes if

$$(IC_{U2}) \left[1 - \frac{\theta}{2^{n_2}}\right] + \left[1 - \frac{\theta}{2^N}\right] - c^S \geq \Lambda_{2+n'_1} \left[\theta \left(1 - \frac{1}{2^{n_2}}\right) + \left(1 - \frac{\theta}{2^{n_2}}\right)\right] + \Lambda_{2+n'_1} \left[\theta \left(1 - \frac{1}{2^N}\right) + \left(1 - \frac{\theta}{2^N}\right)\right] - c^R.$$

Comparing (IC_{U2}) with (IC_{S2}) , it is easy to see that (IC_{U2}) may be violated even when (IC_{S2}) is not. The left hand side of (IC_{U2}) , which is the young unspecialized manager's payoff if he specializes, includes two extra terms relative to an old manager's payoff (the left hand side of (IC_{S2})). First, since the old are in their final working period in the organization, they do not get rents next period (because they will retire). By contrast, a young manager has one more period in the hierarchy after the current one in which he can earn a rent. While this effect alone would give the young manager more incentive to specialize, the young manager also incurs the cost of specialization. Since this cost exceeds any single period rent, the net benefit from specialization is relatively lower for the unspecialized young than for the specialized old in the same position.

Also the right-hand sides differ. The benefit from competing is relatively higher for the unspecialized young because the young manager will be active in the third period, and expects the high rents from his own human capital (in addition to the rents from the control over the resource

which are also enjoyed by an old manager who competes). Taking both sides together, in any given position in the hierarchy, young unspecialized managers have a higher incentive to compete than old specialized managers. Therefore, (IC_{U2}) may be violated even if (IC_{S2}) is satisfied.

If so, the only reason young managers will not compete is if they do not expect to win the competition. Again, the highest young manager has the greatest likelihood of success because he is closest to the critical resource, and has the greatest number of subordinates below him. He will lose the winner-take-all competition if

$$(CC_{U2}) \quad \Lambda_{2+n'_1} n_2 < 1 + n'_1.$$

The left hand side is the amount the young unspecialized manager can produce with the rest of his cohort who follow him, the right hand side is the amount the entrepreneur can produce with the loyal old managers. Either (IC_{U2}) or (CC_{U2}) has to be satisfied for the young manager not to compete. So only one of the constraints will (generically) be binding, a fact that will greatly simplify the ex ante maximization problem.

It is now easy to see why specialized old managers are placed above unspecialized young managers in the hierarchy. Old managers benefit less than the young by competing, hence they are better suited to occupy the higher positions in the hierarchy where the incentive to compete is highest. Interestingly, such a policy of access also improves the incentive of the unspecialized young. For a given length of hierarchy, this policy ensures the highest young manager has the fewest possible subordinates, and is furthest away from the resource. Not only does such a policy give him the lowest return from competing even if successful, it also decreases his chance of being successful since the entrepreneur will be supported by the loyal Praetorian Guard of the specialized. As a result, the policy allows the entrepreneur to hire the most young. It follows that the number of specialized managers retained at the top of the hierarchy at date 2 is the entire number who specialized in

the first period, i.e., $n'_1 = n_1$. In the appendix, we show more generally that a seniority-based promotion policy where all specialized old managers are retained and placed in their prior-period order above unspecialized young managers provides managers the best incentives. Thus, we have

PROPOSITION 1. The level of production that can be achieved from any point in time onwards by a vertical hierarchy weakly, and sometimes strictly, increases in the number of managers who are active at that point and made specific investments in the past.

Now let us examine incentives at date 1.

C. Incentives in period 1

The hierarchy starts out in the first period, so all managers are unspecialized and young. The incentive constraints of the young manager immediately below the entrepreneur will be tightest, and we can write them as

$$(IC_{U1}) \quad \left[1 - \frac{\theta}{2^{n_1}}\right] + \left[1 - \frac{\theta}{2^{n_1+n_2}}\right] - c^S \geq \Lambda_2 \left[\theta \left(1 - \frac{1}{2^{n_1}}\right) + \left(1 - \frac{\theta}{2^{n_1}}\right)\right] +$$

$$\Lambda_2 \left[\theta \left(1 - \frac{1}{2^{n_1+n_2^R}}\right) + \left(1 - \frac{\theta}{2^{n_1+n_2^R}}\right)\right] + \gamma \Lambda_2 \left[\theta \left(1 - \frac{1}{2^N}\right)\right] - c^R.$$

$$(CC_{U1}) \quad \Lambda_2 n_1 < 1.$$

where n_2^R is the number of young managers who will be given access at date 2 by the manager who successfully breaks away at date 1. Since only one of these constraints has to be satisfied, the one that is satisfied with the most number of managers n_1 will be the limiting one.

D. IR constraint

Finally, we have ignored IR constraints thus far. A manager's rents increase as the number of subordinates below him increases. Thus, the manager who is at the bottom of the hierarchy today expects the lowest rents of his cohort if he specializes because the rest of his cohort will continue to be placed above him in the future. Therefore, the relevant IR constraint is that of the lowest young manager. Moreover, since the number of young hired increases over time, promotional prospects improve, and the most binding IR constraint is of the lowest unspecialized young manager in the first period which is¹⁷

$$(IR) \quad [1 - \frac{\theta}{2}] + [1 - \frac{\theta}{2^{n_2+1}}] \geq c^S.$$

E. The Entrepreneur's Maximization Problem.

We can now write down the entrepreneur's maximization problem

$$\max_{\theta, n_1, n_2} \theta(1 - \frac{1}{2^{1+n_1}}) + (1 - \frac{\theta}{2^{1+n_1}}) + [\theta(1 - \frac{1}{2^{1+n_1+n_2}}) + (1 - \frac{\theta}{2^{1+n_1+n_2}})] + \gamma\theta(1 - \frac{1}{2^N})$$

subject to the managers having the incentives to specialize, i.e, s.t.

(1) CC_{U1} or IC_{U1}

(2) CC_{U2} or IC_{U2}

(3) IR

Now consider two important cases.

F. Case 1: Uniformly high threat of competition in hierarchy.

Let expropriability, Λ , be high and constant with distance from the resource because all managers need significant access in order to function. This would be the case for example, in a firm where

the critical resource is an idea or code that all managers need to know to be productive. The only way to prevent managers from competing is to convince them they cannot be successful in the winner-take-all competition. The relevant constraints in this case are $(CC_{U1}), (CC_{U2})$, and IR.

Let the *Int* operator be the greatest integer less than its argument. Then CC_{U1} implies $n_1^* = \text{Int}[\frac{1}{\Lambda}] = 1$ because $\Lambda > \frac{1}{2}$ by assumption 2. So the entrepreneur will hire only one young manager at date 1, because more will overwhelm her in competition. In period 2, the entrepreneur is confident of her own loyalty as well as this employee's loyalty, so CC_{U2} indicates she can hire $n_2^* = \text{Int}[\frac{2}{\Lambda}] \geq 2$.

At the outset, the entrepreneur would like to set the degree of differentiation, θ , as high as possible so as to extract the most rents from the managers. So the optimal degree of differentiation, θ^* , will be set so that managers are left with enough to just satisfy the IR constraint. Interestingly, the necessity of motivating early employees may set in place a degree of differentiation that will persist long after the original rationale vanishes.

Ignoring integer constraints, the comparative statics are (all proofs are in the Appendix)

PROPOSITION 2. (i) $dn_2^*/d\Lambda \leq 0$; (ii) $d\theta^*/d\Lambda \leq 0$.

The first result is obvious: the hierarchy is kept small when expropriability increases so that the entrepreneur and loyal specialized managers can keep the young unspecialized in check. More interestingly, the reduced size of the date-2 hierarchy will decrease promotional prospects. The entrepreneur will then be forced to give employees a greater share of every unit produced so that they continue to have the incentive to specialize. Thus, when expropriability increases, the entrepreneur is forced to lower the degree of differentiation, and her rents decrease even further.

Proposition 2 throws some light on Saxenian's [1996] comparison between computer industry firms along Route 128 in Massachusetts and in Silicon Valley, California. Route 128 firms tended to be large, vertically integrated organizations with very distinctive cultures. There was very little

lateral movement between these organizations. In Silicon Valley, by contrast, firms were smaller, and possessed a more homogenous culture, with a lot of fraternizing and job hopping.

Saxenian provides a sociological interpretation of these findings, contrasting the buttoned-down, risk-averse East Coast culture with the more casual and risk-loving West. While we do not deny these aspects may have played a role, we can provide an explanation based more on economics. Since in the computer industry the main critical resource is ideas, the degree of expropriability is potentially very high. Computer firms are very much aware of this problem and require their employees to sign non-compete clauses. But California, unlike Massachusetts, has a historical tradition of not enforcing these clauses. This institutional difference represents a rare case of an exogenous difference in expropriability in the same industry. Consistent with the predictions of Proposition 2, in Massachusetts where expropriability is lower, organizations are bigger and show more differentiated technologies and culture – as evidenced by the lower job mobility across firms.

