

Teams, Hierarchies, and Innovation

Karen F Bernhardt-Walther*

This version: 11/26/2010

Abstract

I present a model that examines how the optimal organizational design of a firm depends on the complexity and uncertainty of the tasks the firm faces. I show that teams are optimal when firms face high task uncertainty or high task complexity. Hierarchies are optimal if a few tasks reoccur frequently, i.e., if task uncertainty and complexity are low. Since most previous work on organizational design implicitly or explicitly assumes some specific distribution of tasks, analogues of several existing models are subsumed by the model presented here. I argue that high task uncertainty characterizes innovation activity and analyze implications of the model for the organization and management of innovation. Finally, I use the model to study the interactions of organization with technology adoption, use of incentives, and labor markets. I show that some stylized facts about a firm's behavior, such as compensation patterns and technology adoption, can be understood as byproducts of the firm's optimal organizational response.

1 Introduction

Firms address tasks and projects that vary in their complexity and by how frequently similar tasks reoccur. This paper argues that the organizational structure of a firm depends on the complexity and uncertainty of the tasks the firm faces and on the interaction of task complexity and task uncertainty with other drivers of organizational design. I show that through the firm's organizational choices, task complexity and task uncertainty affect indirectly which tasks are solved in an economy, which human capital investments are made, and the equilibrium wage in an economy.

I follow Garicano and Rossi-Hansberg [2006], Kremer [1993], and others, and model a firm as the owner of a task or problem distribution. In order to complete tasks and generate

*The University of Chicago Booth School of Business. I thank Luis Garicano, Lars Stole, Emir Kamenica, and Canice Prendergast for their encouragement and guidance, Robert Akerlof, Joseph Blasi, Alessandro Bonatti, Eric Budish, Florian Ederer, Robert Gibbons, Gustavo Manso, Matthew Notowidigdo, and Bruce Weinberg for valuable discussions. I also thank participants of the Economics Student Luncheon at the Booth School of Business and participants of the AltEcon conference for their comments and feedback. Contact: karen.bernhardt@gmail.com

revenue, the firm has to hire workers who possess human capital in the form of knowledge, skill, or experience. Tasks are characterized by the skills needed to solve them. Different tasks may require different skill sets, and more knowledgeable workers are more expensive for the firm to hire. I assume that the firm knows the distribution of tasks it faces, but that there is an uncertainty about the exact skill required to solve any particular task. This uncertainty is not resolved until the task is successfully completed. The firm can attempt to complete a task more than once. Indeed, in many cases it is optimal for the firm to use multiple attempts: If all tasks were resolved in the first attempt, the skill set used in the first attempt would be larger than necessary for many tasks, resulting in excessive cost.

In this economy, the firm's optimization problem consists of choosing the maximal number of attempts it is willing to conduct and which combination of human capital the workers should have who perform each attempt. A solution to this optimization problem has a natural interpretation as an organizational hierarchy. The number of attempts corresponds to the number of layers in the hierarchy, and the human capital for each attempt characterizes the worker or manager to be hired for that level of the hierarchy. Since workers on each level complete some of the tasks they confront, each level has to attempt fewer problems and hence requires fewer workers than the previous level. The resulting organizational form is directional and pyramidal, i.e., hierarchical. When it is optimal for the firm to attempt all problems only once using a large knowledge set, I assume that the large knowledge set is provided by multiple workers as a team.

The previous literature typically models knowledge as a continuous parameter. This assumption makes the firm's optimization problem tractable, but it requires restricting assumptions on what knowledge workers in the economy have and on the probability distribution of problems. For example, Garicano and Rossi-Hansberg [2006] assume that knowledge of different workers is nested and that the probability distribution decreases exponentially in the complexity of problems. Kremer [1993], Becker and Murphy [1992] assume that knowledge of different workers is disjoint and (implicitly) that the same highly complex problem occurs every period. These restrictions exclude, for example, three-tiered organizations where middle managers have a different expertise than workers, but the principal's skill set encompasses both. Moreover, comparative statics in these models are limited to where the particular distributional assumption is valid, and many distributions are excluded from the analysis.

I deviate from the previous literature and model human capital as discrete units. In the real world, knowledge, skill, or experience are often described using discrete labels: a college degree, being a tax lawyer or a certified accountant, or having twelve years of executive experience - all refer to a particular bundle of human capital. It is these labels that I have in mind when I refer to discrete "units of knowledge."

The main advantage to working with discrete units of knowledge is that no restrictions on feasible knowledge sets or task distributions are needed to make the model tractable. A large variety of organizational forms can arise endogenously, and comparative statics and interaction with environmental factors can be analyzed across different types of task distri-

butions and different organizational forms. For example, I show that task complexity and task uncertainty can explain how technology adoption and organizational change interact differently for different firms.

Among the task distributions not previously considered in the literature are those with high uncertainty about the knowledge needed to solve any particular task. I argue that such task distributions are characteristic for significant innovation. As Bill Buxton observed, “almost anyone who has actually built a product [...] will tell you that they really didn’t know enough to start until they had finished.”¹ The economic literature on innovation has so far focused mostly on innovation as a contractual problem, see Ederer [2008], Holmstrom [1989], Manso [2007]. This model presents a starting point for a complementary “organizing for innovation” analysis.

The organization of firms is an important driver in an economy: It affects the development of new products and services (Bresnahan et al. [2002], Foss [2003]), the employee’s investment in human capital (Prendergast [1993]), the wage distributions in the economy (Garicano and Rossi-Hansberg [2006]), and trading between equally developed economies (Gene M. Grossman [2000]). Among the many determinants of organizational design that have been studied are incentive, monitoring, and moral hazard problems (Holmstrom and Milgrom [1991], Qian [1994]), career concerns (Holmstrom [1999], Gibbons and Waldman [1999]), communication and learning cost (Bolton and Dewatripont [1994], Garicano and Rossi-Hansberg [2006]), and volatility of the firm’s environment (Sutton [2004], Rivkin and N.Sigglekow [2005]). I pursue a complementary approach and study how the organizational design of a firm depends on what the firm does, i.e., how it depends on the distributions of tasks the firm faces.

The remainder of the paper is structured as follows. I formally introduce the model in section 2.1, and solve it completely for an economy with two units of knowledge in section 2.2. I analyze the general model in section 2.3. In particular, I show that teams are optimal if complex tasks, which require a large skill set to be solved, occur frequently, and hierarchies are optimal if some simple tasks reoccur repeatedly. This reproduces the results found in Kremer [1993] and Garicano [2000], respectively. In section 3, I explore task distributions with high task uncertainty and their link to innovation in detail. I show that high task uncertainty, even of relatively simple tasks, results in team formation. I also show that theory predicts that complexity-driven teams behave differently from uncertainty-driven teams, and I present consistent anecdotal evidence. In section 4, I return to the general case where firms may address any task distribution. I use the model as a launching point to study the differential interaction with other drivers of organizational design. In particular, I consider the interaction between organization and technology adoption, use of incentives, and labor markets. Section 5 concludes the paper.

¹“Sketching User Experiences: Getting the design right and the right design”, p.78

2 Optimal Organizational Form

This section presents the theoretical core of the paper. Subsections 2.1 through 2.3 presents the model set-up, a complete solution for an economy with two units of knowledge, and results for the general model. Subsections 2.4 through 2.6 discuss special cases of general model known from the previous literature, empirical predictions of the model, and anecdotal evidence that is consistent with the model.

2.1 The Firm’s Optimization Problem

2.1.1 The Model Set-Up

A firm is the owner of distributions or urns of problems. A firm may own more than one urn, but I assume that all urns owned by one firm have the same problem distribution. Throughout the paper, I use the terms “tasks” and “problems” interchangeably. In each time period, one problem is drawn from each urn and its solution is attempted. Revenue is generated if the problem is solved. The only input into problem solving is human capital, referred to as knowledge. I assume that there are N discrete units of knowledge in this economy, which I label A , B , etc. In practice, bundles of particular knowledge, skill, or expertise are often described by discrete labels such as “two years of college education”, having a “law degree”, or being an expert in “building security”. It is such bundles that I have in mind when discussing “units of knowledge”. The coarseness and interpretation of these units depends on the context of the example.

A firm can access a knowledge set $K \subset \{A, B, \dots\}$ at cost c_K by hiring agents. Any particular knowledge set may be provided by one or more agents, but c_K reflects the total cost to the firm.² Depending on the context, larger knowledge sets may represent a team of workers or a highly skilled worker. However, I will generally speak of a team of workers when the complete knowledge set, i.e., K with $|K| = N$, is required in some organization.³ I assume that more extensive knowledge sets are more expensive, i.e., $K \subset K'$ implies $c_K < c_{K'}$. Unless otherwise stated, I do not impose any other conditions on the relationship between the c_K for different knowledge sets. In particular, the cost function for large knowledge sets can reflect coordination cost or synergy effects. In other words, if K is the disjoint union of K_1 and K_2 , both $c_K > c_{K_1} + c_{K_2}$ and $c_K \leq c_{K_1} + c_{K_2}$ are possible. I assume that the labor market is flexible enough that firms only pay for knowledge sets in those periods where they

²Formally, let $w(K_i)$ denote the cost of hiring an individual worker with knowledge set K_i . Let $coord(\{K_i\}_{i=1}^t, t)$ be the coordination cost incurred when t agents with knowledge sets K_1, \dots, K_t work together; $coord(K_i, 1) = 0$ for all K_i . Then the cost to the firm of accessing knowledge set K is

$$c_K = \min_{t, \{K_i\}_{i=1}^t} \sum_{i=1}^t w(K_i) + coord(\{K_i\}_{i=1}^t, t) \text{ s.t. } K = \bigcup_{i=1}^t K_i$$

³Assuming either bounded rationality or returns to specialization would formally result in large knowledge sets optimally being provided by multiple workers.

access them.⁴

Different problems require different units of knowledge to be solved. To indicate the kind of knowledge a particular problem requires, problems are labeled in the corresponding lower-case letters. For example, a problem “ ab ” can be solved by an agent with the units of knowledge A and B . The interpretation of this description of problems also depends on the context of the example. For example, a particular legal case may need the units of knowledge “tax law” and “court proceedings”, while a particular problem faced by a handyman may require drilling, leveling, and painting. Although the coarseness or the scaling of these labels differs, they are meaningful within their respective contexts. What matters for the model is that those concerned with attempting the problem can agree on and list the knowledge or skills needed to solve the problem ex-post. Solving problem \mathfrak{p} generates revenue $v_{\mathfrak{p}}$. I assume that all problems of the same kind, i.e., those that require the same kind of knowledge to be solved, generate the same revenue.

I put aside any concerns that having more knowledge than is needed to solve a problem can be an obstacle to finding the solution, and assume that for all problems every agent with a knowledge set that contains the minimal necessary knowledge set can solve the problem equally well. I use $K_{\mathfrak{p}}$ to denote the minimal knowledge set that can solve problem \mathfrak{p} , and $\mathcal{P}(K)$ to denote the collection of problems solvable by the knowledge set K . In particular, $\mathfrak{p} \in \mathcal{P}(K)$ indicates that an agent with knowledge set K can solve problem \mathfrak{p} . If an agent finds that he cannot solve a problem, the problem is placed in an auxiliary urns from which other agent(s) can draw problems. Alternatively, unsolved problems can be discarded. A job is an assignment within the firm, characterized by which urn(s) to draw from and in which urn(s) to place unsolved problems.

The firm’s optimization problem is to choose which jobs to create and which knowledge sets to access for each job in order to maximize profit per urn. A solution to this optimization problem is call an “organizational form.”^{5,6}

The timing is as follows.

⁴Alternatively, one can assume that the size of the firm, expressed in the number of urns it owns, is sufficiently large that the (auxiliary) urns from which an agent draws problems are never empty. The purpose of this assumption is that integer constraints about the number of agents involved do not bind.

⁵Each firm is exogenously assigned their urn or distribution. There is no entry. If firms were allowed to trade problems - either from the original distribution or those they cannot solve - and there was free entry, then the price to access a particular distribution would equal the profit derived by a firm in my set-up. The organizational structure would not be affected.

⁶If the problems firm solves does not depend on the firm’s choice of organizational form, then maximizing profit per urn is equivalent to maximizing returns to capital investment. To see this, assume that a firm chooses between two organizational forms OF_1 and OF_2 , that both solve the same collection of problems. Let V denote the total value generated from all problems solved by either organizational form, and let C_1 and C_2 denote the cost of each organizational form per urn. Then

$$\frac{V - C_1}{C_1} > \frac{V - C_2}{C_2} \Leftrightarrow V - C_1 > V - C_2.$$

Time Period	Event
T = 1	Firm learns its problem distribution.
T = 2	Firm chooses which jobs to create and which knowledge to assign to each job.
T = 3	The firm hires appropriate agents and matches them to the jobs. Workers are paid.
T = 4	Production takes place, revenue is generated, and profits are realized.

2.1.2 First-Best Outcome and Problem Triage

The first-best outcome for a firm is to address every problem with the minimal knowledge set that can solve it. Then each problem \mathbf{p} is addressed by an agent with knowledge $K_{\mathbf{p}}$ and the firm generates the profit $v_{\mathbf{p}} - c_{K_{\mathbf{p}}}$. This first best scenario occurs, for example, if the distribution the firm faces consists of one problem that is drawn every period with probability one.

If more than one problem occurs with positive probability, the firm can try to diagnose which knowledge is required to solve each problem before attempting to solve it. I call a mechanism through which such diagnosis is possible “problem triage”. Under triage, every problem \mathbf{p} is delegated to an agent with the corresponding minimal knowledge set $K_{\mathbf{p}}$. Again, first best efficiency is achieved.

I believe that in practice such triage mechanisms exist for many problem distributions. However, triage is often imperfect and costly. For the first part of the paper I assume that the cost of triage is prohibitive. In section 4.1.2, I allow firms to invest in imperfect triage and consider the impact on organizational forms.

2.1.3 The Optimal Organization Problem

If perfect triage of problems is not feasible, then the firm has to specify what knowledge the agents should have who first attempts to solve a problem. An agent with a larger knowledge set is more likely to solve the problem, but is also more expensive to hire. Therefore, in many cases, not every problem is solved in the first attempt. In this case, the firm decides whether unsolved problems should be discarded or whether a second attempt should be made to solve the problem. In the latter case, the firm specifies which knowledge the agents should have who then attempts to solve the problem, and so forth. Our intuition may be that in an optimal arrangement every problem drawn should be treated in the same way. In this case, the firm’s choice can be characterize by a *contingency sequence* of knowledge sets $\{K_1, K_2, \dots, K_l\}$, where knowledge set K_i is used to address the problems not solved during the first $i - 1$ attempts. Any problems not solved by K_l are discarded. This intuition is true, but not obviously so.

The realm of possible arrangements is very large, because only the distribution of problems at the first attempt is given exogenously. Through the choice of an organizational form, the firm chooses the distributions for all auxiliary urns. If some distributions are more valuable to address than others, then it may be optimal for the firm to not treat all problems drawn from the original distribution in the same way. Instead, the firm may accept a worse-than-average outcome for some problems if the benefit for other problems exceeds the loss. In this case, the firm may ask a worker to place a certain fraction of the problems the worker could not solve in one auxiliary urn, the other problems in another urn. The firm may also ask workers with different knowledge sets to address the same distribution or to place problems they could not solve in the same auxiliary urn. Choosing the optimal organizational form has to take all these possibilities into account.