Thus, better enforcement of property rights (lower Λ) can tilt rents towards the entrepreneur – since she does not have to provide as much in incentives to ward off potential breakaway, she can differentiate more to extract surplus from worker-managers. This might appear to vindicate Marx’s position that the enforcement of private ownership claims to capital enables capitalists to “exploit” workers (see Roemer [1988] for a lucid exposition). However, this comparative statics result can be reversed in other circumstances as we will now see.

G. Case 2: High threat of competition in close positions

Consider now a firm where only the managers in direct contact with the resource have sufficient access to appropriate while expropriability falls off dramatically with distance. This is the case, for example, in a firm where client relationships are the key resource. Only the partner dealing with clients has an ability to walk off with them, while the lowly associate who does all the ground work has no contact with clients and little ability to expropriate. In particular, if Λ_2 is very high,

the CC constraint will bind in period 1. This implies that $n_1 = 1$; only one manager will be given access. But because this manager will specialize, he can loyally fill the position in period 2 without competing. Since expropriability is lower for the remaining subordinate positions, the IC constraint rather than the CC constraint is likely to bind. Substituting $n_1 = 1$ in (IC_{U2}) , ignoring the IR constraint for the moment, and solving the maximization problem (see appendix), we get a bang-bang solution. The optimal degree of differentiation, θ^* is 1 if

$$(4) \quad \gamma[4\Lambda_3 - 2][1 - \frac{1}{2^N}] + [4\Lambda_3 - 3] + (2\Lambda_3 - 1)\frac{1}{2^N} \geq 0,$$

and $\theta^* = \underline{\theta}$ otherwise. Given θ^* , n_2^* is the maximum value consistent with (IC_{U2}) .

The entrepreneur's profits increase both in the number of managers n_2 and the extent of differentiation, θ . However, as size increases, the highest young manager will have more subordinates, and therefore more incentive to compete. Similarly, an increase in differentiation not only gives the young manager less incentive to specialize, but also increases his surplus conditional on winning the competition – more differentiation is a double-edged sword since it concentrates rents on the possessor of the critical resource, thus making expropriation more attractive. Since (IC_{U2}) becomes tighter as either θ or n_2 increase, θ^* and n_2^* are strategic substitutes.

Condition (4) simply indicates when the entrepreneur gets more bang for the buck by increasing θ . Since the left hand side of (4) increases in Λ_3 , θ^* increases in Λ_3 , so n_2^* must decrease to keep inequality (IC_{U2}) satisfied. Therefore,

PROPOSITION 3. (i) $d\theta^*/d\Lambda_3 \geq 0$; (ii) $dn_2^*/d\Lambda_3 \leq 0$.

Now with better property rights (lower Λ_3), the profit maximizing entrepreneur will increase the size of the organization, thus increasing the societal pie. But, in order to preserve managerial incentives not to compete in a longer hierarchy, she will have to *reduce* the rents (lower θ)

she extracts off each manager. Thus, stricter enforcement of the entrepreneur’s private property rights can lead to more surplus for worker-managers, and in contrast to the previous case, less “exploitation”.

We ignored the IR constraint in determining the above solution. If IR is satisfied at the optimal, then it will have no effect. If not, the trade-off will move in favor of n_2^* at the expense of θ^* , since a longer date-2 hierarchy will improve anticipated promotional prospects at date 1, while a lower θ will tilt rents towards the manager.

H. Discussion

In the literature there are two main roles for promotion (see Prendergast [1999] for an excellent survey): as a way for the firm to allocate more talented workers to higher positions with greater marginal influence (e.g., Rosen [1982]), or as a reward for performance (e.g., Lazear and Rosen [1981], Rosen [1986]). Talented workers are not necessarily the most senior ones, but seniority may be a useful and objective index of talent that reduces influence activities (Milgrom and Roberts [1990]). In our model, however, we abstract from issues of productivity – all managers are equally productive. Promotion is then simply a way of filling sensitive positions with the loyal.

Interestingly, concern about expropriation enables the hierarchy to commit to rewarding specialization, and we do not have to rely on repeated games to explain why the firm does not renege on promised promotions (see Bull [1987]). In this, our work resembles that of Lazear [1981] or Akerlof and Katz [1989] who argue that deferred compensation paid to older workers also serves as an incentive for younger workers. Our work, however, differs in that compensation emerges from the endogenous organizational structure rather than as part of an optimal pay package. Thus, it does not require any exogenous source of commitment by the firm.

The upward-sloped profile of wages is also consistent with Becker [1975], who points out that a reduced initial salary is a way for young workers to post a bond for the training provided by the

firm. In our model, however, it is the worker who bears the cost of the initial training, and his low wage is simply a reflection of his low bargaining power, while his upward wage profile indicates his growing power in the firm.

In the interests of space, we skip other cases in the three period model that add little additional insight. We now examine the steady state.

III. The Limits to Growth

The artificial end-date prevented us from determining a steady state in the previous section. We remedy this now by reverting to the infinite period world.

We define a steady state equilibrium as a subgame-perfect equilibrium of the game where the state (the number of old specialized managers) is repeated every period.¹⁸ We solve for the steady state equilibrium in Markov strategies assuming that there are no constraints on the initial state, so that the entrepreneur can start out with any number of specialized old managers. If the hierarchy can progress beyond initial growth traps (see later), this will indeed be the ultimate size of the hierarchy.

In a steady state equilibrium, each cohort will have equal size. There will be an equal number of specialized managers above the highest entering young manager, and he can never win in competition against them because $\Lambda < 1$. Therefore, the only limit to the size of the organization is that the IC constraint should hold for the specialized head manager. Let $n^*(\theta) = \lceil \log\{\lceil \theta\Lambda_1(2 - \gamma)/(1 - \gamma) - 1 \rceil / \lceil \Lambda_1(1 + \theta - \gamma)/(1 - \gamma) - 1 - c^R \rceil \} / 2 \log 2$. Then,

PROPOSITION 4. The vertical hierarchy has a cohort size in steady state Markov Perfect equilibrium

of (i) $n^V = \infty$ if $c^R \geq \Lambda_1(2 + \theta - \gamma)/(1 - \gamma) - 1$ (ii) $n^V = n^*$ if $c^R < \Lambda_1(2 + \theta - \gamma)/(1 - \gamma) - 1$.

If the condition in (i) holds, then the IC constraint for the highest old manager always holds regardless of the hierarchy's length. Hence the hierarchy's steady state length is infinity. Otherwise,

the head manager cannot have more than a certain number below him, else he will compete. This is the situation in (ii).

This result has been obtained conditional on a given choice of the level of differentiation. An entrepreneur who wants to maximize the long run steady state level of profit will choose θ to maximize the present discounted value of future profits, i.e., she will solve

$$(5) \quad \max_{\theta} \frac{1}{1-\gamma} \theta \left[1 - \frac{1}{2^{2n^V(\theta)}} \right].$$

We then have

PROPOSITION 5. If $\underline{\theta} \geq \frac{(1-\Lambda_1+c^R)(1-\gamma)}{\Lambda_1}$ then $\theta^* = \underline{\theta}$ and $n^V(\underline{\theta}) = n^*$, otherwise $\theta^* = \frac{(1-\Lambda_1+c^R)(1-\gamma)}{\Lambda_1}$ and $n^V = \infty$.

Intuitively, it is more profitable for the entrepreneur to increase the steady state size of the hierarchy by reducing θ than to squeeze more out of each manager by increasing θ . Therefore, so long as the size of the hierarchy is less than infinity, the entrepreneur will reduce θ to its minimum possible level. Of course, if the potential size of the hierarchy reaches infinity at some level of θ above $\underline{\theta}$, the entrepreneur has no need to reduce θ further. In short, the entrepreneur differentiates the vertical hierarchy minimally.

When $\theta^* = \underline{\theta}$ (and the steady state length of the hierarchy is finite) we can easily derive the comparative statics for steady state size:

COROLLARY 1. An increase in the cost of competition, c^R , a decrease in expropriability, Λ_1 , or a decrease in the probability of survival, γ , increase the steady state length of the hierarchy.

Thus, better protection of property rights, both in the sense of increasing the cost of competing, and reducing expropriability, increases the steady state length of the hierarchy. Moreover, the more likely agents are to survive after retirement, the more they value long-run rents from property

compared to short-run income from employment, and the greater the incentive to break away and compete.

A. Discussion

A persistent result is the adverse effect of an increase in expropriability on the size of the hierarchy whether at an interim date or in steady state. This suggests that when property rights are weak, vertical hierarchies are small because the threat of expropriation does not allow them to grow and production is stunted. As property rights become better enforced, vertical hierarchies are larger and production is higher. When property rights are perfectly enforced ($\Lambda = 0$), however, there is nothing distinctive about the hierarchy in our framework and its size becomes indeterminate.

This provides a testable implication of our model. Countries with poor enforcement of property rights should have small organizations and low per-capita income. Countries with better enforcement of property rights should have larger organizations and higher per-capita income. Finally, countries with an excellent enforcement of property rights have organizations of indeterminate size but also enjoy very high per capita income.

One might also think that the extent of expropriability would vary across industries within a country. Property rights to physical assets are enforced in all but the most rapacious regimes. However, a more sophisticated legal system is needed to enforce property rights to intangible assets such as ideas or client relationships. This suggests a more subtle testable implication: the relative size of firms in industries with intangible assets should increase when the efficiency of the judicial system improves.

In their study of firm size across industries and 15 European countries, Kumar, Rajan and Zingales [1999] test these implications. They find that asset intensive industries where expropriability is likely to be lowest have larger firms. They also find that countries with more efficient judicial systems have larger firms. Furthermore, after correcting for industry and country effects, they find

that the relative size of firms in industries with intangible assets increases when the efficiency of the judicial system improves.

As we have argued, expropriability can also depend on the extent to which competition is allowed. If regulations, or the state, assure a monopoly to the incumbent, then expropriability is small, regardless of the enforcement of property rights. Therefore, state or natural monopolies are likely to have long hierarchies. Not surprisingly, the first long hierarchies in history were the army and the administration, two state monopolies. Interestingly, when property rights tend to be weakly enforced, these tend to be the only organizations of any size.

Finally, our model indicates the dangers for an entrepreneur in attempting to keep too much for herself in the early stages by differentiating the organization extensively. Not only do managers have little incentive to specialize, but also a high degree of differentiation concentrates a large amount of rents in the hands of the one who controls the critical resource making expropriation more attractive to subordinates. Analogously, the centralization of resources in a political dictatorship can undermine the survival of the dictator because it provides incentives for a coup by others. In order to achieve a stable organization, the entrepreneur has to sprinkle rents throughout it. This is precisely what a low degree of differentiation achieves.

IV. The Path to the Steady State

To conclude our analysis of vertical hierarchies, let us ask how the hierarchy will grow to this steady state. As we have seen, specialized managers will not compete when the hierarchy is of smaller length than the steady state size, i.e., on the path to the steady state. Therefore, the entrepreneur only needs to ensure the young managers do not compete. We will focus on the case when the resource is highly expropriable so that the entrepreneur has to ensure each period that she has the backing of enough specialized managers to defeat the highest unspecialized manager.

Under these circumstances, we have already seen that the entrepreneur can give access to only

one unspecialized manager in the first period. Importantly, if the entrepreneur cannot be active in the first period, she will not be able to start the organization since she can only hire unspecialized managers who will break away and dispossess her. Thus, it is the entrepreneur's ability to man the most sensitive positions at the outset, before managerial loyalties are established, that enables her to start an organization with strangers. Moreover, control over access alone will not be enough, since the threat of competition is high, so the organization cannot be started by an arm's length owner. Only when the hierarchy has specialized employees can outside ownership arise.

If the entrepreneur was young to begin with, she can be active for one more period in which case the number hired in the second period satisfies $\Lambda_3 \cdot n_2 \leq 2$. This yields $n_2 = \text{Int}[\frac{2}{\Lambda_3}] \geq 2$. Even though the entrepreneur retires now, the number of young managers that can be hired will grow if $n_2 > 2$.¹⁹ In period t , the number hired will be

$$(6) \quad n_t = \text{Int}\left[\frac{n_{t-1}}{\Lambda_{n_{t-1}+1}}\right].$$

Using (6) and ignoring integer problems, the rate of growth of the hierarchy, which is $[(n_t + n_{t-1}) - (n_{t-1} + n_{t-2})]/[(n_{t-1} + n_{t-2})]$, simplifies to $\frac{1/\Lambda_{n_{t-1}+1} - \Lambda_{n_{t-2}+1}}{1 + \Lambda_{n_{t-2}+1}}$.

It is easily checked that the growth rate increases over time. If expropriability decreases to a constant Λ beyond a certain distance from the resource (as seems natural), the growth rate becomes a constant $\frac{1}{\Lambda} - 1$, independent of size. This is Gibrat's Law, derived not from industry market structure and opportunities (see Sutton [1997] for an excellent review of the literature), but from endogenous internal constraints on firm growth.

A. Growth Traps

Despite the apparent inevitability of growth depicted above, there are situations when a hierarchy is doomed to remain small. If the hierarchy in its initial stages is expected to grow slowly (because

of the high risk of expropriation), it may not provide managers enough future rents to give them an incentive to specialize. As a result, the hierarchy may not even start up. Another way of saying this is that improvements in the enforcement of property rights may draw forth more specialized organizations in an economy.

A hierarchy will also stagnate if at any point it cannot hire more new employees than existing old. For example, $n_2 = 2$ if expropriability is high enough, so only two unspecialized managers in the second period. But the entrepreneur must retire in the third period as must his first specialized manager. This leaves the hierarchy with only two specialized managers in the third period, the same as it had in the second period. The hierarchy is trapped at a level of four managers (two specialized and two unspecialized).

A similar growth trap is encountered if the entrepreneur is old by the time he acquires the critical resource. Even though he can hire one manager initially, the entrepreneur has to retire in the next period, leaving just one active specialized manager again. The hierarchy cannot grow beyond two employees (one old, one young) because the entrepreneur does not remain active long enough to build a sufficiently large cadre to grow the firm.

B. Evading growth traps

The important point to retain is that even though a hierarchy of substantial length may be feasible, actual organizations may never reach that size. But this suggests that temporary shocks can move a hierarchy out of, or into, a growth trap and can have long-term consequences.

For example, a hierarchy can escape a growth trap when an entrepreneur can employ people such as family who will not compete because they have strong social costs of doing so. This may explain why behind the success of many professionally-managed firms lie families who were critical in its initial stages. Hobsbawm [1996, p241] suggests such this was important in the early development of European firms in the period 1848-1875: “The economic advantages of a large

family ...were of course still substantial. Within the business it guaranteed capital, perhaps useful business contacts, and above all reliable managers”.²⁰

Another phenomenon our model may explain is why so many venerable firms today got their start in times of war. During wars, governments typically confer monopolies on a chosen few (to prevent “waste”). Since government-sanctioned monopoly implies low expropriability, firms can grow substantially in such times without fear. But this temporary advantage may be enough to get them over the growth trap so that even after the war ends and expropriability returns to normal, these firms retain their size and growth rates.

Adverse temporary shocks can also have permanent effects by decimating a firm’s stock of organizational capital. Suppose a temporary shock (such as an economic depression or a financial crisis) depresses the prospects of future rents within an organization, and hence the young managers’ incentive to make specific investments. This will lead to a drastic reduction in the number of specialized managers available next period. The number of specialized will have to be rebuilt painstakingly slowing the recovery. The firm’s feasible size as well as its rate of growth could fall considerably. In extreme situations, firms could find themselves back in a growth trap with no way out, even if opportunities returned to their old values.

V. Horizontal Hierarchies

Thus far, we have focused on vertical hierarchies where the primary problem is the incentive to breakaway and compete. One alternative arrangement stands out as a remedy – the horizontal hierarchy where the head has multiple subordinates in the tier immediately below her (see Figure V).

Since none of the managers have subordinates in this hierarchy, the active entrepreneur alone can defeat any manager who competes (and managers cannot form coalitions against her because they are not specialized to each other). So through a process of divide and conquer - preventing

managers from having subordinates that they can take with them – the entrepreneur ensures competition is not a concern.²¹

Of course, to continue with a horizontal hierarchy, on retirement the entrepreneur has to transfer control to a manager. In the last period of the entrepreneur’s active life, his subordinates have all specialized. Since all these subordinates have knowledge of the resource, the entrepreneur can transfer control to one of them.

Let us assume that control is transferred via an auction where the gains from trade are distributed among players according to their Shapley value. If the cohort size is n the Shapley value of a generic specialized manager is $\frac{1}{n(n+1)}S$, where S is the surplus accruing to the coalition of the potential purchasers and seller.²² Since there are n managers who can purchase, the surplus that accrues to the seller is $\frac{n}{n+1}S$. This is also the price P the chosen purchaser pays for control.

The rest of the managers who specialized will have to be fired since they are now old, and do not have the incentive to specialize to the new boss. They will be replaced by young hires, who have better incentives to specialize because they have eyes on the prize of becoming the head next period.

Therefore, the surplus to be split in the auction is the total production next period minus the surplus that goes to next period’s n managers $(1 - \frac{\theta}{2})$ each, plus the price P at which today’s purchaser will sell the resource at the end of the period; i.e.,

$$(7) \quad S = n + 1 - (1 - \frac{\theta}{2})n + P.$$

Substituting $S = \frac{n+1}{n}P$ in (7) and solving for P , we get $P = n[1 + \frac{n\theta}{2}]$. The equilibrium steady state price is much more than the total surplus extracted by the head in a period (of $(1 + \frac{n\theta}{2})$). This is because the price embeds some of the surplus extracted off future cohorts of managers.