Two features of the model help to reduce the number of potentially optimal organizational forms significantly. First, in this model it is never optimal for a firm to mix distributions. Recall that distributions at the vertices of the distribution space where one problem occurs with probability one are the only distributions that a firm can address and achieve first best. The closer a distribution is to a vertex of the distribution space, the smaller the efficiency loss compared to first-best. Mixing distributions, i.e., forming a convex combination, always increases the efficiency loss. As a consequence, it is never optimal for a firm to create particular distributions so that they can be mixed. Second, if a worker learned something about problems that he cannot solve, then different unsolved problems would be part of different distributions. According to the first observation these unsolved problems should not be “mixed”, and should not be placed in the same auxiliary urn. But such learning would constitute a form of triage. In the absence of a triage mechanism, all problems which could not be solved by the same knowledge sets share the same distribution.

As a consequence, it is indeed optimal for the firm to treat all problems in the same manner. In particular, in an optimal organizational form all problems are attempted by the same first knowledge set. All problems not solved in the first attempt should be addressed by the same knowledge set in the second attempt, and so forth. This result is summarized by the following proposition:

Proposition 1 Characterization of optimal organizational form

If triage is prohibitive, then the optimal knowledge arrangement can be described by a finite sequence of length l of the form (K_1, \dots, K_l) . Moreover, it cannot be optimal for knowledge set to be contained in a knowledge set that occurs at an earlier stage in the contingency sequence. In other words, if $K_i \subset K_j$ for any two knowledge sets in the sequence, then $i < j$.

All proofs are delegated to the appendix. The main purpose of this result is to simplify the search for the optimal organizational structure when solving the firm’s optimization problem. Nonetheless, the proposition provides some insight. First, if we think of problems originating through production or client contact, the result implies that only one kind of worker works in production or with clients. As Garicano [2000] argued, this is consistent with the practice

in many real-world organizations. Second, since every agent in an optimal contingency sequence solves at least some of the problems he faces, fewer agents are required in every consecutive stage. This immediately implies a pyramidal shape familiar from hierarchies. Figure 2 below illustrates this point for $N = 2$. Finally, the optimal contingency sequence is one-dimensional, i.e., all unsolved problems are passed on to the same next knowledge set. This is, of course, due to the absence of triage in this proposition. Given how often real-world organizations depart from this one-dimensionality and forward problems to different agents depending on some outcome or observation, the result indicates how wide-spread and important various forms of triage are.

A firm addressing a problem distribution f with a contingency sequence of knowledge sets $\{K_1, K_2, \dots, K_l\}$ solves all problems that can be solved by at least one knowledge set contained in the sequence. So the revenue generated is $\sum_{\mathbf{p} \text{ s.t. } \exists K_i: \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p})v_{\mathbf{p}}$. The cost incurred by each stage of the contingency sequence equals the share of time that this knowledge set is required. This is equal to the share of problems not solved by a preceding knowledge set. Thus, the profit generated by the firm is given by

$$\Pi(f, \{K_1, K_2, \dots, K_l\}) = \sum_{\mathbf{p} \text{ s.t. } \exists K_i: \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p})v_{\mathbf{p}} - \sum_{i=1}^l \left(1 - \sum_{\mathbf{p} \text{ s.t. } \exists K_j: j < i \text{ and } \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p}) \right) c_K.$$

Note that when keeping the contingency sequence fixed, Π is linear in f .

The optimization problem of a firm facing the problem distribution f is then given by

$$\max_{l, \{K_1, K_2, \dots, K_l\}} \Pi(f, \{K_1, K_2, \dots, K_l\}).$$

I say that a contingency sequence $\{K_1, K_2, \dots, K_l\}$ becomes *more prevalent* due to a change in the environment or an intervention, if the mass of distributions f with regard to a uniform measure increases for which $\{K_1, K_2, \dots, K_l\}$ is optimal.

This is a rich and inclusive model. In particular, I show in section 2.4 that distributions like those considered in Garicano [2000], Kremer [1993], and others are part of the space considered. Where the same organizational questions are addressed, the results presented here coincide with the results in these papers.

2.2 Illustration: Two Units of Knowledge

To illustrate the model, I discuss the case where there are two units of knowledge, labeled A and B . There are then three knowledge sets $\{A\}$, $\{B\}$, and $\{A, B\}$ as well as three potential kinds of problems a , b , and ab . To simplify notation, I write the knowledge sets without brackets.

2.2.1 Distribution Space and Organizational Forms

The distribution of problems a firm faces is described by the probabilities of these problems, denoted by p_a , p_b , and p_{ab} , respectively, and can graphically be represented by the two-dimensional equilateral triangle spanned by $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ in \mathbb{R}^3 as shown in figure 1.

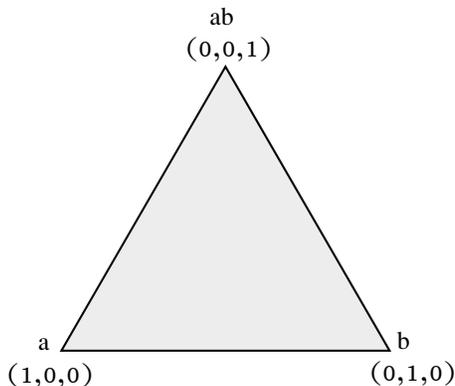


Figure 1: Each point in the triangle spanned by $p_a + p_b + p_{ab} = 1$ in \mathbb{R}^3 corresponds to a problem distribution (p_a, p_b, p_{ab}) . The corners of the triangle correspond to the three (pure) distributions where only one of the three problems occurs. The simplex represents the space of all possible problem distributions.

In this economy, there are three knowledge sets which firms can hire to form contingency sequences. The one unit knowledge sets A and B reflect individual worker. But depending on the context, the knowledge set AB can represent a team, an expert, or a skilled worker. The maximal length of a contingency sequence here is three. Each contingency sequence larger than one naturally corresponds to a hierarchy: Problems are only passed in one direction and there are fewer agents at each stage than in the preceding one. Figure 2 illustrates how knowledge sequences correspond to organizational structures.

Among the many potential possible contingency sequences only ten are potentially optimal for $N = 2$. For example, a contingency sequence where a knowledge set appears later in the sequence than a strictly larger set is never optimal, and is hence excluded. The ten potentially optimal ones are listed in table 1 together with their cost and value functions and the corresponding organizational form.

2.2.2 Optimal Organizational Form

If $v_a, v_b > c_{AB}$, then addressing a distribution with the knowledge set AB generates a positive profit for all distributions. Therefore, production is profitable for each distribution

Knowledge sequence

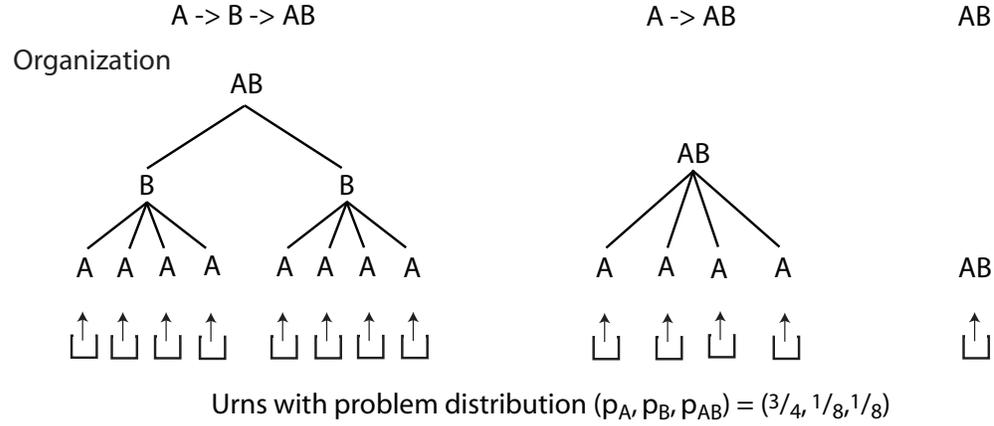


Figure 2: Given a distribution, any knowledge sequence implies a particular organizational structure. This figure shows how the knowledge sequence and the organizational form relate for the distribution $(p_a, p_b, p_{ab}) = (3/4, 1/8, 1/8)$.

Contingency Sequence	Revenue	Cost	Organizational form
no production	0	0	no production
A	$p_a v_a$	c_A	individual worker
B	$p_b v_b$	c_B	individual worker
$A \rightarrow B$	$p_a v_a + p_b v_b$	$c_A + (1 - p_a)c_B$	two-layer hierarchy
$B \rightarrow A$	$p_a v_a + p_b v_b$	$c_B + (1 - p_b)c_A$	two-layer hierarchy
AB	V	c_{AB}	Team
$A \rightarrow AB$	V	$c_A + (1 - p_a)c_{AB}$	flat hierarchy
$B \rightarrow AB$	V	$c_B + (1 - p_b)c_{AB}$	flat hierarchy
$A \rightarrow B \rightarrow AB$	V	$c_A + (1 - p_a)c_B$ $+ (1 - p_a - p_b)c_{AB}$	multi-layer hierarchy
$B \rightarrow A \rightarrow AB$	V	$c_B + (1 - p_b)c_A$ $+ (1 - p_a - p_b)c_{AB}$	multi-layer hierarchy

Table 1: This table shows all ten potentially optimal contingency sequences and their corresponding organizational forms for $N = 2$. All organizational forms in the second half solve all problems and would generate the same revenue $V = p_a v_a + p_b v_b + p_{ab} v_{ab}$ for a problem distribution (p_a, p_b, p_{ab}) .

and all problems get solved for each distribution.⁷ In particular, the optimal organizational form is among the five listed in the lower half of table 1. Comparing the profit generated by these contingency sequences pairwise yields a set of inequalities which characterize the regions in which each of these five contingency sequences is optimal. For example, the contingency sequence AB is optimal if the following four inequalities are satisfied.

$$\begin{aligned} \mathbf{AB} > \mathbf{A} \rightarrow \mathbf{AB} &: p_a < \frac{c_A}{c_{AB}} \\ \mathbf{AB} > \mathbf{B} \rightarrow \mathbf{AB} &: p_b < \frac{c_B}{c_{AB}} \\ \mathbf{AB} > \mathbf{A} \rightarrow \mathbf{B} \rightarrow \mathbf{AB} &: p_a(c_B + c_{AB}) + p_b c_{AB} < c_A + c_B \\ \mathbf{AB} > \mathbf{B} \rightarrow \mathbf{A} \rightarrow \mathbf{AB} &: p_b(c_A + c_{AB}) + p_a c_{AB} < c_A + c_B \end{aligned}$$

Figure 3 a) shows the regions characterized by these inequalities for each of the five contingency sequences. Figure 3 b) shows the corresponding organizational forms that are optimal.

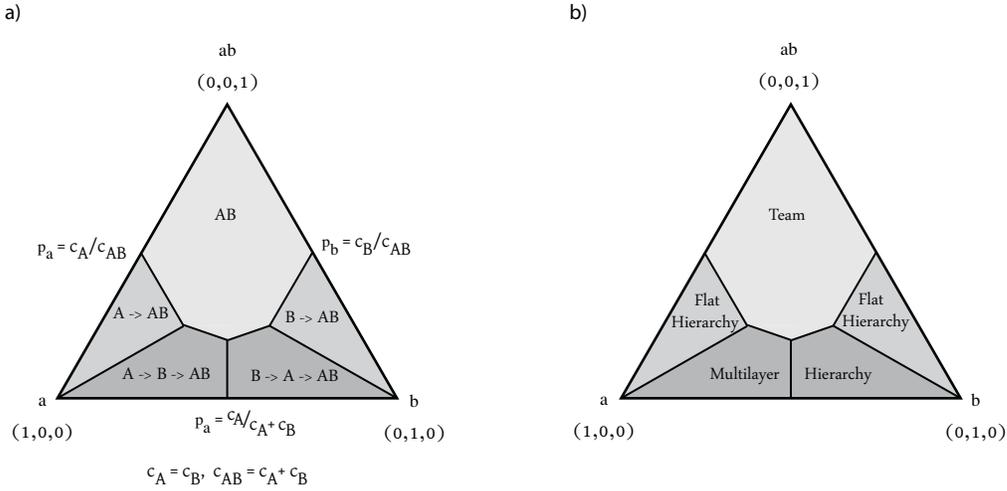


Figure 3: Points in the triangle represent distributions of problems p_a, p_b, p_{ab} . Problems are solved if corresponding knowledge addresses the problem. The firm hires agents with knowledge sets. The firm decides which agents to hire, and in which order agents should attempt to solve problems. This figure assume $v_a, v_b > c_{AB}$, so that all firms solve all problems they face. Figure a) shows which contingency sequence is optimal for which problem distribution. Figure b) shows the corresponding organizational form.

If $v_a < c_{AB}$ or $v_b < c_{AB}$, then it is possible that not all problems get solved for every distribution. In this case, all ten potentially optimal contingency sequences have to be

⁷To see this, assume that there is a distribution for which not all problems get solved. Then there is a remainder distribution out of which no problem gets solved. This remainder distribution, however, corresponds to some point in the distribution space. By assumption, some problem of every distribution has to be solved. Therefore, all problems for all distributions get solved.

considered. Figure 4 shows a general solution where all ten contingency sequences are optimal somewhere. Note that distributions where A is the optimal organizational form correspond to the region on the $b-ab$ line where no production is optimal, because the distribution of problems not solved by A corresponds to a point on the $b-ab$ line.

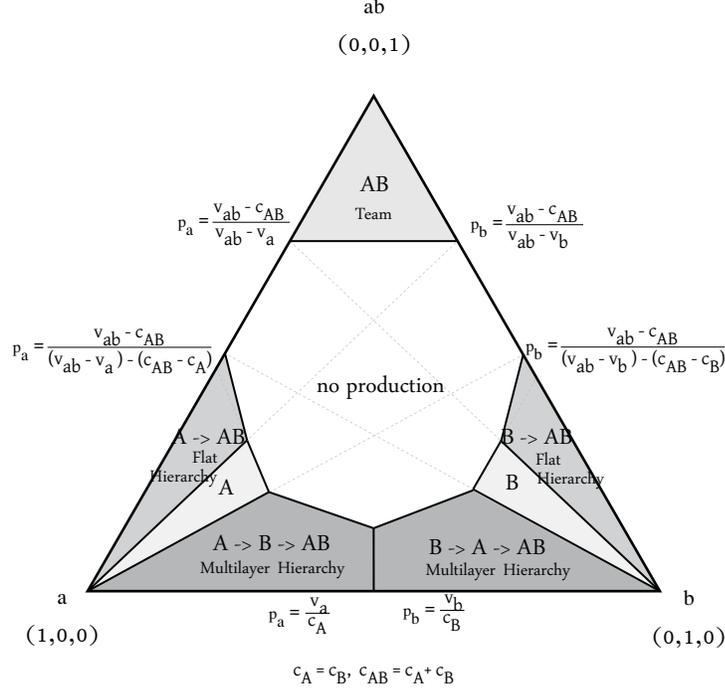


Figure 4: If $v_a, v_b < c_{AB}$, then not all firms solve all problems they face. All ten contingency sequences listed in table 1 are potentially optimal. The optimal contingency sequence for each distribution is inscribed in the corresponding region. Each such sequence corresponds to an optimal organizational form, see figure 2.