The lifetime expected surplus accruing to each young manager is

$$(8) \quad \left[1 - \frac{\theta}{2}\right] + \frac{1}{n(n+1)}S = \left[1 - \frac{\theta}{2}\right] + \frac{1 + \frac{n\theta}{2}}{n} = 1 + \frac{1}{n}.$$

The expected surplus falls in the number of managers because each manager faces more competition in the auction for ownership. Interestingly, the surplus does not depend on the level of differentiation. What a manager loses to the entrepreneur in the first period as a result of a higher θ , he recovers (in expectation) as a head in the second period.

Since the surplus has to exceed c^S for the manager to specialize, it must be that

$$(9) \quad 1 + \frac{1}{n} \geq c^S.$$

The entrepreneur's objective is to maximize profits

$$(10) \quad 1 + \frac{n\theta}{2} + n\left(1 + \frac{n\theta}{2}\right);$$

which is increasing in both θ and n .

Since the constraint (9) does not depend on θ , the entrepreneur will choose $\theta^* = 1$ and set the optimal value n^H so that the IR constraint (9) is just met. Therefore, comparing with the result in proposition 5, we have

PROPOSITION 6. The degree of differentiation in a horizontal hierarchy always weakly exceeds and sometimes strictly exceeds the degree of differentiation in a vertical hierarchy.

This highlights an important difference between a vertical and horizontal organizations. Vertical organizations find it optimal not to differentiate too much because this concentrates rents at the top and limits their potential size. In the horizontal organization, however, the incentive to

expropriate is controlled by the strategy of divide and rule and is unaffected by how much rents are concentrated at the top. Moreover, rents are naturally spread through the process of sale; what is lost by a manager in a horizontal organization in wages is picked up in expected ownership so the manager's individual rationality constraint is unaffected by differentiation. As a result, horizontal organizations differentiate maximally. It follows that the wages of the lowest manager, $(1 - \frac{\theta}{2})$, are lower in a horizontal organization than in the vertical organization, and if the lowest managers are on their IR constraint, the former organization's expected income profile is more steeply sloped.

A. Will Control Be Sold?

Thus far, we have assumed that the entrepreneur will sell control. Given the impossibility of commitment, however, this sale will take place only if the entrepreneur has no better option other than selling. Perhaps, however, she could continue to maintain control.

Apart from the vertical hierarchy, there are potentially other forms of organization that could allow the entrepreneur to maintain control of the resource. For example, on retirement from the horizontal hierarchy, the entrepreneur could place an experienced manager as head and continue to retain control of the resource. As we argue in the appendix, this hybrid (and other hybrid forms) do not provide managers adequate incentives to specialize when $\theta = 1$, and are therefore not feasible.²³

This leaves the entrepreneur only one alternative: to convert the horizontal hierarchy into a vertical hierarchy. It is easy to show that anticipating such a conversion, no manager in the initial horizontal hierarchy has the incentive to specialize. Therefore, for the prospect of selling to be seen as credible ex ante, the entrepreneur should prefer selling control to one manager rather than retaining control through organizational change. While the size of the vertical hierarchy (and hence the prospective rents to the entrepreneur from conversion) decrease in the ease of expropriability (high Λ , low c^R), the size and rents in the horizontal hierarchy are unaffected by these parameters

since expropriability is not a concern. Therefore, conversion is dominated, and the entrepreneur's promise to sell control can be credible only when the ease of expropriability is very high. We have

PROPOSITION 7. Consider a critical resource with parameters $\Lambda \equiv \{\Lambda_1, \dots, \Lambda_n\} = \bar{\Lambda}$, and $c^R = \bar{c}^R$.

If the entrepreneur prefers setting up a horizontal hierarchy at the outset, she also prefers to do so whenever $\Lambda \geq \bar{\Lambda}$ and $c^R \leq \bar{c}^R$.

This proposition implies that there is a threshold of expropriability above which the horizontal hierarchy will always be preferred.

B. Comparison of Horizontal and Vertical Hierarchies

Consider now the difference between a vertical and a horizontal hierarchy. The main incentive problem in a vertical hierarchy is expropriability. This limits the size of the hierarchy. The nature of the hierarchy, however, puts positional power in the hands of managers. So managers are secure in their rents and have the incentive to specialize.

In a horizontal hierarchy, expropriability is dealt with by a process of divide and conquer. No subordinate has enough power to overcome the head. But this gives them very little positional power, and therefore little incentive to specialize. Hence the need of an internal sale of control so as to motivate the incoming cohort to specialize. Moreover, size has to be limited so that each manager has a reasonable chance of getting control. In other words, size is constrained by individual rationality in horizontal hierarchies, while it is constrained by incentive compatibility in vertical hierarchies.

From proposition 7, we should expect horizontal hierarchies to predominate in sectors where expropriability is very high. This accords with casual empiricism. The critical resources in human capital intensive industries are strategies, client lists, and ideas, that are very hard to protect, thus the flat structure of law and consulting firms. Interestingly, these are also the firms where

older managers sell their ownership stakes to select proven younger managers (see Bhide [1996], for example). By contrast, organizations with very steep organizational pyramids typically seem to be found in physical capital intensive industries like automobiles or natural or officially-sanctioned monopolies like utilities and the government. If the property rights to physical capital are easily protected from expropriation, and monopolies protected by definition from competition, these are indeed organizations with a low Λ and our model predicts the observed vertical hierarchies.

Horizontal hierarchies will have stronger cultures or more distinctive technologies of production than vertical hierarchies. This is because the expected rents from ownership are widely spread in a horizontal hierarchy so the entrepreneur can get the initial rents from differentiating the hierarchy without destroying the incentives of managers. As a result, owners “exploit” new employees more in a horizontal hierarchy, but this is compensated in the long run because employees have a chance of becoming owners and “exploiting” succeeding generations of employees.

Vertical and horizontal organizations differ in their treatment of the specialized old. In the vertical hierarchy, they are a valuable resource who can man critical positions, allowing the entrepreneur to expand the hierarchy. In the horizontal hierarchy, apart from the favored few who get ownership, they are detritus to be discarded. Unlike the unspecialized young, the old have no prospect of ownership, and will not make the additional investment needed to make themselves useful. Thus up-or-out policies follow naturally from differences in the nature of the hierarchies.²⁴ An immediate corollary is that vertical hierarchies will take time to reach their potential size because they have to build on past specialization, while horizontal hierarchies can reach their potential quickly.

Finally, the sale of the resource to employees substantially alters incentives in a horizontal hierarchy while it does little in a vertical hierarchy. The reasons are instructive. First, in a horizontal hierarchy the head extracts a significant portion of the human capital that subordinates

contribute. So the critical resource in a horizontal organization is effectively a claim on a large portion of the human capital of future generations of subordinates. Clearly, the prospect of getting some of it can improve investment incentives substantially. By contrast, in a vertical hierarchy, current managers absorb much of the surplus they generate each period, so very little of the human capital of future generations is embodied in the critical resource. Second, in a horizontal hierarchy, once the head is loyal, subordinates have little incentive to compete. By contrast, in a vertical organization, ensuring the loyalty of the head only pushes the problem one step down. Perhaps most important, absentee ownership is not possible in a horizontal hierarchy while it is in a vertical. Therefore, if the sale of control does little to change the size of the vertical hierarchy there is no incentive for the entrepreneur to sell because, given the imperfectly competitive nature of the auction, she always gets more by retaining her stake.

We now discuss what our work implies for the difference between organizations and markets.

VI. Markets and Organizations

What is the difference between hierarchies (organizations) and markets in our model? We follow Coase's [1937] depiction of the hierarchy as an entity where transactions are driven more by authority or power than by prices. In our paper, the entrepreneur controls access to the critical resource as well as access to the specialized human capital of the managers. In this sense, she has some power. But managers are not powerless. In a vertical hierarchy, the specialized get positional power, while in the horizontal hierarchy the specialized get (a chance at) the power from ownership. This permits a distinction between those inside the organization (those who have access) and those outside, as well as between those who have belonged for some time and have some power (the specialized) and those who have just joined.

Many of these distinctions cannot be drawn in the Property Rights approach, the seminal work of Grossman and Hart [1986] and Hart and Moore [1990] (GHM).²⁵ That approach is essentially a

static approach where the ownership of the critical resource and the legal right to make away with it in the future gives the owner power. Since the Property Rights approach does not deem current access necessary in any way (ownership is important only in that it helps control future access), and the model is static, it does not help distinguish between people inside the organization and people outside other than the owner(s).

There is another important distinction in our work. Organizations cannot exist in the Property Rights framework if the legal rights of ownership cannot be enforced. This is why that framework emphasizes physical assets (or alienable intangible ones like patents), as well as a well functioning legal system, to explain organized activity like firms. But then what does one make of consulting or law firms where the assets are largely human capital, or how can one explain the existence of organizations in economies (or sectors) where property rights are not well enforced? This is where the distinction between ownership (which gives the owner the right to determine access now and in the future) and control (which only gives the right to determine current access) is important. As we show, the ability to control access today may be enough to project control rights into the future, so a whole gamut of unique resources that cannot be owned but can only be controlled, even temporarily, offer similar residual rights as does ownership, and allow the emergence of organizations.²⁶ This distinction between the role of ownership and the role of control is very useful if we want to model large corporations where ownership is separated from control.

The idea that the organizational core consists of unique capabilities and resources has become crystallized in the Resource Based View of the firm (Penrose [1959], Wernerfelt [1984], Wernerfelt and Montgomery [1988], Hamel and Prahalad [1990]). While the Resource Based View admits to a greater variety of resources than does the Property Rights View, unlike our model it does not indicate how these resources, especially intangible ones, provide authority around which organizations can be built, or how they internally constrain organizational size.

Finally, the superiority of organizations in our model does not rely on the permanent scarcity of the critical resource. Organizational capabilities are built over time through specialization. In particular, once the vertical hierarchy is built, even if the critical resource is no longer unique, the hierarchy can produce more at that point than can a competing entrepreneur starting with a similar resource even though technology does not constrain specialization or production possibilities. This is because the web of specific investments built into the vertical hierarchy over time cannot be reproduced instantaneously, through legal arrangement or by regulating access. This web could also be termed its organizational capital (Klein [1988]).

VII. Conclusion: Is the Hierarchy a Firm?

At the core of our organization is a unique source of value, the critical resource, and three mechanisms – access, specialization, and ownership – that tie agents to it. These mechanisms confer on the organization a power of fiat which differs from ordinary market contracting. According to Coase [1937], this is what distinguishes organizations from markets. Thus ours could be viewed as an economic definition of an organization’s boundaries.

The legal definition of the firm, however, is primarily in terms of ownership. Thus our notion of organization sometimes differs from the legal definition. For instance, our theory can not only encompass extra-legal organizations like the Mafia, but also networks of “independent” firms such as an automobile manufacturer and its dedicated suppliers. At the same time, entities the law defines as a firm may not fit our definition of an economic organization. For instance, a bond trading group at an investment bank, tied to the bank only through its use of capital (which is easily obtained elsewhere), is not really an integral economic part of the bank, though for all legal and cash flow purposes, it is. The bank’s headquarters has no real power over the group and, for all practical purposes, trades at arm’s length with it.²⁷

In our view, that our notion of economic organization does not fully overlap with the legal definition of the firm is not a weakness of our theory. Instead, it points to inadequacies in the legal definition. For instance, regulations such as anti-trust impose limitations on entities that belong to a firm because it is believed they operate in a concerted manner, different from arm's length interactions in a market. If entities that are not part of the same firm operate in a concerted manner, there is no reason why regulations (and the law) should not treat them as a single firm. By contrast, entities that are currently treated by the law as part of the same firm but operate at arm's length should be treated differently.

By distinguishing between an “economic” definition and a legal definition of the firm, we can also understand the possible adverse efficiency consequences of the current legal definition. For example, the corporate opportunity doctrine restricts the ability of managers to personally take advantage of opportunities that come to them while they are agents of the firm. Thus, there is a strong incentive for organizations to conform to the legal definition of a firm so that they can enjoy the protection the law gives them against opportunistic employees, even if this is not the optimal form of organization from an economic perspective.

As physical assets become less important and give way to human capital, the boundaries of the corporation defined in terms of the ownership of physical assets are becoming less meaningful. It is becoming increasingly difficult to classify new hybrid entities such as EcoNets (keiretsu-type alliances of internet firms) in traditional boxes. A deeper analysis of the economic nature of firms is an necessary first step in understanding and dealing with these new entities. We hope our model represents a small move in this direction.

Appendix

A. Derivation of manager k 's payoff

We prove this by induction.²⁸ When $k = n$, (1) shows that (2) holds. Suppose (2) holds for managers $k + 1, \dots, n - 1$. All we need to show then is that it holds for manager k . It is easy to see that the payoff of the k th manager, M_k , is

$$(11) \quad \pi_k = \frac{1}{2} \left[\left(n - \sum_{j=k+1}^n \pi_j \right) - (k - 1) \right] + \frac{1}{2} (1 - \theta).$$

The first term in brackets is the payoff remaining to be divided after managers below M_k have taken their share, while the second term is the total output without M_k . Substituting from (2) using the induction hypothesis

$$(12) \quad \pi_k = \frac{1}{2} \left[\left(n - \sum_{j=k+1}^n \left(1 - \frac{\theta}{2^{n-j+1}} \right) - (k - 1) \right) + \frac{1}{2} (1 - \theta) \right] = \frac{1}{2} \left[2 - \theta + \sum_{j=k+1}^n \frac{\theta}{2^{n-j+1}} \right].$$

Simplifying, we get (2).

B. To prove that in any position a specialized manager has less incentive to compete than an unspecialized manager

To prove that in any position a specialized manager has less incentive to compete than an unspecialized manager we need to write down the conditional IC constraints (the competition constraints does not differ between the two).

The specialized manager's conditional IC constraint is

$$(13) \quad \left[1 - \frac{\theta}{2^{n-k+1}} \right] \geq \Lambda_k \left[\theta \left(1 - \frac{1}{2^{n-k+1}} \right) + \left(1 - \frac{\theta}{2^{n-k+1}} \right) + \Pi^0(y^k) \right] - c^R.$$

The left hand side of (13) is the payoff the specialized manager receives by working for an extra period (his last one) in the firm. The right hand side is the present value of the payoff he obtains if he competes. When he competes, in the first period he gets a rent for his human capital as the head manager (producing with all his specialized subordinates) as well as a rent for his control over the critical resource. After that he will get the present discounted value of the future rents from the hierarchy because of his control of the resource, $\Lambda_k \Pi^0(y^k)$. Discounted rent Π is indexed by O , the age of the manager when he replicates, because the age of the manager may affect the development of the hierarchy in future periods. Also, future rents are a function of y^k , the number of specialized young managers who follow him. The number of specialized old managers who follow him is irrelevant for future rents because they will all retire this period.

For an unspecialized manager in the k^{th} position, the IC constraint is given by

$$(14) \quad \left[1 - \frac{\theta}{2^{n-k+1}}\right] + \left[1 - \frac{\theta}{2^{n_s - h(n_s, y^0, y^k) + 1}}\right] - c^S \geq \Lambda_k \left[\theta \left(1 - \frac{1}{2^{n-k+1}}\right) + \left(1 - \frac{\theta}{2^{n-k+1}}\right) + \Pi^Y(y^k)\right] - c^R,$$

where n_s is the length of the hierarchy next period if all managers specialize this period, and h is the tier the manager will be placed in next period (which depends on the length of the hierarchy, the number of managers in his cohort, y^0 , the number of unspecialized managers below him, y^k).

Comparing the unspecialized manager's payoff with the specialized manager's one we have:

i) regardless of the placement policy the left hand side of (13) is bigger than the l.h.s. of (14) because of Assumption 1; ii) the right hand side of (14) is bigger than the r.h.s. of (13) because the first terms are equal and $\Pi^Y(y^k)$ is greater than or equal to $\Pi^0(y^k)$ because the young competing manager has the option to decide how to use his own human capital next period unlike an old manager who has to retire.

C. Proof that in case 2 the solution is bang-bang

The maximization problem is

$$(15) \quad \max_{\theta, n_2} \theta(1 - \frac{1}{2^2}) + (1 - \frac{\theta}{2^2}) + [\theta(1 - \frac{1}{2^{2+n_2}}) + (1 - \frac{\theta}{2^{2+n_2}})] + \gamma[\theta(1 - \frac{1}{2^N})]$$

subject to

$$(16) [1 - \frac{\theta}{2^{n_2}}] + [1 - \frac{\theta}{2^N}] - c^S \geq \Lambda_3[\theta(1 - \frac{1}{2^{n_2}}) + (1 - \frac{\theta}{2^{n_2}})] + \Lambda_3[\theta(1 - \frac{1}{2^N}) + (1 - \frac{\theta}{2^N})] - c^R.$$

Let f be the objective function, g the constraint and L the Lagrange multiplier. Since we assumed N is finite, the optimal n_2 is finite and it has to satisfy $f_{n_2} + Lg_{n_2} = 0$. On substituting values, we find

$$(17) \quad \frac{-f_{n_2}}{g_{n_2}} = \frac{1}{4\Lambda_3 - 2},$$

which is a constant. On the other hand,

$$(18) \quad \frac{-f_{\theta}}{g_{\theta}} = \frac{\frac{3}{2} - \frac{2}{2^{2+n}} + \gamma(1 - \frac{1}{2^N})}{\frac{1}{2^n} + \frac{1}{2^N} + \Lambda_3[2 - \frac{2}{2^n} - \frac{2}{2^N}]}.$$

As long as $\frac{-f_{\theta}}{g_{\theta}} > \frac{1}{4\Lambda_3 - 2}$ the objective function increases with θ . The terms containing n simplify and this condition can be written as

$$(19) \quad \gamma[4\Lambda_3 - 2][1 - \frac{1}{2^N}] + [4\Lambda_3 - 3] + (2\Lambda_3 - 1)\frac{1}{2^N} > 0,$$

which corresponds to (4). Interestingly, (4) does not depend on θ . Thus, if (4) holds $\theta^* = 1$.