The reason that firms facing task distributions in the center of the distribution space do not produce is not that the firm's tasks are not valuable. Indeed, the parameters assume that each of the three problems is valuable, i.e., $v_p - c_{K_p} > 0$, and production occurs at all three vertices. Instead, the problem solutions are not *valuable enough* to overcome the uncertainty about the minimal knowledge required to solve each problem, i.e., through attempting problems more than once or with a larger knowledge set.

2.2.3 Discussion

Although $N = 2$ is a stark simplification, the above results on optimal organizational form warrant three basic observations. First, wage within hierarchies is often found to be increas-

ing with rank (Rosen [1982], Baker et al. [1994]). The literature offers several explanations for this.⁸ This model suggests another, simpler, explanation: If the labor market wage for the higher ranked knowledge set is not larger than the one for the lower ranked knowledge set, then the firm may be better off with a different organizational form. This comparative static is illustrated in figure 5, which shows how the optimal organizational form changes for firms facing different task distributions as the cost c_A decreases relative to c_B .

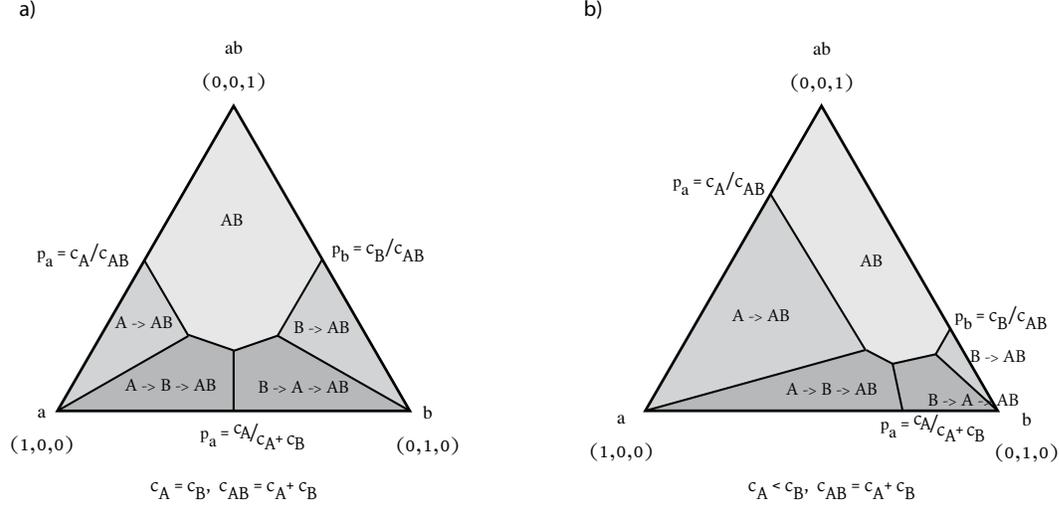


Figure 5: Comparative static: As the relative cost of accessing A versus accessing B decreases, contingency sequences with A as the first stage knowledge set become more prevalent. This offers an explanation why higher-ranked tasks are often found to be better paid in hierarchies: If the labor market wage for the higher ranked knowledge set was not larger than the one for the lower ranked knowledge set, then many firms would be better off with a different organizational form. The figure assumes that $c_{AB} = c_A + c_B$.

Similarly, only if market wages for different knowledge workers are fairly homogeneous, it is optimal for many firms to organize the workers as a team. Assume for a moment that $c_A = 0$. Then a worker with knowledge set A would solve all a -problems for free and only pass on b and ab problems. In this case, almost all firms would be strictly better off by forming a hierarchy with A at the first stage of the knowledge contingency sequence.

Second, Garicano [2000], Garicano and Rossi-Hansberg [2006] observed that distributions in which a few problems occur frequently are common in the real world. Figure 4 offers a hint why this may be the case: Distributions near the vertices are generally the most profitable

⁸For example, Rosen [1982] and Garicano and Rossi-Hansberg [2006] consider recursive production technologies through which superior talent at the top of a hierarchy affects all subordinate layers. Prendergast [1993] shows that a higher wage in a higher rank serves as an incentive for a worker to invest in firm-specific capital when the productivity gain from that investment is large enough to induce a firm to promote the worker after the investment.

to address. If the cost of accessing knowledge is such that not all problem distributions are addressed, then production addressing near-uniform distributions in the center of the distribution space would not be observed. This is in particular the case, if the cost c_A and c_B are market wages and total potential demand exceeds labor supply.

Finally, note that returns to knowledge $V(K) = \sum_{p \in \mathcal{P}(K)} f(p) \cdot v_p - c_K$ are not a meaningful indicator of the optimal organizational form. In particular, all organizational forms that solve all problems generate the same revenue. If it is optimal to solve all problems, then changing the value parameters v_a , v_b , and v_{ab} does not affect the optimal organizational form, but it affects the returns to knowledge $V(K)$.

As an alternative to using the change in revenue to measure returns to knowledge, one may consider using changes in probability that a random problem is solved, $Prob(K) = \sum_{p \in \mathcal{P}(K)} f(p)$. However, this indicator also fails, in particular for convex probability curves. The reason for this is that it fails to take into account the interaction with the knowledge sets at later stages. The probability curve of the conditional distribution $f|_{p \notin K_1}$ of all remainder problems at the second stage is not necessarily convex even if the original probability curve is convex. If the second-stage probability curve is concave, then a multi-stage contingency sequence can be optimal even if the first stage probability curve is convex. In particular, this is true for small values of p_{ab} and relatively balanced values of p_a and p_b . This explains the divergence from the basic concave-convex intuition, as shown in figure 6. The concave-convex intuition fails further if marginal cost of knowledge is not constant.

Although the above observations generalize to $N > 2$, the results shown in the above figures should be taken with a grain of salt, given the oversimplification that $N = 2$ represents. Indeed, a few observations from the $N = 2$ case do not generalize to $N > 2$. For example, for $N = 2$ the number of units of knowledge in each stage is weakly increasing in all potentially optimal organizational forms. This is not true for general N .⁹

2.3 General Results

The model with general $N > 2$ is, of course, less tractable. Nonetheless, the optimal organizational form can be characterized to some extent and completely described in representative special cases.

It is useful to first characterize economy-wide outcomes, before establishing criteria when an organizational form is optimal for a particular distribution.

Proposition 2 The region in the space of problem distributions where one particular organizational form is optimal is convex.

⁹For example, if ab and c are the only problems with positive probability of $p_{ab} = .9$ and $p_c = .1$ and the marginal cost of accessing knowledge is constant, then $AB \rightarrow C$ is the optimal organizational form. In particular, the cost of using the two units A and B for .1 fraction of the time when problem c occurs is less than the cost of using the unit C for .9 fraction of time when problem ab occurs. So $AB \rightarrow C$ outperforms $C \rightarrow AB$.

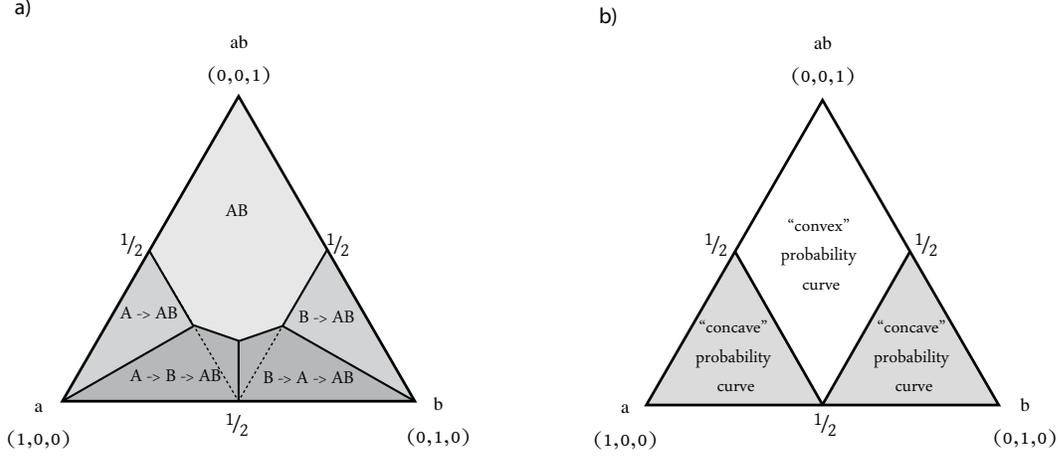


Figure 6: The left figure shows the prevalence of optimal organizational forms assuming that all problems are solved for all distributions and the marginal rate of knowledge is constant, i.e., $c_{AB} = 2c_A = 2c_B \leq v_A, v_B$. The right figure shows the regions in which the probability curve is “convex” and “concave”, respectively. Being limited to two units of knowledge, the probability curve is “concave” if the first unit of knowledge contributes more toward the likelihood of a problem getting solved than the second unit. This is the case if and only if the problems that can be solved by the first unit occur with probability greater than one half.

This result holds in general for any cost function c_K and any value function v_p . The intuition behind proposition 2 is driven by the linearity of the profit function for any organizational form. Because of this linearity, inequalities between profit functions under different organizational forms are preserved under convex combinations of probabilities. As a consequence of this result, it suffices to show that a particular organizational form is optimal for a small number of distributions to establish that the organizational form is optimal in the convex hull of the distributions.

The next result provides a sufficient criterion for when hierarchies arise.

Proposition 3 Sufficient Criterion for Optimality of Hierarchies

Assume that the marginal cost of knowledge is constant. If there exists a knowledge set K such that

$$\text{Prob}(\mathbf{p} \in \mathcal{P}(K)) > \frac{|K|}{N}, \quad (1)$$

then a team with complete knowledge is not optimal. In particular, if the firm solves all problems, a multistage contingency sequence is optimal.

A less than complete knowledge set at the first stage of the organization has the benefit of solving some problems at lower cost. The cost, however, consists in some problems being attempted more than once. Condition (1) implies that the probability of solving a problem is larger than the share of knowledge $|K|/N$, so the savings from not using complete knowledge exceed the loss due to attempting some problems more than once.

I showed in the discussion 2.2.3 above, that already for $N = 2$ the converse result does not hold, because the probability of solving a problem at the second stage of a contingency sequence can be concave even if the original probability curve is convex.¹⁰

The next result establishes an alternative criterion. To formulate it, I introduce the notion of task complexity. I say a task \mathbf{p} is of *complexity* n if the minimal knowledge set $K_{\mathbf{p}}$ that can solve \mathbf{p} has cardinality n , i.e., \mathbf{p} requires exactly n units of knowledge to be solved. Moreover, a task \mathbf{p}' is *strictly more complex* than \mathbf{p} if the minimal knowledge set required to solve \mathbf{p}' strictly contains the minimal knowledge set required to solve \mathbf{p} , i.e., $K_{\mathbf{p}} \subsetneq K_{\mathbf{p}'}$. Note that “strictly more” refers to the knowledge set not to its cardinality. For example, the problem $\mathbf{p}' = abc$ is strictly more complex than $\mathbf{p} = ab$, however no comparison is made with $\hat{\mathbf{p}} = bd$.

Theorem 1 Sufficient Criterion for Optimality of Teams

Assume that the marginal cost of knowledge is constant and that a firm solves all problems it faces. For any collection of problems \mathcal{S} let \mathcal{S}^X denote the collection of problems that are strictly more complex than some problem in \mathcal{S} . Then a team with complete knowledge is the optimal organizational form, if for every knowledge set K and every subset $\mathcal{S} \subset \mathcal{P}(K)$ of problems solvable by K

$$\frac{Prob(\mathbf{p} \in \mathcal{S})}{Prob(\mathbf{p} \in \mathcal{S}^X \setminus \mathcal{S})} \leq \frac{|K|}{(N - |K|)}, \tag{2}$$

that is, if the share of problems solvable by K relative to the correspondingly more complex problems not solvable by K is small relative to the number of units in K .

What stands out about theorem 1 is that the inequality (2) has to hold for all collections \mathcal{S} of problems solvable by the knowledge set K . The reason for this is that at a late stage in the contingency sequence some problems have already been solved. At a late stage in a contingency sequence, it is thus not clear which problems a worker with knowledge set K solves and which ones the worker passes on without specifying the contingency sequence. To require that (2) holds for all collections \mathcal{S} means that the inequality is satisfied for all contingency sequences. Indeed, the claim of the theorem fails if inequality (2) only holds

¹⁰A weaker version would require convexity of the conditional probability curve along the optimal path of learning at every stage of every non-pathological contingency sequence. This weaker version is true in some sense. But it is difficult to make its meaning precise: There always exists a contingency sequence such that adding a unit of knowledge at some stage does not increase the likelihood that a random problem being addressed at this stage is solved. More importantly, it is difficult to test every conceivable conditional probability curve for convexity.

for selected \mathcal{S} , such as the collection that contains all problems solvable by K or only the most complex problem solvable by K .¹¹

The intuition behind theorem 1 reflects the trade-off between wasting time of a worker with insufficient knowledge to solve many problems and wasting time of an agent with a more extensive knowledge set. This can be seen by resolving the fractions on both sides of the inequality through cross-multiplication. Then the former is described by the right-hand side of the inequality: A worker with knowledge set K is by construction not able to solve any strictly more complex problems than those in \mathcal{S} . The latter is reflected in the left-hand side of the inequality: at least $N - k$ units of knowledge are not used for problems in \mathcal{S} . If the left hand side is less than the right hand side then it is cheaper to approach all problems with all N units of knowledge. Note that the marginal cost of knowledge is constant, and therefore cancels on both sides of the inequality.

Although inequality (2) has to hold for all collections \mathcal{S} of problems solvable by any knowledge, the result applies to a wide range of distributions. A firm that faces the most complex problem with probability one satisfies the criterion, as does a firm that faces every possible problem with uniform non-zero probability. So teams with complete knowledge are the optimal organizational form in both cases. Proposition 2 then implies that all convex distributions of these two examples are also best addressed by an all-knowing team. Indeed, theorem 1 and proposition 2 together imply that if firms solve all problems they face, then the region in which a team with complete knowledge is the optimal organizational form has always positive measure.¹²

The criterion of theorem 1 is sufficient but not necessary. Many distributions that are

¹¹For example, consider the distribution where only the problems a , b , ab , and $abcd$ occur with positive probabilities of $p_a = p_b = 1/5$, $p_{ab} = 1/4$ and $p_{abcd} = 7/20$. Then for the knowledge sets A and B condition (2) is weakly satisfied:

$$\frac{\frac{1}{5}}{\frac{1}{4} + \frac{7}{20}} = \frac{1}{3} \leq \frac{1}{4-1}.$$

Moreover, (2) is satisfied for ab when considering either the most complex problem only or all problems solvable:

$$\begin{aligned} \frac{\frac{1}{4}}{\frac{7}{20}} &= \frac{5}{7} < \frac{2}{4-2} \\ \frac{\frac{2}{5} + \frac{1}{4}}{1} &= \frac{13}{20} < \frac{2}{4-2} \end{aligned}$$

However, a team is still not optimal. For $\mathcal{S} = \{a, ab\}$, the knowledge set AB does not satisfy inequality (2)

$$\frac{\frac{1}{5} + \frac{1}{4}}{\frac{7}{20}} = \frac{9}{7} > \frac{2}{4-2}.$$

Indeed, the optimal organizational structure is $AB \rightarrow ABCD$.

¹²To show this, it suffices to find $2^N - 1$ convex independent distributions for which teams are the optimal organizational form. By proposition 2, teams are also optimal for all distributions in the convex hull of these distributions. Hence, the collection of all distributions for which teams are optimal are of positive measure. The $2^N - 1$ convex independent distributions can be constructed as follows. For each problem p of complexity n , let g_p denote the distribution that places weight $n/2N$ on p and weight $2N - n/2N$ on the problem of complexity N . Then by theorem 1, all $2^N - 1$ distributions g_p are optimally addressed by a team with complete knowledge.

optimally addressed by an all-knowing single-stage contingency sequence do not satisfy the criterion. For example, the distribution in which each of the N problems of complexity $N - 1$ occur with equal probability is optimally addressed by a single-stage contingency sequence whenever $N > 2$.¹³ But the inequality of proposition 1 is not satisfied because there does not exist a strictly more complex problem that occurs with positive probability for any of the N problems.

To summarize, figure 7 shows how theorem 1 and propositions 2 and 1 apply to the $N = 2$ economy. In particular, comparing figure 7 with figure 3 shows that both criteria (1) and (2) are applicable to a wide range of distributions, but that neither of them is necessary.

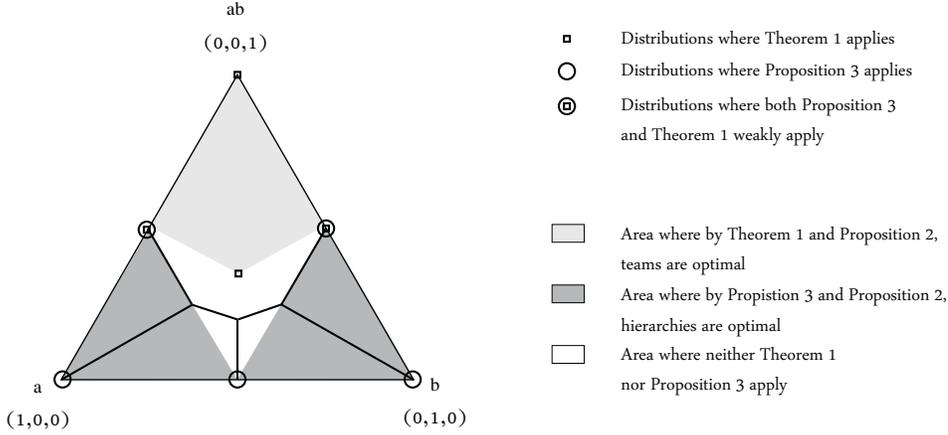


Figure 7: The criteria for optimality of hierarchies and teams, respectively, are widely applicable in the $N = 2$ economy. The boxed and circled distributions represent distributions where theorem 1 and proposition 2 are easily verifiable. The shaded areas represent the respective convex hulls of these distributions, where according to proposition 2 teams and hierarchies are optimal. The white area shows where neither the sufficient criterion for hierarchy formation nor the sufficient criterion for team formation are satisfied.

¹³Formally, the only knowledge sets that can occur in the optimal contingency sequence are of cardinality $N - 1$ or N . The former solve one problem each, the latter solves all problems. If there is only one stage with N units of knowledge, then the cost of solving all problems is $N \cdot c$. If there is more than one stage, then at every stage there have to be at least $N - 1$ units of knowledge. The total cost for such a contingency sequence are hence equal or greater than

$$c \cdot \left[(N - 1) + \frac{N - 1}{N} \cdot (N - 1) \right] = c \cdot \left[N + \frac{(N - 1)^2}{N} - 1 \right].$$

But $\frac{(N-1)^2}{N} > 1$ for all $N > 2$. Hence, the cost for any feasible contingency sequence of length larger than one exceeds the cost $c \cdot N$ of a single stage contingency sequence.

2.4 Special Cases and Relation to Previous Literature

As an application of the above results, I want to consider in turn three special distributions, two of which are closely related to known examples from the literature.

2.4.1 Hierarchy: When Complex Problems are Rare

In many real world examples of problem solving some problems occur much more frequently than others. Often, the re-occurring problems are also the simpler ones. In these cases, proposition 3 implies that a hierarchy represented by a multi-stage contingency sequence is the optimal organizational form. This is consistent with the many hierarchies observed in practice, for example, in an urgent care setting where doctors address only those patients who could not be helped by a registered nurse, in a law firm, where senior partners only get involved if a case is unusual or complex enough that a junior partner cannot handle it, or in manufacturing where managers deal with “exceptions,” not routine details.

Such distributions are the focus of Garicano [2000]. This influential paper treats knowledge and problems as continuous variables and focuses on how organizations are shaped by the trade-off between the cost of communication between layers of a hierarchy and the cost of learning which enables workers to solve more problems. In contrast, the focus of this paper is to show how the optimal organizational form varies with the distribution of problems a firm faces. In particular, Garicano [2000] does not distinguish between the complexity of problems and the uncertainty about problems.

Setting these differences aside, the three settings considered by Garicano [2000] can be translated into this model’s notation. The settings are (1) distributions where problems of the form a, b, c , and so forth occur with (exponentially) decreasing probability, (2) distributions where non-overlapping problems of increasing difficulty, e.g., a, bc, cde , and so on, occur with uniform probability, and (3) distributions with strictly more complex problems, i.e., a, ab, abc , etc., occur with (exponentially) decreasing frequency. For these distributions, Garicano [2000] shows that the optimal organization of knowledge is characterized by a pyramidal hierarchy in which only one kind of worker draws problems and all other workers solve problems others could not. This coincides with the results in this paper. Garicano [2000] also shows that within the pyramidal hierarchy, more frequent or simpler problems are solved first. These results are driven by the paper’s particular distributional assumptions. In particular, theorem 1 implies that a team with complete knowledge is optimal for a distribution where problems of the form a, ab, abc , and so on occur with uniform probability $1/N$.

2.4.2 Team: When Complexity induces Strong Complementarity

Another common scenario is a complex problem that reoccurs with high probability. Let a, b, c, \dots, z_n denote the components necessary to solve a problem of complexity n and

assume that the problem $abc\dots z_n$ is drawn with probability one. Naturally, a single-stage contingency sequence with knowledge set $K = \{A, B, C, \dots, Z_n\}$ is optimal and achieves first-best. If n is large, K is provided by a team. Such team formation has been previously studied in the literature. For example, Becker and Murphy [1992] study the trade-off between coordination and specialization in team formation. In this model’s language, assume that each unit of knowledge i can be provided with a different quality level q_i , so that the overall value generated takes the form $v_{abc\dots z_n} \cdot (\min_i q_i)$ or $v_{abc\dots z_n} \cdot \prod_i q_i$. Becker and Murphy [1992] consider homogeneous agents who can provide multiple units of knowledge, but the returns to specialization or the quality of each unit provided, i.e., q_i , decreases the more units an agent provides. Conversely, the coordination cost within a team increases with the number of agents in the team. This implies a trade-off for the team formation between coordination cost and returns to specialization, which is the focus of Becker and Murphy [1992]. In contrast, Kremer [1993] assumes that every agent can only provide one unit of knowledge but that agents are heterogeneous in the quality q with which they provide their unit of knowledge. He discusses how agents should be matched to form teams. The model presented here takes a more basic approach by only allowing $q \in \{0, 1\}$.

2.4.3 Team: When Uncertainty induces Strong Complementarity

A less studied case is a firm who faces a high uncertainty about the likely (minimal) knowledge set that can solve a problem. This describes a firm that faces an unexpectedly volatile environment, or which is figuratively or literally exploring uncharted territory. Formally, this can be represented by the uniform distribution over all problems, i.e., assume that all $2^N - 1$ problems occur with equal likelihood. Then theorem 1 above implies that an all-knowing single-stage contingency sequence is optimal. In particular, for large N , a team is the optimal organizational form. Most of the team formation results mentioned in the previous example extend from complexity-driven teams to such uncertainty-driven teams. All other things equal, however, investments in or returns to specialization are lower for uncertainty-driven teams, since each unit of knowledge is used with probability roughly $1/2$ rather than one.¹⁴ Differences in investment behavior between complexity and uncertainty-driven teams are discussed in more detail in section 3.1.

Note that the team formation in response to a distribution with high task uncertainty cannot be deduced from knowing the optimal organizational forms for the previous two cases, not even for $N = 2$. This is illustrated in figure 8 a) and b). Knowing the optimal organizational forms for the previous two cases is equivalent to knowing the optimal organizational forms along the edges of the distribution space for $N = 2$, as shown in a). An “educated guess” about the optimal organizational form in the interior of the distribution space is shown in

¹⁴To be precise, the probability that any particular unit of knowledge is needed is given by

$$1 - \frac{2^{N-1} - 1}{2^N - 1} \approx 1 - \frac{1}{2} = \frac{1}{2}.$$

b). The “guess” uses symmetry between A and B , and projection onto the edges.¹⁵ The main difference between the guess and the actual optimal organizational forms is exactly the region with high uncertainty about the minimal knowledge needed.

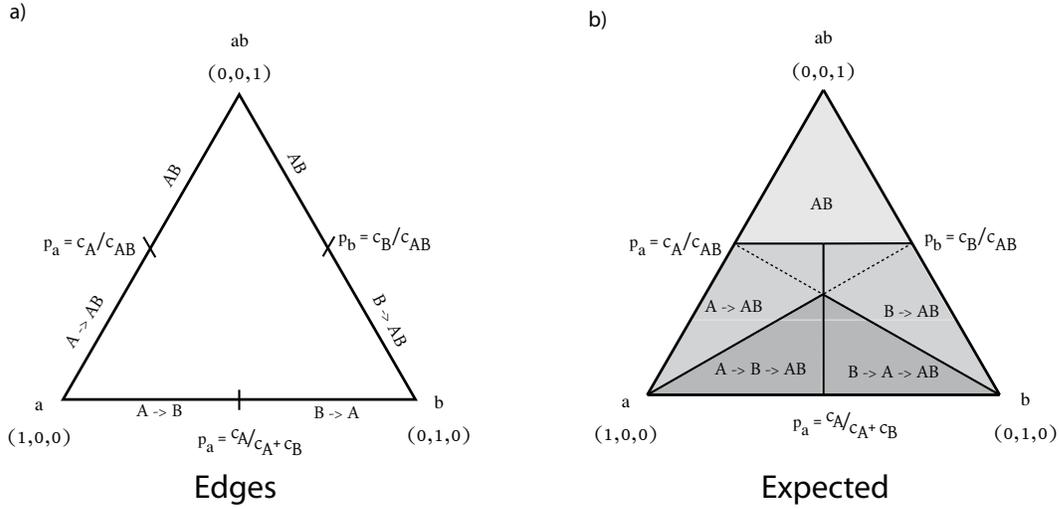


Figure 8: Optimal organizational form in the interior of the distribution space cannot be “guessed” from knowing the optimal organizational form along the edges. Figure a) show the optimal organizational forms along the edges. These are consistent with hierarchy formation discussed by Garicano [2000], and with team formation analyzed by Kremer [1993]. Figure b) shows regions in which different organizational forms are optimal if the results from the edges extended to the entire distribution space, using the symmetry between A and B and projection onto the edges. The expected outcome differs from the actual result shown in figure 3 mainly in the center region where there is a high task uncertainty. Both graphs assume that all problems get solved and that A and B are symmetric in cost and revenue value.

2.5 Discussion: Uncertainty in the Environment

Uncertainty, as considered in this model, is about which knowledge set is necessary to solve a problem. This uncertainty has different empirical implications from the uncertainty typically considered in the literature, namely, the uncertainty about the outcome after a particular effort has been exerted.

Consider a construction project. There may be considerable uncertainty about the geology

¹⁵Projection onto the edges means that if $\{A\}$ is the first knowledge set in an optimal sequence, then the distribution of the auxiliary urn in which the agent with knowledge $\{A\}$ places all unsolved problems must correspond to a point on the edge between B and AB . The organizational form optimal for addressing the auxiliary urn must be the same as the one that faces the distribution at that point on the edge.

of the construction site, which may require hard-to-predict changes to the design in the process of the construction. For example, the design for the Getty Center Art Museum in Los Angeles had to repeatedly be adapted due to a challenging geology that included canyons, slide planes, and earthquake fault lines, (Bajari and Tadelis [2001]). Alternatively, the value created by the completed construction may be highly uncertain. For example, the value of a new apartment building may be uncertain.

A higher uncertainty about the required input means that more architects, engineers, and geologists are more likely to be involved in the project. In contrast, a higher uncertainty about the value generated does not affect the organization of the construction.

The intuition of this example extends to this model. Formally, assume that once a problem is solved, the generated revenue is $\varrho \cdot v_{\mathbf{p}}$, where $\varrho \sim F[0, 2]$ is a random variable. I assume $E[\varrho] = 1$, so that the expected revenue from solving problem \mathbf{p} is unchanged. The smaller the variance of ϱ is, the more predictable is the environment.

Proposition 4 Assume that the marginal cost of knowledge is constant. Assume that firms are risk-neutral and that all distributions are profitably addressed. Then

1. An increase in the uncertainty about the necessary knowledge set increases the prevalence of teams.
2. An increase in the uncertainty about the value generated by a particular effort does not affect the optimal organization of a firm.

If the uncertainty about the knowledge necessary to solve a problem increases, then the underlying task distribution of the firm must have shifted toward the center of the distribution space. If all firms experience a similar shift, the model implies that the prevalence of teams increases. In contrast, an increase in the uncertainty about the value does not affect optimal organization forms. Technically, both parts of proposition 4 are straight-forward. The insight of this proposition comes from the contrasting empirical predictions.

2.6 Application to Organizational Design

Complexity and uncertainty as drivers of organizational decisions have been studied empirically. For example, Bajari and Tadelis [2001] show that procurement contracts for construction projects are more likely to be “cost plus” rather than “fixed price” if there is a higher uncertainty about whether and how the design of the project may have to be changed in the process. Masten [1984] studies make-or-buy decisions in aerospace production, and finds that more complex items are more likely to be produced internally.

Organizational change and team management practices have also been analyzed, for example in relationship with retention and job satisfaction (Carol S. Weisman [1993]), labor productivity (Cappelli and Neumark [2001]), and economic costs and benefits (Batt [2001])

as well as how such benefits are determined by firm characteristics such as worker heterogeneity, firm size, union membership, industry, etc. (Black and Lynch [2001], DeVaro and Kurtulus [2006], Hamilton et al. [2003]).