Otherwise, $\theta^* = \underline{\theta}$.

D. Proofs of Propositions

Proof of Proposition 2

Ignoring integer constraints, $n_2^* = \frac{2}{\Lambda_3}$, thus $\frac{dn_2^*}{d\Lambda_3} \leq 0$. Λ_3 does not enter either the objective function or the IR constraint, thus its effect is only through n_2^* . Since $\frac{dn_2^*}{d\Lambda_3} \leq 0$, the IR constraint requires $\frac{d\theta^*}{d\Lambda_3} \leq 0$.

Proof of Proposition 3

The left hand side of (4) is increasing in Λ , hence, θ^* can flip from $\underline{\theta}$ to $1 \geq \underline{\theta}$. Thus, (i) follows.

The comparative static for n_2^* can be derived off inequality (16). Since the objective function is increasing in θ^* and n_2^* , we know that the constraint is binding. Thus, we can rewrite it as:

$$(2\theta - \frac{\theta}{2n_2}) + [1 - \frac{\theta}{2N}] - c^S - \Lambda_3[\theta(1 - \frac{1}{2n_2}) + (1 - \frac{\theta}{2n_2})] - \Lambda_3[\theta(1 - \frac{1}{2N}) + (1 - \frac{\theta}{2N})] + c^R = 0.$$

For a given θ^* , an increase in Λ_3 reduces the left hand side of (20), thus for the equality to hold n_2^* should drop, since the derivative of (20) with respect to n_2^* is $\frac{\theta^*}{2^2} \log 2(1 - 2\Lambda_3) < 0$. At the same time, an increase in Λ_3 leads to an increase in θ^* because of (i). As before, in order for (20) to hold n_2^* should drop, since the derivative of the left hand side of (20) with respect to θ^* equals $-\frac{1}{2n_2}(1 - \Lambda_3) - \frac{1}{2N}(1 - \Lambda_3) - \Lambda_3(1 - \frac{1}{2n_2}) - \Lambda_3(1 - \frac{1}{2N}) < 0$. Thus, (ii) follows from both the direct and indirect effects of Λ_3 on n_2^* .

Proof of Proposition 4

Let n be the steady state cohort size, then CIC constraint for the highest specialized manager is

$$[1 - \frac{\theta}{2^{2n}}] \geq \Lambda_1[\theta(1 - \frac{1}{2^{2n}}) + (1 - \frac{\theta}{2^{2n}})] + \Lambda_1 \frac{\gamma}{1 - \gamma} \theta(1 - \frac{1}{2^{2n}})] - c^R.$$

This can be rewritten as

$$(21) \quad \frac{\frac{\Lambda_1(1-\gamma+\theta)}{1-\gamma} - 1 - c^R}{\theta[\frac{\Lambda_1(2-\gamma)}{1-\gamma} - 1]} < \frac{1}{2^{2n}}.$$

If $c^R \geq \frac{\Lambda_1(1-\gamma+\theta)}{1-\gamma} - 1$, then the above inequality is always satisfied because under assumption 2 $\frac{\Lambda_1(2-\gamma)}{1-\gamma} - 1 > 0 \forall \gamma \in [0, 1]$ and thus the CIC constraint imposes no limits to the hierarchy. This proves (i).

If $c^R < \frac{\Lambda_1(1-\gamma+\theta)}{1-\gamma} - 1$, then inequality (21) can be rewritten as

$$n \leq \frac{\log\{[\theta\Lambda_1(2-\gamma)/(1-\gamma) - 1]/[\Lambda_1(1+\theta-\gamma)/(1-\gamma) - 1 - c^R]\}}{2\log 2} = n^*.$$

Thus, the CIC is satisfied if $n = n^*$. This proves (ii).

Proof of Proposition 5

Differentiating (5) with respect to θ after substituting n^V we get

$$\frac{\Lambda_1(2-\gamma)}{1-\gamma} - 1 - \frac{\Lambda_1}{1-\gamma} < 0.$$

The entrepreneur's profits decrease in θ but only when size is finite. Once size becomes infinite, i.e., $\theta < \frac{(1-\Lambda_1+c^R)(1-\gamma)}{\Lambda_1}$ so that the head's CIC constraint is satisfied for any value of n , then there is no need to reduce θ further. So if $\underline{\theta} \geq \frac{(1-\Lambda_1+c^R)(1-\gamma)}{\Lambda_1}$ then $\theta^* = \underline{\theta}$ and $n^V = n^*$. Otherwise $\theta^* = \frac{(1-\Lambda_1+c^R)(1-\gamma)}{\Lambda_1}$ and $n^V = \infty$.

Proof of Corollary 1

From the definition of n^* we obtain

$$(22) \quad 2^{2n^*} = \frac{\theta[\Lambda_1(2-\gamma)/(1-\gamma) - 1]}{\Lambda_1(1-\gamma+\theta)/(1-\gamma) - 1 - c^R}.$$

Since the left hand side is a monotone function of n , we can derive the comparative statics on firm's size just differentiating the r.h.s. w.r.t. the parameters.

$$\frac{\partial 2^{2n^*}}{\partial c^R} = \frac{\theta[\Lambda_1(2-\gamma)/(1-\gamma) - 1]}{[\Lambda_1(1-\gamma+\theta)/(1-\gamma) - 1 - c^R]^2} > 0;$$

$$\frac{\partial 2^{2n^*}}{\partial \Lambda_1} = \frac{\theta}{1-\gamma} \frac{[\theta - 1 - c^R(2-\gamma)]}{[\Lambda_1(1-\gamma+\theta)/(1-\gamma) - 1 - c^R]^2} < 0;$$

$$\frac{\partial 2^{2n^*}}{\partial \gamma} = \frac{\theta \Lambda_1 [-(1-\Lambda_1)(1-\theta) - c^R]}{[\Lambda_1(1-\gamma+\theta) - (1-\gamma) - c^R(1-\gamma)]^2} < 0.$$

Why the hybrid is not feasible.

To see why the hybrid is not feasible, let n be the cohort size. Then, the surplus the entrepreneur splits equally with the head is what the head produces plus what they can extract from subordinate managers, totalling $1 + \frac{n}{2}$. Thus, the surplus a young incoming manager expects over his lifetime (including his initial period as a subordinate) is

$$(23) \quad \frac{1}{2} + \frac{1}{n} \left[\frac{1}{2} + \frac{n}{4} \right] = \frac{3}{4} + \frac{1}{2n}.$$

The expected surplus decreases in the number of managers in the cohort because the surplus conditional on becoming head – which has a fixed component – is divided among more candidates. But this creates a problem. From (23), the expected surplus when $n \geq 2$ is less than or equal to 1, which by assumption 1 is less than the cost of specialization, c^S . Since the hybrid requires at least two managers in the second tier (else it is a vertical hierarchy), we have to conclude that it is not feasible when rewards come solely through promotions.

Other hybrids that involve a horizontal organization with vertical “divisions” provides worse incentives than does a pure horizontal organization. A sketch of the intuition follows: Let a head have n_s subordinates in the second tier and let each of these have k subordinates in a vertical line (for a total of $k * n_s + 1$ managers in the hierarchy). Regardless of whether one of the second tier is promoted or purchases control when the head retires, the managers who are subordinates of the remaining $n - 1$ managers in the second tier are fired along with those managers. Now consider the incentives of a manager at the bottom of the hierarchy. He gets $\frac{1}{2}$ in the first period, and has a probability $\frac{1}{n}$ of being promoted by one rung in the second period and getting $\frac{3}{4}$. With probability $\frac{n-1}{n}$, he will be fired. So he gets an expected income of $\frac{1}{2} + \frac{1}{n}\frac{3}{4}$. Since n has to be at least 2 for this to not be a vertical hierarchy, it is immediate that the expected rents are below 1 and hence below c^S .

Proof of Proposition 7

(Sketch) For $\Lambda \rightarrow 1$ and $c^R \rightarrow 0$ the vertical hierarchy is infeasible when the entrepreneur is inactive. Also when the entrepreneur is active the per-period entrepreneurial rent in the horizontal hierarchy is always greater than the per-period rent in the vertical hierarchy. Thus, in the limit the horizontal hierarchy is always preferred.

Therefore, all we need to show is that the entrepreneur’s preference for the vertical hierarchy is decreasing in Λ and increasing in c^R . The incentive constraints in the vertical hierarchy become tighter as Λ increases and c^R decreases, while the incentive constraints for the horizontal hierarchy remain unchanged provided that the entrepreneur can credibly promise to sell. But as Λ increases and c^R decreases the entrepreneur’s incentive to sell also increases. Hence the proposition.

University of Chicago and NBER

University of Chicago, NBER, and CEPR

References

- Aghion, Philippe and Jean Tirole , “Formal and Real Authority in Organizations” *Journal of Political Economy*, CV (1997), 1-27.
- Akerlof, George and Lawrence Katz, “Workers Trust Funds and the Logic of Wage Profiles” *Journal of Political Economy*, 104 (1989), 525-36.
- Alchian, Armen and Harold Demsetz, “Production, Information Costs and Economic Organization”, *American Economic Review*, LXII (1972), 777-705.
- Baker, George, Robert Gibbons, and Kevin J. Murphy, “Relational Contracts and the Theory of the Firm,” Harvard Business School working paper, 1997.
- Becker, Gary, S., *Human Capital* (Chicago, IL: University of Chicago Press, 1975).
- Bhide, Amar V. , “Building the Professional Firm: McKinsey & Co: 1939-1968” , Harvard Business School working paper, 1996.
- Bhide, Amar V. , *The Origin and the Evolution of New Businesses* (New York, NY: Oxford University Press, 2000).
- Brown, Charles, James Hamilton , and James Medoff, *Employers Large and Small*, (Cambridge, MA: Cambridge University Press, 1990).
- Bolton, Patrick and Dewatripont, Mathias, “The Firm as a Communication Network” *Quarterly Journal of Economics* 109 (1994), 809 - 840.
- Bull, Clive , “The Existence of Self-Enforcing Wage Contracts,” *Quarterly Journal of Economics*, 102 (1987), 147-159.
- Burt, Ronald , *Structural Holes* (Cambridge, MA: Harvard University Press, 1992).

- Cheung, Steven, "Property Rights in Trade Secrets," *Economic Inquiry*, 20 (1982), 40-53.
- Coase, Ronald, "The Nature of the Firm," *Economica* 4 (1937), 386-405.
- Garicano, Louis, "A Theory of Knowledge-Base Hierarchies," University of Chicago Working Paper, 1998.
- Granovetter, Mark S., "The Strength of Weak Ties: A Network Theory Revisited," in R. Collins ed, *Sociological Theory* (San Francisco, CA: Jossey-Bass, 1984).
- Grossman, Sanford and Oliver Hart, "The Costs and the Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, XCIV (1986), 691-719.
- Hall, Peter, *Cities in Civilization*, (New York, NY: Pantheon, 1998).
- Hamel, G. and C. Prahalad, "The Core Competence of the Corporation" *Harvard Business Review* 68 (1990), 79-91.
- Hansmann, Henry, *The Ownership of Enterprise*, (Cambridge, MA: Belknap 1996).
- Harhoff, Dietmar and Thomas J. Kane , "Financing Apprenticeship Training: Evidence from Germany," NBER Working Paper # W4557, 1993.
- Hart, Oliver and John Moore, "Property Rights and the Nature of the Firm" *Journal of Political Economy*, XCVIII (1990), 1119-1158.
- Hobsbawm, Eric, *The Age of Capital*, (New York, NY: Vintage Books, 1996).
- Holmstrom, Bengt, "The Firm as a Subeconomy" *Journal of Law, Economics and Organization*, 15 (1999), 74-102.
- Holmstrom, Bengt and Paul Milgrom, "The Firm as an Incentive System" *American Economic Review*, 84 (1999), 972-991.

- Holmstrom, Bengt and J.Roberts, “The Boundaries of the Firm Revisited,” *Journal of Economic Perspectives*, 12 (1998), 73-94.
- Jackson, Tim *Inside Intel; Andrew Grove and the Rise of the World’s Most Powerful Chip Company*, (New York, NY: Dutton Books, 1997).
- Jensen, Michael, C. and Willaim Meckling, “Theory of the Firm: Managerial Behavior, Agency Costs and Capital Structure”, *Journal of Financial Economics* 3 (1976), 305-360.
- Kahn, Charles and Gur Huberman, “Two-Sided Uncertainty and Up- or-out Contracts” *Journal of Labor Economics* 6 (1988), 423-44.
- Klein, Benjamin, “ Vertical Integration as Organizational Ownership: The Fisher Body-General Motors Relationship Revisited,” *Journal of Law, Economics and Organization* XXXI (1988), 199-213.
- Kumar, Krishna, Rajan, Raghuram, R. and Luigi Zingales, “What Determines Firm Size?,” NBER Working Paper #7208, 1999.
- La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny , “Trust in large organizations,” *American Economic Review*, 87 (1997), 333-338.
- Landes, David, “What Do Bosses Really Do?”, *Journal of Economic History*, 46 (1986): 585-623.
- Lazear, Edward P., “Agency, Earnings Profiles, Productivity, and Hours Restrictions,” *American Economic Review* 71 (1981): 606-620.
- Lazear, Edward P. and Sherwin Rosen , “Rank-Order Tournament as Optimum Labor Contracts”, *Journal of Political Economy*, 89 (1981): 841-864.
- Liebesskind, John P. , “Knowledge, Strategy, and The Theory of the Firm,” *Strategic Management Journal* (1996).

- Lucas, Robert E., Jr., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9 (1978): 508-523.
- Mailath, George J. and Andrew Postelwaite, "Workers Versus Firms: Bargaining Over a Firm's Value," *The Review of Economic Studies*, 57 (1990), 369-380.
- Marglin, Stephen, "Knowledge and Power," in Frank H. Stephen, ed. *Firms, Organization and Labour. Approaches to the Economics of Work Organization*, (London: Macmillan, 1984).
- Milgrom, P. and J. Roberts, "Bargaining Costs, Influence Costs, and the Organization of Economic Activity," in *Perspectives on Positive Political Economy*, J. Alt and K. Shepsle eds., (Cambridge, UK: Cambridge University Press, 1990).
- Penrose, Edith T., *Theory of the Growth of the Firm* (New York, NY: J. Wiley & Sons, 1959).
- Prendergast, Canice, "The Provision of Incentives in Firms," *The Journal of Economic Literature* 37 (1999), 7-63.
- Prescott, Edward and Mark Visscher, "Organization Capital," *Journal of Political Economy* 88 (1980), 446-461.
- Rajan, Raghuram R. and Luigi Zingales, "Power in a Theory of the Firm," *Quarterly Journal of Economics* CXIII (1998), 387-432.
- Rebitzer, John B., and L.J. Taylor, "When Knowledge is an Asset: Explaining the Organizational Structure of Large Law Firms," MIT Working Paper, 1997.
- Rodriguez-Palenzuela, Diego "New Ideas Need New Space," Universitat Pompeu Fabra Working Paper, 1999.
- Roemer, John E. "Free to Lose : An Introduction to Marxist Economic Philosophy," (Cambridge, MA: Harvard Univ Press, 1988).

- Rosen, Sherwin, "Authority, Control, and the Distribution of Earnings" *Bell Journal of Economics*, Autumn (1982), 311- 323.
- Rosen, Sherwin, "Prizes and Incentives in Elimination Tournaments," *Journal of Political Economy*, 76 (1986), 701-715.
- Rumelt, R.P., " Theory, Strategy, and Entrepreneurship," in David Teece ed. *The Competitive Challenge* (Cambridge, MA: Ballinger, 1987).
- Saxenian, Annalee , *Regional Advantage*, (Cambridge, MA: Harvard University Press, 1996)
- Stole, Lars, and Jeffrey Zwiebel , "Organizational Design and Technology Choice under Intra-Firm Bargaining," *American Economic Review*, XXCVI (1996), 195-222.
- Sutton, John, "Gibrat's Legacy," *Journal of Economic Literature*, 35 (1997), 40-59.
- Teece, David , "Profiting from Technological Innovation," *Research Policy* 15 (1986), 286-305.
- Wernerfelt, Birger, "A Resource Based View of the Firm," *Strategic Management Journal* 5 (1984), 171-180.
- Wernerfelt, Birger and Cynthia Montgomery, "Diversification, Ricardian Rents, and Tobin's q," *Rand Journal of Economics* 19 (1988), 623-632.
- Williamson, Oliver, *Market and Hierarchies: Analysis and Antitrust Implications*, (New York, NY: The Free Press, 1975).
- Williamson, Oliver, *The Economic Institutions of Capitalism*, (New York, NY: The Free Press, 1985).

Notes

¹There is a large literature that discusses this problem (see, for example, Williamson [1975], Cheung [1982], Landes [1986], Liebeskind [1996], Rumelt [1987], Teece [1986], Mailath and Postelwaite [1990], and Rebitzer and Taylor [1997]). Teece [1986] proposes that a firm mitigates expropriation by owning a set of complementary assets that are critical to production while Rebitzer and Taylor [1997] argue that law firms reward those with the highest threat of expropriation with higher rents.

²After breaking away, each manager produces 0.49. Pietro gets half of what Ram produces for a total of $0.49 * 1.5$ each period.

³In this example we assume that managers have a horizon of only two periods and that the cost of expropriating the technology is zero. Both assumptions are relaxed in the model.

⁴Examples of technology-based theories of the firm include Lucas [1978], Rosen [1982], Bolton and Dewatripont [1994], and Garicano [1998].