Unfortunately, team management practices and task complexity or uncertainty have rarely been aspects of the same study. This author is aware of only one work that records both task complexity and organizational change.¹⁶ Brent Boning [2007] studies the adoption of group incentive pay and problem-solving teams in U.S. mini mills. They find that:

Observation 1 Teams are more beneficial in complex and uncertain environments than in simple production processes.

They also find that group incentives are almost uniformly adopted in their sample, because the only measurable output is the production line's final output. However, problem-solving teams are only beneficial to production when the steel bar output is complex.¹⁷ If the output is not a complex product, "following standard operating procedures appears to suffice."

There is, however, anecdotal evidence that:

Observation 2 A firm's successful response to an increase in the uncertainty it faces frequently involves a change in organizational structure toward team work.

For example, at the end of the 1970s the Ford Motor Company faced an uncertain future in a changing environment. Ford was losing owner loyalty, the share of import cars was rising, and the preferences of American customers were changing. There was a high uncertainty about what kind of car could help Ford gain market share and renewed enthusiasm from customers. Under pressure, Ford executives changed the organization of car design. Instead of a conveyor-belt process where designers handed their intermediate results to engineers who in turn passed it on to manufacturing, and so on, experts from sales, marketing, manufacturing, and design interacted directly with one another in the development of Ford's next model. The resulting Ford Taurus was widely successful, and literally "the car that saved Ford," Taub [1991].

¹⁶Evidence on the interaction between task uncertainty and organization form is rarer still. Graubner [2006] provides some suggestive evidence. He compares organizational, managerial, and marketing practices in 33 medium sized and large non-in-house management consultancy firms in Germany and Northern Switzerland. He finds that these practices vary widely in the specificity of their marketing, involvement of senior partners, the likelihood that a project undergoes a shift in focus, and the total number of hierarchy levels. In particular, he finds a correlation between non-specific marketing and a higher likelihood of senior partner involvement in projects. More specific marketing is more likely to attract client with specific projects. Thus marketing specificity can be interpreted as a proxy for task uncertainty. Interpreting higher senior partner involvement as more team-like structure, this finding is consistent with this model. However, the interpretation is stretched too thin to be conclusive.

¹⁷The mini mills studied turn scrap steel into steel bar products. Problem solving teams in this study are defined as "a group of workers who meet regularly with a goal of solving production problems that managers or workers identify." The authors measure complexity by classifying the output products into one of four categories ranging from simple bars to intricate steel products with small tolerances.

At the end of the 1980s, the hearing aid manufacturer Oticon was confronted with a similar uncertainty and change in the environment. Oticon’s existing expertise in behind-the-ear hearing aids stood in contrast with the emerging in-the-ear paradigm. There was a high uncertainty about the direction the company should take. In 1991, the firm transitioned from a functional department-based organization to one completely flat and project-based. This reorganization resulted in a burst of new technology developments and innovative products, saving Oticon from the brink of bankruptcy, Foss [2003], Verona and Ravasi [2000].

On a smaller scale, a study of airplane pilots has shown that unusual and critical situations are handled best when pilot and copilot work together as a team, instead of following the hierarchy implied by their relative ranks. Earl Weener, chief engineer for safety at Boeing, observed that “for a long time it’s been clear that if you have two people operating the plane cooperatively, you will have a safer operation than if you have a single pilot flying the plane and another person who is simply there to take over in case the pilot is incapacitated.”¹⁸

Conversely, it also seems to be true that companies switch to a more hierarchical organization if uncertainty decreases. For example, after a long period of uncertainty about the future of cellular phones - would they be toys, portable hand-held devices, a more sophisticated two-way radio? - markets seemed to stabilize in the mid and late 1980s with increasingly predictable consumer needs and product features. During this period the four companies then dominating the cellular industry, AT&T, Motorola, Matsushita, and Nokia, all changed their cell phone development from a team environment toward a more hierarchical setting, Richard K. Lester and Malek [1998].

3 Innovation as an Organizational Capacity

In this section, I explore the organizational design when addressing task distributions with high task uncertainty in more detail. The analysis applies to all such distributions, but the main application that I have in mind is “organizing for innovation.” In the first subsection I highlight how task uncertainty affects task-specific investments, in the second I discuss anecdotal evidence about organizational tensions as an obstacle to innovation.

3.1 Task Specific Investments

I have shown that both task uncertainty and task complexity result in team formation. In this subsection, I show that complexity- and uncertainty-driven teams can be distinguished empirically by considering the timing of investments in task-specific tools.

Formally, assume that the firm can choose from a menu of investments $(C_i, r_i(\mathbf{p}))$. Each such investment increases the revenue generated when problem \mathbf{p} is solved to $(1 + r(i, \mathbf{p})) \cdot v_{\mathbf{p}}$. If the firm makes this investment after a problem is drawn, the cost are C_i . If the investment

¹⁸Source: “Outliers: The Story of Success”, Malcom Gladwell, Hardcover edition 2008, p. 185

is made ex-ante, the cost are δC_i , with $\delta < 1$. This discount may reflect a discount in bulk ordering or the value of being able to make the investment at an opportune moment. I call an investment i *generic*, if $r_i(\mathbf{p})$ is independent of \mathbf{p} , and *specific* to a problem $\hat{\mathbf{p}}$ if $r_i(\mathbf{p}) = 0$ everywhere except for $\mathbf{p} = \hat{\mathbf{p}}$. Most investments in practice are on a continuum between these extremes. I assume that the firm only has to make this investment once per problem and that the benefit accrues independently of which agent addresses the problem.¹⁹ I say that an investment i is *valuable* for problem \mathbf{p} if $r_i(\mathbf{p}) \cdot v_{\mathbf{p}} \geq C_i$.

The firm decides for each i whether or not to make the investment. For each investment it makes, the firm also chooses between investing ex-ante and investing ex-post. The benefit of investing ex-post is that the investment is only made for problems for which the investment is valuable. If this benefit exceeds the higher ex-post investment cost, then the firm invests ex-post. Formally, if the firm invests in i , the investment is made ex-post if

$$\left[\sum_{\mathbf{p}} f(\mathbf{p}) r_i(\mathbf{p}) \cdot v_{\mathbf{p}} - \delta \cdot C_i \right] < \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}) [r_i(\mathbf{p}) \cdot v_{\mathbf{p}} - C_i].$$

Simplifying this condition yields the following proposition:

Proposition 5 Assume $\delta < 1$. If an investment i is made, then the investment i is made ex-ante if

$$\delta \leq \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}).$$

If a valuable investment i is specific to problem \mathbf{p} , then the investment is made ex-ante if and only if the probability of problem \mathbf{p} satisfies $f(\mathbf{p}) \geq \delta$.

The proposition states that an investment is made ex-ante if the ex-post investment cost from all the cases where the ex-post investment is valuable exceeds the unconditional ex-ante investment cost. The result is intuitively obvious. It highlights, however, the importance the task distribution f plays in the timing of the investment. In particular, the proposition implies that (2) generic investments that are valuable for all problems are always made ex-ante, that (2) task-specific investments for frequent tasks are made ex-ante, and that (3) task-specific investments for rare tasks are made ex-post. In other words, a complexity-driven team which frequently confronts the same complex tasks invests mostly ex-ante, while an uncertainty-driven team which almost never faces the same problem twice makes generic investments ex-ante, but task-specific investments ex-post.

Anecdotal evidence indeed shows that uncertainty- and complexity-driven teams differ in more than just their investment patterns.

Observation 3 Task complexity and task uncertainty result in the formation of different teams and in different investment pattern.

¹⁹For the purpose of this discussion, it is useful to assume that the investment decision does not interact with the optimal organizational form. But I believe that this interaction is important in practice and may be a fruitful direction for future research.

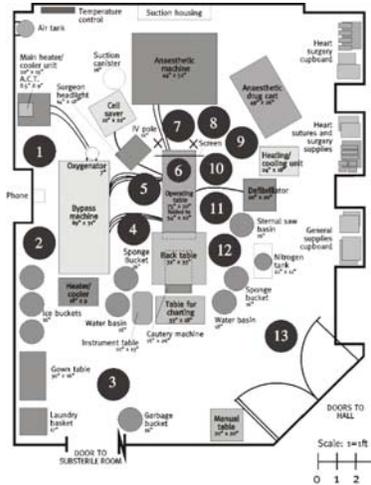
Team Work

a)



Operating Team at Work

- | | |
|--------------------------|-------------------------------|
| 1 Primary perfusionist | 6 Patient |
| 2 Secondary perfusionist | 7 Anaesthetist |
| 3 Circulating nurse | 8 Anaesthetic resident |
| 4 First scrub nurse | 9 High risk anaesthetic nurse |
| 5 Surgeon | 10 First surgical assistant |



Operating Room Layout Example

- | |
|------------------------------|
| 11 Second surgical assistant |
| 12 Second scrub nurse |
| 13 Circulating nurse |

b)



IDEO Team at Work.



Figure 9: a) An operating team at work and the layout of an operating room used for pediatric open heart surgery. The figure shows the fixed arrangement repeated for each operation and names a variety of machines used in the operating room. b) A team at the industrial design company IDEO at work. There is no particular layout for a team's workspace. Some of their most frequently used tools include "Sharpie markers, giant Post-its for the walls, and rolls of old-fashioned butcher-shop paper on the tables" as well as "foam core, blocks, tubing, duct tape, whatever might be helpful."

Figure 9 contrasts a surgery team with a team working for the industrial design company IDEO.²⁰ The figure highlights the differences in how both teams operate. Surgery is performed by an operating team, because it is a highly complex task. The skills required are relatively predictable. Every person in the operating room performs a specific sub-task. The position of every team member in the operating room is predetermined. The surgeon and the anesthesiologist lead the team and have the overall responsibility. The work environment and interactions within the team reflect a high degree of task specificity.

Most of the tools and machines used during surgery are not acquired for the care of any individual person. The majority of investments are made by the hospital before any patient using them is diagnosed. Many of the tools routinely remain in the operating room to be used during almost every operation.

In contrast, a team working on an industrial design project always faces a large uncertainty about the knowledge that is required to complete the project.²¹ For example, when a team at Design Continuum started to work for Reebok to develop a new athletic shoe, no one anticipated that experience with inflatable splints and knowledge about IV bags used in hospitals would prove crucial.²² Teams are formed for each project and dissolved after completion. Team members are chosen for their widely varying backgrounds as much as for their differing personalities. Who leads a project is not pre-determined by background or seniority. Instead leadership often emerges out of personal excitement about a project. For these uncertainty-driven teams, there is no particular lay-out of the floor plan or workspace. Offices get rearranged frequently to suit the needs of a particular project or an employee's personality.

Tools readily available at IDEO are generic and inexpensive. They include "Sharpie markers,

²⁰Sources: a) The Pediatric Cardiac Surgery Inquest Report, Chpt. 3, http://www.pediatriccardiacinquest.mb.ca/ch03/diagram3_2.html, Palestine Children Relief Fund, http://palestinote.com/cfs-filesystemfile.ashx/___key/CommunityServer.Blogs.Components.WeblogFiles/news/0576.Steve-Sosebee-_2D00_-Italian-doctors-1.jpg

b) <http://blog.gentry.io/ideo-seen-through-a-leica-m3>.

²¹Indeed, usually in the early stages of a project is used to "get to know" the project, the client, and their business. This process differs from clinical diagnosis: At a hospital a diagnosis is generally used to classify a medical problem. The diagnosis is used to determine treatment or which expert a patient should be referred to. The process following the diagnosis is contingent on the outcome of the diagnosis. In contrast, "getting to know" projects at an industrial design company is about the acquisition of knowledge needed to solve the problem. Projects are rarely "passed on." The same team members work on a project from beginning to end.

²²In 1988, shortly after Nike launched "air technology" shoes, Reebok hired DesignContinuum to find a good response. The project team that was formed came to the conclusion that using another form of "energy return" would not have the impact Reebok was looking for. The idea to develop an inflatable shoe to create a custom fit emerged because one team member had previously worked on inflatable splints. Another team member had worked on medical equipment before and realized how medical IV bags could be modified into shaped air bags. Finally, through brainstorming sessions with designers who had worked with diagnostic instruments the details of inflating and deflating the shoe were developed, Hargadon [2003]. Within a year the shoe was launched, and despite the relatively high price of \$170, over 20 million pairs were sold worldwide in the following four years, see *History of Reebok* at <http://www.hoopsvibe.com/sneaker-vibe/the-history-of-reebok-in-the-sneaker-industry-ar30040.html>.

giant Post-its for the walls, and rolls of old-fashioned butcher-shop paper on the tables” as well as “foam core, blocks, tubing, duct tape, whatever might be helpful.”²³ Other tools and materials are purchased during the project. For example, if a team is working on remote controls, it may buy all remote control models available to study their usage and to take them apart.

Overall, these two (kinds of) teams differ significantly, in particular in their investment behavior. Three alternative hypotheses may come to mind to explain these differences. First, “projects” in the operating room may be more urgent than those at IDEO, so that materials and tools have to be available “just in case”, even if they are not needed for every surgery. While this is true for emergency surgeries, most surgeries do not take place in an emergency situation. Indeed, doctors sometimes require a special tool or individually adapted aid and postpone the surgery until the tool or aid is available.²⁴ Second, one may argue that the tools required by the IDEO team are much more task-specific and are therefore only purchased after a project has began. But many tools used in the operating room, including clamps for blood vessels, suction tubes, and surgical staplers, have less usage outside the operating room than most things purchased by IDEO teams have outside of IDEO. Finally, in the operating room “lives are at stake” which one may argue explains higher ex-ante investments. However, ex-ante investments are not confined to life-or-death situations. For example, many dentists invest in an x-ray machine before they face their first patient who requires an x-ray exam.

3.2 Application to Organizing for Innovation

Bill Buxton, a designer and principal researcher at Microsoft, observed, “almost anyone who has actually built a product [...] will tell you that they really didn’t know enough to start until they had finished.”²⁵ Solving problems with a high uncertainty about the knowledge required to solve the problem is at the heart of significant innovation. I use the term “significant innovation” to refer to innovation that is not necessarily radical in the sense that it competes with existing products²⁶, but it is more novel than incremental improvements of existing products. The most successful innovations rely on existing technology and knowledge but combine them in novel ways.²⁷ Such innovation is rarely the work of a “lone

²³Source: *The Art of Innovation*, p. 62

²⁴For example, surgeons generally wait to determine a patient’s blood type and then call in that type of blood from the blood bank rather than having all types available in the operating room. Only in extreme emergencies, type O- blood is given without waiting for the result of a blood test.

²⁵“Sketching User Experiences: Getting the design right and the right design”, p.78

²⁶For example, the launch of the iPod did not result in fewer sales of Apple’s existing products such as iMac computers.

²⁷For example, electronic mail was created in 1972 by combining two existing codes for an intra-computer messaging application and an inter-computer file transfer protocol. It immediately changed how the APRAnet, the predecessor of today’s Internet, was used. The Reebok Pump sport shoe used modified medical IV bags to implement “splint in a shoe” air bladders. Using an existing IV bag manufacturer as the supplier, development and preparation for production took only six months each. In its first year, the Reebok Pump generated one billion USD in revenue. Finally, the IBM PC launched in 1981 was famously

genius” but usually due to the effort of several people. How should the people engaged in innovation work together and interact with one another? Implementing innovation is a question of organizational design.

If high task uncertainty is a valid characterization of significant innovation, then this model implies that (uncertainty-driven) teams are the corresponding organizational form. In contrast, the firm’s more predictable economic activities, e.g., production, purchasing, product distribution, are optimally addressed by a hierarchical structure or, for complex tasks, by complexity-driven teams. This difference in optimal organizational form makes implementing significant innovation alongside other economic activities challenging.²⁸ While there can be two divisions within a firm that are organized differently and employ distinct managerial expertise, any competition for company resources between these divisions will be amplified by the differences between the divisions. Coordination between the divisions and aligning them with the firm’s interests will be costly. As a consequence, firms may separate the pursuit of significant innovation from their other activities (“start-up within a firm”) or not pursue significant innovation at all.

This is consistent with the following two observations:

Observation 4 Significant innovation is often separated from the firm’s other activities either in time or space.

A particularly strong separation occurs when a firm consults an industrial design company. For example, a large laundry detergent manufacturer approached IDEO to help them find a way of making laundry detergent an attractive product for families in suburban China. IDEO, however, has neither particular expertise in China nor particular knowledge about laundry detergents. But IDEO’s organization is optimized for rare problem solving. The model presented here suggests that it was easier for IDEO personnel to learn about detergents and about laundry needs in Chinese communities than for the detergent manufacturer to adapt its organizational structure.

Firms sometimes implement policies explicitly designed to free employees from the regular organizational structure in order to facilitate innovation activities. Two such policies are skunk works and bootlegging.

built with “off the shelf parts”. Thus by-passing usual IBM testing procedure, development was completed in a year. The PC’s success is widely known and highlighted by its 1984 choice as Time Magazine’s “Person of the year.” In contrast, completely novel ideas often need a long time to be perfected in production and be accepted in the marketplace. The zipper constitutes an extreme example: Its idea required over sixty years of maturation - from the first zipper-like product patented in 1851 to the first zipper that functioned smoothly and could be cheaply produced in 1914. It took another twenty years until the product was fully accepted by the public.

²⁸In the literature on the management of innovation this tension is known as the flexibility-efficiency tension or the exploration-exploitation trade-off, Richard K. Lester and Malek [1998], Harry Boer [2005], Mats Magnusson [2008]. This literature is mostly concerned with managerial practices that can mitigate this tension. In contrast, this paper provides a micro-founded theoretical framework explaining the drivers of this tension.

Skunk works²⁹ generally refer to authorized projects that have been given a large degree of autonomy. They usually involve a relatively small and highly qualified team and are often located away from the firm's headquarters or main production site to facilitate an "out of the box" approach. For example, throughout its history IBM has allowed skunk works to deviate from its general organizational structure which was built to establish "barriers to hasty [product] introductions."^{30,31} IBM's best-known skunk work project is, of course, the IBM PC.³²

In contrast, bootlegging refers to employees spending time on non-authorized projects to the benefit of the firm with few or no resources at all. Some firms explicitly permit bootlegging. For example, 3M, Hewlett-Packard, and Google allow their employees to spend 10 to 20% of their time on "pet-projects" in the hope that they eventually result in a return for the company. Post-it notes, Gmail, Google News, Orkut and AdSense are all products that were originally developed by employees while bootlegging. Apple's graphic calculator NuCalc is an example for a bootleg project developed "under the radar", i.e., not in officially permitted bootleg time. Bootlegging is not a practice constrained to global innovation leaders. In extensive interviews with R&D managers and researchers at 57 industrial firms in Germany, France, and the United Kingdom, Augsdorfer [1996] found that bootlegging was practiced in 47 out of the 57 firms. Only in two sectors, namely Health care and Mechanical Engineering, less than 70% of firms practiced bootlegging.

Observation 5 Many firms do not generate significant innovation successfully in-house.

This stylized fact is supported by the many firms that fail or are bought out after having launched one (successful) product and possibly a few incremental improvements, and - by

²⁹The term originated with the Lockheed Martin Advanced Development Program when in 1943 under the direction of Kelly Johnson a new jet fighter was designed and build for the Air Force in 143 days. Lockheed Martin to this day showcases Kelly Johnson's accomplishment and emphasizes that "[w]hat allowed Kelly to operate the Skunk Works so effectively and efficiently was his unconventional organizational approach. He broke the rules, challenging the current bureaucratic system that stifled innovation and hindered progress.", see <http://www.lockheedmartin.com/aeronautics/skunkworks/>.

³⁰Source: *Once Upon a Time in Computerland*, p. 124

³¹For example, in the 1960s, a task force was allowed to "sidestep the company's bureaucracy to work on solutions - not reports", see *Blue Magic*, p.20, to develop the concept for "System/360" computers. System/360 was the first suite of compatible computers allowing corporate customers to upgrade their hardware as their needs grew without having to rewrite their (customized) software. The system/360 architecture became an industry standard. In the late 1980s, a task force operating under the code name "Silverlake" stepped outside the box of IBM's regular development cycle and short-cut development times to build a prototype of a new medium-sized computer, which later became as AS/400 one of IBM's best-selling products, see *The Silverlake Project*.

³²The project was approved by IBM's management in 1980 after IBM's first desktop computer had earlier failed in the market place. The group was located in Boca Raton, Florida, far away from IBM's headquarters in Armonk, New York. It deviated from many of IBM's development principles, sidestepped IBM's hierarchy, and reported directly to IBM CEO Frank Cary. Within a year IBM announced the "IBM personal computer", short "PC". At the time when the IBM PC was announced in August 1981, IBM had expected to sell about 200,000 units of the personal computer in the first three years, see *Big Blues*, p. 89. Instead, the PC became a huge success with 200,000 units sold in some months. In 1982, "Time Magazine" named the personal computer it's "Person of the Year."

the same argument - by the firms that grow their product line through acquisitions. For example, Adobe Systems is one of the largest software companies in the world, as well as one of the oldest. Adobe's portable document file format (pdf) is omni-present. Book layouts are created with InDesign, design artists around the world rely on Illustrator, and few photos are published without Photoshop editing. In 2009, Adobe Systems generated \$2.95 Billion USD revenue.³³ However, as Buxton [2007] points out, only two of the seven major current products (InDesign, Illustrator, Dreamweaver, Photoshop, Acrobat, Adobe Audition, Flashplayer) were developed in-house, namely Illustrator and Acrobat.³⁴ All other products originated outside of Adobe. Between 1990 and 2009, Adobe System merged once (with Aldus in 1994) and acquired 27 companies. Of these 28 companies only two, namely Aldus and Macromedia, had more than one product or product idea.

An alternative explanation for both of the above observations is that significant innovation is difficult to contract on. Because input is difficult to monitor and output is associated with a high variance, it is difficult to incentivize agents to engage in innovation, in particular in a multi-tasking setting. Such contractual concerns constitute a long-known obstacle to innovation, (see Holmstrom [1989], Manso [2007], Ederer [2008]). However, in practice contractual concerns appear less of an obstacle to innovation than organizational concerns. Each year since 2004, the Boston Consulting Group and the magazine *BusinessWeek* conduct a Senior Management Survey on Innovation. Every year, more respondents list "Lack of coordination within company" as an obstacle than "Compensation not tied to innovation results", sometimes significantly more.

4 Organization and Environment

In this section I return to considering all task distributions. I explore how a firm's organizational form interacts with the firm's environment, specifically, with the technology, incentive concerns, and the labor market that characterize the environment.

4.1 Organization and Technology

Changes in firm's technological environment can exogenously induce a shift toward more complex tasks or place a on finding solutions quickly. Firms can also endogenously adopt technology to decrease communication and coordination costs, or to lower the costs of problem diagnosis. Each of these changes may affect a firm's optimal organizational structure. I focus on the second kind of changes, where technology adoption is an endogenous choice of the firm.

³³Source: Adobe Reports Fourth Quarter and Fiscal 2009 Results, <http://www.adobe.com/aboutadobe/pressroom/pressreleases/pdfs/200912/Q409Earnings.pdf>.

³⁴The launch of Illustrator in 1987 marked the reinvention of Adobe Systems as a software application company. Acrobat has been the only major in-house product development since then.

4.1.1 Organizational Change, Communication, and Coordination Cost

Empirical evidence suggests that adoption of information technology (IT), decentralization of workplace organization, and demand for skilled workers are within-sector complements in productivity, (Caroli and van Reenen [2001], Bresnahan et al. [2002]). When integrating communication technology cost into this paper’s model of organizational design, a pattern of complementarities naturally arises. However, within this model, both adoption of IT technology and organizational change are driven by the firm’s task distribution.

For specificity, consider the economy with $N = 2$ units of knowledge. I call coordination the cost differential $coord = c_{AB} - (c_A + c_B)$.³⁵ In the model so far, workers at different levels of a hierarchy do not communicate. Unsolved problems are passed on, and every worker’s attempt to solve a problem starts from “scratch.” Now assume that through a new technology agents who cannot solve a problem can create a report about the problem. A worker at a higher level of the organization only needs time τ to solve the problem if he has access to the report. The organization is better off with the new technology if and only if $\tau < 1$. I assume that the report is always useful at later stages of the organization, and that amendments to the report by workers at intermediate stages do not affect the reading cost. For example, the cost for the flat and the multi-layer hierarchy in the $N = 2$ economy are, respectively, given by

$$\text{Cost}(A \rightarrow AB; coord, \tau) = c_A + \tau(1 - p_a)(c_A + c_B + coord) \quad (3)$$

$$\text{Cost}(A \rightarrow B \rightarrow AB; coord, \tau) = c_A + \tau(1 - p_a)c_B + \tau p_{ab}(c_A + c_B + coord) \quad (4)$$

The corresponding expressions for the non-communication case can be found in table 1. This specification aligns with others used in the literature.³⁶

If τ alone decreases, so does the overall cost of adopting a hierarchical organization. Thus, hierarchies become more prevalent. If $coord$ alone decreases, accessing large knowledge sets becomes relatively less expensive. Thus, firms switch to flatter organizational forms. However, as technology improves, both coordination cost $coord$ and communication cost τ will simultaneously decrease. If τ decreases faster than $coord$, then the overall effect of both technologies results in an increased prevalence of flat hierarchies:

Proposition 6 Assume that every distribution is profitable to address. Assume that both communication cost τ and coordination cost $coord$ depend on a technological parameter s such that $coord'(s), \tau'(s) < 0$. If $coord(s)/\tau(s)$ is increasing in s , then flat hierarchies

³⁵If $coord < 0$, it is more natural to think of it as a “synergy” between A and B .

³⁶For example, Garicano [2000] and Garicano and Rossi-Hansberg [2006] assume that communication cost takes $1/h = \tau$ time while problem solving takes no time at all. Bolton and Dewatripont [1994] consider $\tau = \tau(x)(\hat{m}_i + \lambda + am_i)$ as a function of the number of units m_i that are being reported to worker i , a fixed communication cost λ , the number of units \hat{m}_i that worker i has to process in addition to the units he has read about, and returns to specialization $\tau(x)$ that is larger the more frequently a worker engages in the same task. Becker and Murphy [1992] consider $coord = C(n)$ as a function of the number of team members n .

become more prevalent. In particular, firms may optimally switch from either multi-layer hierarchies or teams to flat hierarchies.

Proposition 6 emphasizes the ambiguous effect technology can have, depending on the channel through which technology acts and on the distribution a firm faces. In particular, an econometrician may find that technological progress results in more hierarchical structures if his sample mostly includes task distributions where teams or flat hierarchies are more likely to occur. In contrast, if his sample includes mostly task distributions for which multilayer hierarchies are optimal, he may find that technological progress results in flatter and less hierarchical organizations.

An econometrician would also find a correlation between higher benefits from technology adoption and organizational change, because more firms are more likely to change their organizational form where the benefit from technology is large:

The benefit from technological change is largest for firms that are optimally organized in at least two stages and that face a distribution where p_a is small. This is due to the nature of communication technology: τ proportionally decreases the cost of the hierarchy from the second stage on-ward. For example, in equations (3) and (4) all but the first summand on the right hand side are weighted by τ . As τ decreases, all firms organized as hierarchies benefit from a lower cost of implementing their organizational form. But firms that face a distribution where problems b and ab are more likely, i.e., where p_a is small, benefit more than other firms.

More firms are more likely to change their organizational form when p_a is small. To see this, consider first firms that switch from a team AB to a flat hierarchy $A \rightarrow AB$. After the change in organizational form, these firms face the smallest values of p_a among all firms organized as a hierarchy: If they faced a distribution with a larger p_a , then a flat hierarchy would have been their preferred organizational form before the change in technology. Because of p_a is small, they also benefit the most from the change in technology. Second, consider firms that switch from a multilayer to a flat hierarchy. Here, the connection between higher benefits from technology adoption and organizational change is a different one: Firms facing distributions with large and small values of p_a change their organizational form. However, the number of firms that change their organizational form is larger for small values of p_a . To see this, consider the indifference line between $A \rightarrow AB$ and $A \rightarrow B \rightarrow AB$. It is given by

$$(1 - p_a)(c_A + coord) = p_{ab}(c_A + c_B + coord).$$

This condition describes a line through the two distributions $(p_a, p_b, p_{ab}) = (1, 0, 0)$ and $(p_a, p_b, p_{ab}) = (0, c_B/(c_A + c_B + coord), (c_A + coord)/(c_A + c_B + coord))$. A change in $coord$ affects a larger range of distributions the smaller p_a is. In other words, if $coord$ and τ co-vary, then the change in $coord$ induces more firms to change their organizational form exactly where the benefit from the change in τ is largest.

4.1.2 Organizational Change and Problem Triage

I use the term “problem triage” to refer to any mechanism that allows a firm to learn about, or “diagnose”, the knowledge required to solve a problem without solving it. For example, an x-Ray may reveal whether an orthopedic surgeon or a physical therapist is more likely able to help a patient. Self-diagnosis can also be a form of problem triage. If clients have private information about their problem, then lower search cost allows them to easier find a firm specialized in their kind of problem. This in turn leads to more firms offering specialized services.

For any problem drawn from a problem distribution f , the triage mechanism reveals some information i about the nature of the problem. All problems for which the mechanism yields the same information form the conditional distribution g_i .³⁷ For example, the mechanism may be able to determine one of the units of knowledge needed,³⁸ or may be able to distinguish between simple and complex problems.³⁹

Information gained through a triage mechanism is always weakly beneficial. It is strictly beneficial if and only if at least one of the output distributions is optimally addressed by an organizational form different from the one that optimally addresses the original distribution. This is illustrated in figure 10.

The induced change in organizational form strongly depends on the particular triage mechanism. For example, a mechanism that can perfectly identify all problems results in every problem being addressed by the first-best minimal knowledge set. Since each problem is then solved in the first attempt, the number of organizational levels observed in the economy strictly decreases. However, an imperfect mechanism may remove enough uncertainty about the required knowledge that uncertainty-driven teams are no longer optimal, but not enough to resolve hierarchical organizations. Such a mechanism may increase the number of observed organizational levels. For example, in the $N = 2$ economy, the mechanism that can detect whether a problem requires knowledge A or B , and may reveal the need for either unit for an ab problem, increases the prevalence of hierarchies.