⁵While it is fairly uncontroversial that the entrepreneur can offer or refuse access to one manager, it is more questionable whether she has the ability to regulate access further down the hierarchy. We have proven similar results under the assumption that the entrepreneur offers direct access only to one manager, who chooses his subordinate among the pool of managers with access, and so on. Our notion that the entrepreneur allocates access is similar in many ways to the notion in Holmstrom [1999], Holmstrom and Milgrom [1994] (particularly in their focus on the allocation of tasks) and in Holmstrom and Roberts [1998] that the firm is a closed system that can allocate incentives without being pressured by the market.

⁶This is because access is short form for a broad category of memberships, delegated control rights and task assignments that are very hard for the entrepreneur to specify precisely, and for courts to enforce. In the language of Grossman and Hart [1986], the right to control access is a “residual” right of control that emanates from control (though not necessarily ownership) of the resource. However, under some circumstances we do allow the resource to be alienable (see later), and thus the right to control access can be sold.

⁷In an earlier version of the paper, we assumed managers specialized to the technology only, and not to their superior. The results were similar to the ones in this paper. In practice, specialization contains an element of specialization to a technology and specialization to surrounding humans, which we hope to capture with the current assumption.

⁸Setting up as competition should also be viewed as a metaphor for all those activities that allow managers to appropriate part of the value of the firm. Thus, while in large organizations it is rare to see top managers departing with all their subordinates, is quite common to see them use their control of their subordinates to capture more

rents from the firm. Thus the basic trade-off – of increasing access so as to increase productive efficiency, and limiting access to reduce the transfer of power – will continue to play an important role in determining the structure and growth of large organizations.

⁹This cost is more than just the effort cost of collecting the information about the resource, the psychic costs of sneaking it away or the reputational costs of being deemed disloyal. It includes the expected legal costs of any law suits the entrepreneur might file, and the added risks and stress of becoming an entrepreneur. For a capital intensive technology, it could include the dead weight cost of raising finance to fund the new enterprise, a cost which would increase with the degree of imperfections in the capital markets. In a family run firm, it includes the social sanctions the family may impose for defying the family patriarch. Finally, in the spirit of Hansmann [1996] (also see Alchian and Demsetz [1972], Jensen and Meckling [1976], Williamson (1975,1985)) it could be the costs of reproducing interdependent contracts with the members of the production team. Taken together, these costs can be substantial.

¹⁰The manager does not have the resource, so he cannot produce with his subordinates. The reservation wage, which we will describe momentarily, is public information. This ensures that bargaining will be efficient, making it more difficult to achieve a stable firm. This is unlike Mailath and Postelwaite [1990] who show that when workers' reservation value is unobservable, workers will not leave a firm because they cannot agree on how to split the future surplus arising from their joint departure.

¹¹Since access is not contractible, superiors cannot commit to include or exclude the manager next period, neither can the latter commit to work. Moreover, whether or not production takes place this period has no bearing on whether production can take place next period. So next period's possibilities have no effect on the bargaining.

¹²We are certainly not the first to note that positions that are the only "connections" to other positions are a source of power. This is explored in Granovetter [1984] and Burt [1992]

¹³The only important property of the bargaining outcome is that a manager's rent increases with his position. This would be true, for example, if we assume that each manager gets the Shapley value. There are two reasons why we chose not to use the the Shapley value in this context. First, it assumes, somewhat implausibly that each manager has an equal right to make offers independent of his position in the hierarchy. Second, the expression for the manager's rent is not recursive, increasing significantly the computational burden.

¹⁴While in reality, compensation can be committed to over short periods like a year, compensation over longer periods like decades, which approximate the duration of a period in our model, is hard to commit to. Compensation also cannot be made contingent on specialization, which is hard to verify.

¹⁵By assumption 2, $\Lambda_2 (n'_1 + n_2) > 1$ whenever $n'_1 + n_2 \geq 2$. Since $n'_1 \geq 1$ if there is at least one old manager and $n_2 \geq 1$ when at least one manager is hired at date 2 (which is necessary for the hierarchy to survive into the next period), $n'_1 + n_2 \geq 2$ and the manager will defeat the entrepreneur if he competes.

¹⁶Bringing all the terms in (IC_{S2}) to the left hand side and differentiating with respect to number of subordinates we obtain $\frac{\theta}{2^{1+n'_1+n_2-k+1}} \log 2(1 - 2\Lambda_k)$, which is negative by assumption 2.

¹⁷The lowest young manager today is placed over all the young next period, hence his incentive to specialize increases in the number of young next period. The number of young increase between the first and second period because there are trivially more specialized managers at date 2 than at date 1 when there are zero. They increase between the second and third period because N is large. In general, the number of young will always increase in every period except possibly the one in which the steady state is reached.

¹⁸In a steady state with no competition, this implies that the size of the entering cohort is constant. An alternative definition of steady state would be an equilibrium where the size of the hierarchy remains constant, but the size of the entering cohort oscillates (an overlapping generations model allows the hierarchy size to be constant with the entering cohort being the same size only every two periods). It is easy to prove that for every equilibrium where the entering cohort size changes between periods, there is an equilibrium where the cohort size is constant and the hierarchy at least as long. Thus, our focus on steady states with constant entering cohort size is without any loss of generality.

¹⁹Since $\Lambda \leq 1$ and $n_2 > 2$, $n_t \geq n_{t-1} > 2$ for $t > 2$. If $n_2 > 2$, it must be that $\frac{2}{\Lambda_2} > 3$, so that $\Lambda_2 < \frac{2}{3}$. Since $\Lambda_k \leq \Lambda_2$ for $k > 2$, we have on substituting into (6) that $n_t \geq \text{Int}[\frac{3}{2} \cdot n_{t-1}] > n_{t-1}$ since $n_{t-1} > 2$.

²⁰The extent to which one can rely on family and friends also depend on the level of trust prevailing in a society. This might explain the finding of La Porta et al. [1997], that large firms are more important in countries with a higher level of trust.

²¹This resembles Marglin's [1984] idea that the organization is imposed by the capitalist to make himself essential to the production process on a continuing basis. But the horizontal hierarchy in our framework is not so much a tool for exploitation as a device to prevent expropriation.

²²Each manager has a probability $\frac{(n-1)!}{(n+1)!}$ of immediately following the seller and being pivotal to the creation of surplus S .

²³The entrepreneur would continue to choose $\theta = 1$ at the outset if γ is low or the feasible cohort size n is relatively large.

²⁴Kahn and Huberman [1988] offer a different explanation; an up or out policy emerges as an optimal mechanism to

prevent firms from renegeing on compensation to workers. Good workers cannot be labeled “bad” and not promoted because they have to be fired, at some cost to the firm.

²⁵ More recently, however, Hart and Moore [1999] depart from this approach by examining the ex ante allocation of access to tasks, an analysis closer in spirit to ours though with a different focus.

²⁶ While the ability to sell possession is essential to provide incentives in a horizontal hierarchy when $\theta = 1$, the ability derived from law to enforce future exclusion is not an essential feature of the resource. So, for example, a lawyer could “sell” his clients by introducing the purchaser to all of them and then take retirement. Another way of describing the general point is that formal authority (see Aghion and Tirole [1997]) that typically is thought of as deriving from legal constructs like ownership may, in fact, emerge from more informal past processes.

²⁷ In this, our views are similar to Baker, Gibbons and Murphy [1997] who see a continuum of structures between the firm and the market.

²⁸ Thanks to Heitor Almeida for suggesting a more illuminating exposition of this proof.

Figure I

Example of Horizontal Hierarchy

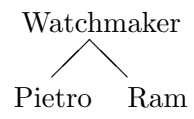


Figure II

Example of Vertical Hierarchy

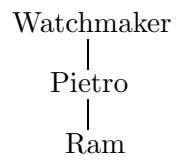


Figure III

Timing

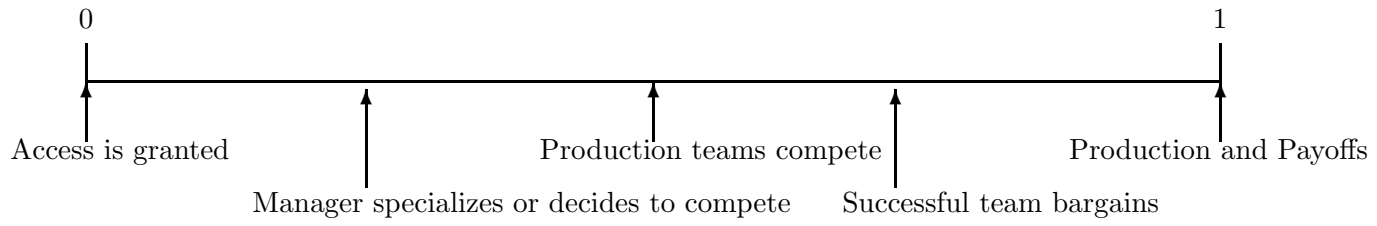


Figure IV

Vertical Hierarchy

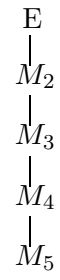


Figure V

Horizontal Hierarchy