Empirically, many benefits from triage mechanisms have been evaluated, such as improved medical outcome, increases consumer welfare, etc. However, there is little empirical evidence

³⁷Formally, a triage mechanism is a decomposition of a problem distribution f into two or more distributions g_i such that

$$f = \sum_i \lambda_i \cdot g_i \quad \text{with} \quad \sum_i \lambda_i = 1.$$

³⁸For $N = 2$, a mechanism that detects A or B formally sends a distribution f to the following weighted sum

$$\mathcal{M} : (p_a, p_b, p_{ab}) \mapsto \frac{p_a}{p_a + p_b} \cdot (p_a + p_b, 0, p_{ab}) \oplus \frac{p_b}{p_a + p_b} \cdot (0, p_a + p_b, p_{ab}).$$

³⁹For $N = 2$, this mechanism can be expressed formally as the following decomposition

$$\mathcal{M} : (p_a, p_b, p_{ab}) \mapsto (p_a + p_b) \cdot \left(\frac{p_a}{p_a + p_b}, \frac{p_b}{p_a + p_b}, 0 \right) \oplus p_{ab} \cdot (0, 0, 1).$$

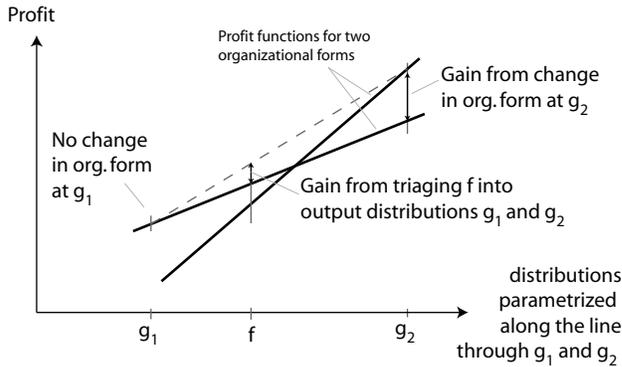


Figure 10: Triage technology is only strictly beneficial when at least one output distribution is optimally addressed by an organizational form different from the one optimal for the original distribution.

for the organizational impact of such triage mechanisms. Following the above analysis, this model predicts that the effect of triage on the number of levels in a hierarchy is ambiguous, but that the overall benefit from triage-induced organizational change is large.

4.2 Organization, Compensation, and Firm-Specific Human Capital Investments

Let us consider a firm that wants its employees to invest in firm-specific human capital, which increases the workers' productivity. As Prendergast [1993] points out, firm and worker face a dual moral hazard problem: The firm cannot prepay the worker to invest in learning, because after having received the payment, the worker has no incentive to actually learn. The worker can also not be compensated ex-post, because the firm has an incentive to renege.

To overcome this moral hazard problem, there are several ways in which the firm can commit to compensate the worker after he has invested. First, the firm can invest to build a reputation for fair compensation. Second, the firm can hire a monitor whose judgment is court-enforceable (Thiele [2009]), or, third, the firm can use a career ladder of sufficiently distinct jobs (Prendergast [1993]). The cost and feasibility of the last two commitment devices vary with the task distribution the firm faces and with the firm's organizational form. For example, if the firm's optimal organizational form is a single-level organization, then promotion to a different job is not available to the firm. I assume that reputation building is always available at a relatively high cost, and I consider how the firm's choice of commitment device depends on the firm's task distribution.

For specificity, assume that there are only two units of knowledge: A is verifiable and non-specific knowledge, whereas B is firm-specific expertise that an agent can only learn once he has joined the firm. Learning B requires effort e by the agent.

Prendergast [1993] shows that agents can be induced to learn B if the firm commits to a job-wage schedule that pays different wages for different jobs. If learning B increases the worker's productivity more in a better-paid job, then the firm has an incentive to promote the worker after he has learned B . The wage increase provides a sufficient incentive for the worker to invest in learning B . Formally, if the firm is organized as a two-layer hierarchy $A \rightarrow AB$, then the firm promotes the worker ex-post from the “ A -job” to the “ AB -job” if and only if

$$\frac{1}{1-p_a} [p_b v_b + p_{ab} v_{ab}] - c_{AB} > p_a v_a + p_b v_b + p_{ab} v_{ab} - c_A.$$

The left-hand side describes the productivity *per worker* at the top level of the hierarchy minus the promised wage c_{AB} , while the right-hand side describes the productivity of an agent with knowledge AB at the lower hierarchy level minus the wage promised for this job, c_A .

For the job-wage schedule to work as a commitment device, the firm has to have sufficiently distinct jobs. Within this model this requirement means that the problem an AB worker solves on a second level in a hierarchy must be sufficiently distinct from those the worker would solve in a first stage. Denote the average value a worker in the “ AB -job” generates by

$$\bar{v}_{ab} = \frac{p_b}{1-p_a} v_b + \frac{p_{ab}}{1-p_a} v_{ab}.$$

Then the above inequality reduces to

$$p_a > \frac{c_{AB} - c_A}{\bar{v}_{ab} - v_a}.$$

In other words, promotion is a valuable commitment device for the firm, if the problem a occurs frequently enough to make second level work sufficiently distinct from that at the first level. This raises the question whether a firm might ever distort its organizational form to be able to use promotion as a commitment device. The next result presents a criterion that rules out organizational distortion.

Proposition 7 Assume A is verifiable, non-specific knowledge and B is firm-specific expertise that an agent can only learn once he has joined the firm. Assume that

$$v_b, v_{ab} < v_a + c_{AB} \left(\frac{c_{AB} - c_A}{c_A} \right).$$

Then it is never optimal for a firm to distort its team organization into a two-tier hierarchy in order to use promotion as an incentive device for A workers to invest in learning B . It can be optimal for a firm to distort its team organization into a three-tier hierarchy, where workers in the “ AB ” job are shielded from solving problems that only require firm specific knowledge.

Next, consider monitoring as a commitment device. Assume the firm can pay the commitment fee m to hire a monitor, and that the monitor can detect the accurate usage of “ B ” with probability θ . The firm pays the worker c_{AB} if the monitor detects the usage of B ,

otherwise the firm pays c_A . The the worker learns B as long as $\theta(c_{AB} - c_A) > e$. Conditional on the worker's participation constraint being satisfied, the firm hires the monitor as long as $p_b v_b + p_{ab} v_{ab} - \theta c_{AB} > m$. In particular, if the monitor detects the usage of B if and only if the worker uses it in solving the problem, and if $p_b = 0$, then the two participation constraints for firm and worker become

$$p_{ab} > \frac{m}{v_a b - c_{AB}} \quad \text{and} \quad p_{ab} > \frac{e}{c_{AB} - c_A}.$$

In other words, monitoring is a valuable commitment device for large values of p_{ab} , i.e., small values of p_a . In contrast, promotion is a valuable commitment device for large values of p_a . Figure 11 summarizes the optimal choice of commitment devices along the line $p_a + p_{ab} = 1$. If promotion is not available as a commitment device, a firm organized as a hierarchy uses monitoring as a commitment device for second-level workers. The problem stream that workers at the second level address contains only ab problems. The commitment device chosen for these workers is thus the same as for the distribution with $p_{ab} = 1$.

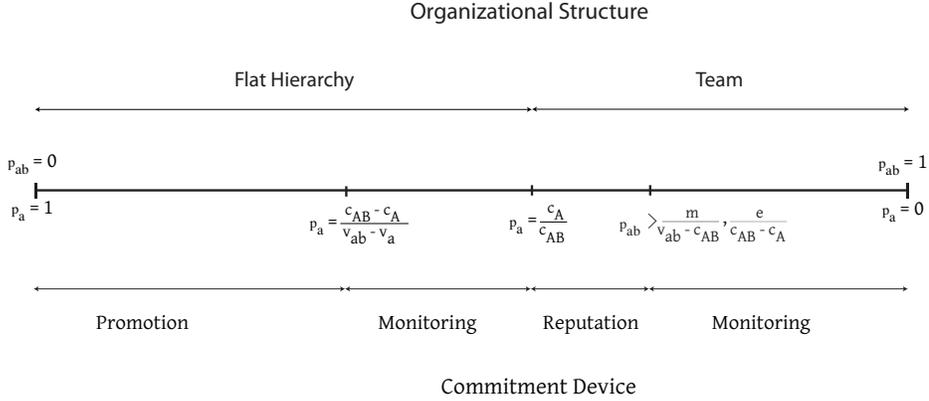


Figure 11: Assume A is verifiable and non-specific knowledge, whereas B is firm-specific expertise that an agent can only learn once he has joined the firm. Then the firm faces a dual moral hazard model, which it may overcome using a commitment device. The optimal choice of a commitment device depends on the task distribution the firm faces, as shown here for the set of distributions for which a and ab are the only two kinds of problems the firm faces, i.e., $p_a + p_{ab} = 1$.

4.3 Organization and Labor Markets

In this final section, I endogenize the cost of accessing knowledge sets and analyze the interaction between organization and labor markets. For this analysis, I restrict attention to the $N = 2$ economy. In this context, I want to think of a larger knowledge set being provided by a better skilled worker, so in particular, even large knowledge sets are provided

by only one worker.⁴⁰

Assume that the population consists of two kinds of workers: Skilled workers, who can learn two units of knowledge, and unskilled workers, who can learn only one unit. I assume that both groups of workers have reservation wage equal to zero, and that both groups of workers are homogeneous. Let m denote the number of unskilled workers in the population, and m_{AB} the number of skilled workers. I assume that unskilled workers can be trained as either an A or a B worker, so that in equilibrium, $c_A = c_B = c$. Then an equilibrium is a quadruple (c, c_{AB}, m, m_{AB}) , such that for prices $c_A = c_B = c$, c_{AB} the total demand of all producing firms for workers who can learn one unit equals m and the demand for workers who can learn two units equals m_{AB} .

To study the interaction between the labor market and the organization of firms, I have to make an assumption about the relative scarcity of firms and labor. If labor supply was unlimited, then all firms would produce and the equilibrium wage for labor would be the workers' reservation wage. The optimal organizational forms prevalent in this case are described by figure 3 in section 2.2. Conversely, if the number of firms for any distribution was unlimited, then only firms facing the most profitable distributions would produce, and these firms would make zero-profits. For any distribution in the distribution space there always exist a more profitable distribution where one problem occurs with probability one. Thus, in this case, every organizational form would be a contingency sequence of length one, and every problem that is solved is addressed by the first-best knowledge set.

To consider the relative scarcity of firms and labor, I assume that there is a mass one of firms uniformly distributed over the distribution space. Measures of labor demand and supply are relative to this mass one of firms. For example, if all firms solve all problems by hiring

⁴⁰If workers can cooperate to combine their knowledge sets to a larger set, thus forming a team, then the cost of accessing a knowledge set c_K is the sum of the wages paid to each worker and, possibly, a communication cost $coord$ that depends on the number of workers combining their knowledge. In this case, a general equilibrium involves a wage function $w(K_i)$ and a function that maps agents with knowledge K_i to teams who combine their knowledge to form a knowledge set K . In equilibrium, the wage function w has to be such that the supply of knowledge sets K equals the demand by firms, given the cost of accessing knowledge c_k as implied by w . Formally, let $w(K_i)$ be the wage a worker with knowledge set K_i receives. If t workers form a team and combine their knowledge sets K_i provided access to the knowledge set $K = \bigcup_i K_i$, then the cost to the firm is

$$\text{Cost}(K|K_1, K_2, \dots, K_t) = \sum_i w(K_i) + coord(t),$$

where $coord(t)$ is the communication cost incurred by combining knowledge sets. $Coord$ is a function of the number of agents involved in the team, $coord(1) = 0$ and $coord'(t) \geq 0$. Then the cost of accessing a knowledge set, as used in the model so far, is given by

$$c_K = \min_{\{K_i\}_{i=1}^t \text{ s.t. } K = \bigcup_i K_i} \sum_i w(K_i) + coord(t).$$

Given the supply of workers with knowledge K_i , a general equilibrium then consists of a wage function $w(K_i)$ and an assignment $K_i \rightarrow K$, such that the demand for knowledge sets K from all firms facing knowledge-access costs c_K equals the supply from all workers forming teams that supply K . Implicit in this equilibrium is that workers' wages cannot depend on the team they work in, only on their knowledge set K_i .

a skilled worker with two units of knowledge, the demand for skilled workers equals one.

For any wage combination (c, c_{AB}) , the analysis of section 2.2 allows us to compute the optimal organizational forms for all task distributions in the distribution space. Since each organizational form implies a particular demand for workers with knowledge A , B , and AB , respectively, we can aggregate demand for each worker over the entire distribution space. By repeating this process for different wage combinations (c, c_{AB}) , we see how aggregate demand varies with wages. This is summarized by constant-demand curves for different kinds of workers in the graphs in figure 12, where the demand for A and for B workers has been aggregated into one demand function for unskilled workers. Thus, the graph shows the demand for skilled and for unskilled workers for each wage combination (c, c_{AB}) .

The graph also allows the reverse analysis and enables us to find the equilibrium wage that supports a particular supply of workers in equilibrium. Given supply (m, m_{AB}) , we can find the corresponding constant demand curves where demand equals m and m_{AB} , respectively. Equilibrium wages can be read off at the intersection of these two constant demand curves. For example, in an economy with a low supply of workers with one unit of knowledge, say $m \approx 0.1$, and a relative large supply of workers with two units of knowledge, say $m \approx 0.6$, the equilibrium is represented by the circled letter “B” in figure 12. The corresponding equilibrium wages are slightly less than v_a for workers with one unit of knowledge and slightly more than v_a for workers with two units of knowledge.

Note that the wages in this model are not determined by a zero-profit condition for firms, but by the demand-supply equilibrium in the labor market. For any given wage (c, c_{AB}) , there may be firms confronting a distribution such that they prefer not to produce. Only the firm on the margin between producing and not producing makes zero profit. All firms which strictly prefer to produce make a positive profit. The equilibrium wage is determined such that the aggregate demand of all firms producing at that wage level equals the labor supply.

For different equilibria, different organizational forms are prevalent. The circled letters “A” through “F” indicate different equilibria. The lower half of figure 12 shows the corresponding partition of the distribution simplex into the regions where different organizational forms are optimal.

The constant demand curves in figure 12 show the subtle interaction between more and less skilled workers: They are partially substitutes, partially complements, depending on the relative supply of both kinds of workers. For example, consider the neighborhood of “D”: Keeping the wage for more knowledgeable workers constant and increasing the wage for less skilled workers increases the demand for skilled workers. Thus, in this region the two kinds of workers are substitutes. In contrast, consider the region in the neighborhood of “F.” Keeping the wage for skilled workers constant and increasing the wage for less knowledgeable workers decreases the demand for skilled workers. Here, they are complements.

This contrast is due to changes on different margins in the problem distribution space

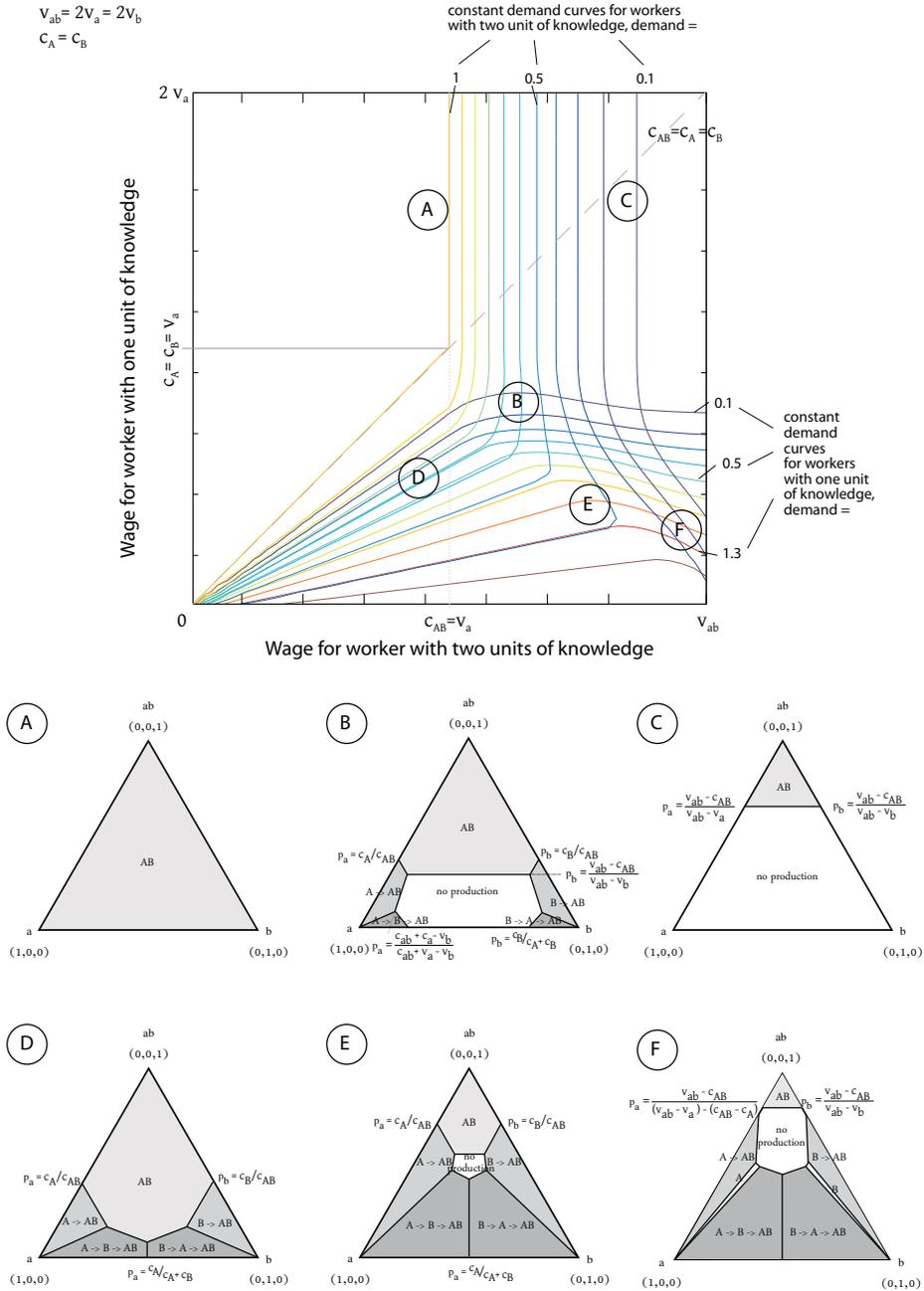


Figure 12: The top graph shows constant demand curves as the labor costs for both kinds of workers vary. Given the supply of the two kinds of workers, the intersection of the corresponding curves corresponds to the equilibrium wage such that demand equals supply. The lower half of the figure shows the prevalence of different organizational forms as supply and wages vary. The values used in these figure are: $v_a = v_b = 11$, $v_{ab} = 22$. Moreover, the wages for the six different scenarios are “A”: $c_A = c_B > c_{AB}$, $c_{AB} = 10$, “B”: $c_A = c_B = 9$, $c_{AB} = 14$, “C”: $c_A = c_B > v_a$, $c_{AB} = 18$, “D”: $c_A = c_B = 6$, $c_{AB} = 10$, “E”: $c_A = c_B = 5$, $c_{AB} = 17$, “F”: $c_A = c_B = 4$, $c_{AB} = 20$.

response to a change in market wages. Complementarity occurs as the margin between producing and not producing shifts for organizations that employ both skilled and unskilled workers. If unskilled workers become cheaper, forming a hierarchy becomes a viable form of production, and demand for skilled workers increases. This is illustrated by the change from “C” to “F”. In contrast, substitution occurs as the margin between different organizational forms changes. As the cost of unskilled workers $c = c_A$ decreases, forming a hierarchy becomes relative less expensive compared with letting all problems be solved by a skilled worker. Thus, the overall demand for skilled workers decreases. This is illustrated by the change from “A” to “D”.

Note that substitution is dominant in regions where supply of skilled workers is relatively large, such that there is a small “skill premium”, whereas complementarity is dominant where supply of skilled workers is relatively small, so that there is large wage differential between more and less knowledgeable workers.

This model may be usefully extended to study the impact of labor shocks on organizational design. For example, the impact of skilled or unskilled immigrants will depend on the existing relative supply of skilled and unskilled workers in the economy and on the kind of task distributions predominantly confronted by firms.

5 Conclusion

The modeling innovation of this paper is to present knowledge as discrete units. This complicates solving the firm’s optimization problem, but, unlike the previous literature, the model does not need to place restricting assumptions on feasible knowledge sets or feasible distributions. The insights gained from this model are threefold.

First, the model encompassing the entire task distribution space applies in particular to distributions with a high task uncertainty, which are characteristic for innovation activity. I have shown that firms confronting such distributions optimally organize as teams, whereas the firms’ more routine activities are optimally addressed by hierarchies. Teams addressing distributions with high task uncertainty also invest in task-specific tools differently than other organizational forms. This organizational and managerial contrast sets uncertainty-driven teams apart from other organizational forms, and makes implementing innovation an organizational challenge. While the literature has analyzed the contractual challenges of innovation, (Ederer [2008], Manso [2007], Holmstrom [1989]), we understand little about the corresponding organizational challenges. This analysis is but a first step toward such an understanding.

Second, the model presents a baseline, against which more complicated drivers of organizational design can be evaluated. I have shown that firms facing different task distributions adopt technology at different rates and choose different commitment mechanisms to overcome moral hazard problems. There are other organizational concerns that the model can

be extended to include. For example, the decision of a firm to vertically integrate depends on having vertically separable units and the frequency of interaction between the units. These factors, in turn, depend on the firm's task distribution. The interaction of organizational design, task uncertainty, and task complexity with other organizational choices such as vertical integration has yet to be explored.

Last but not least, the model emphasizes the empirical relevance of task complexity and task uncertainty when studying organizational design. Even absent of other drivers of organizational design, variations in task complexity and task uncertainty can give rise to different organizational forms. In the presence of other organizational drivers, a differential impact on observables can often be explained by variations in task complexity and task uncertainty.

References

- Peter Augsdorfer. *Forbidden Fruit: An Analysis of Bootlegging, Uncertainty and Learning in Corporate R&D*. Averbury, 1996.
- Patrick Bajari and Steven Tadelis. Incentives versus transaction costs: A theory of procurement contracts. *Rand Journal of Economics*, 32:387–407, 2001.
- George Baker, Michael Gibbs, and Bengt Holmstrom. The wage policy of a firm. *Quarterly Journal of Economics*, 109:921–55, 1994.
- Rosemary Batt. The economics of teams among technicians. *British Journal of Industrial Relations*, 39, No. 1:1–24, 2001.
- Gary S. Becker and Kevin M. Murphy. The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics*, Vol.107, No.4., pp.1137-1160, 1992.
- Sandra E. Black and Lisa M. Lynch. How to compete: The impact of workplace practices and information technology on productivity. *The Review of Economics and Statistics*, 83, No. 3:434–445, 2001.
- Patrick Bolton and Mathias Dewatripont. The firm as a communication network. *Quarterly Journal of Economics*, 109:809–39, 1994.
- Kathryn Shaw Brent Boning, Casey Ichniowski. Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics*, Vol. 25, No. 4:613–650, 2007.
- Timothy Bresnahan, Erik Brynjolfsson, and Lorin Hitt. Information technology, workplace organization and the demand for skilled labor: Firm-level evidence. *Quarterly Journal of Economics*, 117:339–76, 2002.
- Bill Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann, 2007.
- Peter Cappelli and David Neumark. Do "high-performance" work practices improve establishment-level outcomes? *Industrial and Labor Relations Review*, 54, No. 4:737–775, 2001.
- Sandra D. Cassard Marilyn Bergner Rebeca Wong Carol S. Weisman, Dorothy L. Gordon. The effects of unit self-management on hospital nurses' work process, work satisfaction, and retention. *Medical Care*, Vol. 31, No. 5:381–393, 1993.
- Eve Caroli and John van Reenen. Skill-biased organizational change? evidence from a panel of british and french establishments. *The Quarterly Journal of Economics*, 116, No. 4: 1449–1492, 2001.
- Jed DeVaro and Fidan Ana Kurtulus. What types of organizations benefit from team production, and how do they benefit? *Advances in the Economic Analysis of Participatory and Labor-Managed Firms*, 9:3–54, 2006.

- Florian Ederer. Launching a thousand ships: Incentives for parallel innovation. *Job market paper*, 2008.
- Nicolai J. Foss. Selective intervention and internal hybrids: Interpreting and learning from the rise and decline of the oticon spaghetti organization. *Organization Science*, 14:331–349, 2003.
- Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108:874–904, 2000.
- Luis Garicano and Esteban Rossi-Hansberg. Organization and inequality in a knowledge economy. *QJE*, 2006.
- Giovanni Maggi Gene M. Grossman. Diversity and trade. *The American Economic Review*, 90, No 5.:1255–1275, 2000.
- Robert Gibbons and Michael Waldman. A theory of wage and promotion dynamics inside firms. *Quarterly Journal of Economics*, 114:1321–58, 1999.
- Michael Graubner. *Task, Firm Size, and Organizational Structure in Management Consulting - An Empirical Analysis from a Contingency Perspective*. duv, 2006.
- B.H. Hamilton, J.A. Nickerson, and H. Owan. Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*., 11(3):465–497, 2003.
- Andrew Hargadon. *How Breakthroughs Happen: The Surprising Truth About How Companies Innovate*. Harvard Business School Press, 2003.
- Frank Gertsen Harry Boer, Joachim Kuhn. Continuous innovation managing dualities through co-ordination. *Continuous Innovation Network - Working Paper Series*, 2005.
- Bengt Holmstrom. Agency cost and innovation. *Journal of Economic Behavior and Organization*, pages 305–327, 1989.
- Bengt Holmstrom. Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):pp. 169–182, 1999. ISSN 00346527. URL <http://www.jstor.org/stable/2566954>.
- Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* 7, Special, Issue:24–52 [Papers from the Conference, 1991.
- Michael Kremer. The o-ring theory of economic development. *QJE v.108*, n3.:551–575, 1993.
- Gustavo Manso. Motivating innovation. *Working Paper, MIT Sloan School of Management*, 2007.

- Scott Masten. The organization of production: Evidence from the aerospace industry. *Journal of Law and Economics*, 27:403–17, 1984.
- Antonella Martini Mats Magnusson. Dual organisational capabilities: from theory to practice the next challenge for continuous innovation. *International Journal of Technology Management*, 42, Nos. 1/2,, 2008.
- Canice Prendergast. The role of promotion in inducing specific human capital acquisition. *The Quarterly Journal of Economics*, 108, No. 2:523–534, 1993.
- Yingyi Qian. Incentives and loss of control in an optimal hierarchy. *Review of Economic Studies*, 61:527–44, 1994.
- Michael J. Piore Richard K. Lester and Kamal M. Malek. Interpretive management: What general managers can learn from design. *Harvard Business Review*, 1998.
- J.W. Rivkin and N.Sigglekow. Speed and search: Designing organizations for turbulence and complexity. *Organization Science*, 2005.
- Sherwin Rosen. Authority, control, and the distribution of earnings. *Bell Journal of Economics*, 13:311–23, 1982.
- John Sutton. Competing in capabilities. *Clarendon Lectures in Economics 2004*, 2004.
- Eric Taub. *Taurus: The Making of The Car that Saved Ford*. A Dutton Book, 1991.
- Veikko Thiele. Subjective performance evaluations, collusion, and organizational design. *Queen’s School of Business Research Paper*, page Available at SSRN: <http://ssrn.com/abstract=763964>, 2009.
- Gianmario Verona and Davide Ravasi. Unbundling dynamic capabilities: an exploratory study of continuous product innovation. *Industrial and Corporate Change*, 12:577–606, 2000.

A Proofs

A.1 Proof of Proposition 1

Proposition 1 Characterization of optimal organizational form

If triage is prohibitive, then the optimal knowledge arrangement can be described by a finite sequence of length l of the form (K_1, \dots, K_l) . Moreover, it cannot be optimal for knowledge set to be contained in a knowledge set that occurs at an earlier stage in the contingency sequence. In other words, if $K_i \subset K_j$ for any two knowledge sets in the sequence, then $i < j$.

While proposition 1 seems intuitive, its proof is somewhat technical. I prove two intermediate results first. The proposition follows as a corollary.

First, observe that

Claim 1 Without loss of generality, we can assume that every worker only draws problems from one urn and that there is only one worker drawing problems from any urn.

The intuition is as follows. By drawing from different urns an agent works on problems from a problem distribution that is a mix of distributions. The same overall distribution, however, can be achieved by rearranging the input into these urns from workers in the previous stage.

Proof: For workers who draw problems from the original distribution, the claim is true because all urns a firm owns have the same distribution and one problem is drawn per period from each urn.

Consider workers who draw problems from “auxiliary” urns, i.e., who attempt problems that some other worker(s) could not solve. Assume that there are k workers each with knowledge set K_i , that pass unsolved problems to \hat{m} auxiliary urns, from which l workers with knowledge sets L_j draw problems, see figure 13.

Denote the share of problems each worker i leaves unsolved by $f_i \leq 1$, and the fraction of those put in urn $a \in [1, \hat{m}]$ by s_{ia} , so that

$$\sum_{a=1}^{\hat{m}} s_{ia} = f_i.$$

Each urn receives a total to $\sum_i s_{ia}$ problems, with a fraction

$$\hat{s}_{ia} = \frac{s_{ia}}{\sum_{i=1}^k s_{ia}}$$

of all problems in urn a came from knowledge set i .

Also, denote the share that each worker j draws from the urn a by t_{aj} , so that the worker j is busy for

$$g_j = \sum_{a=1}^{\hat{m}} t_{aj} \leq 1$$

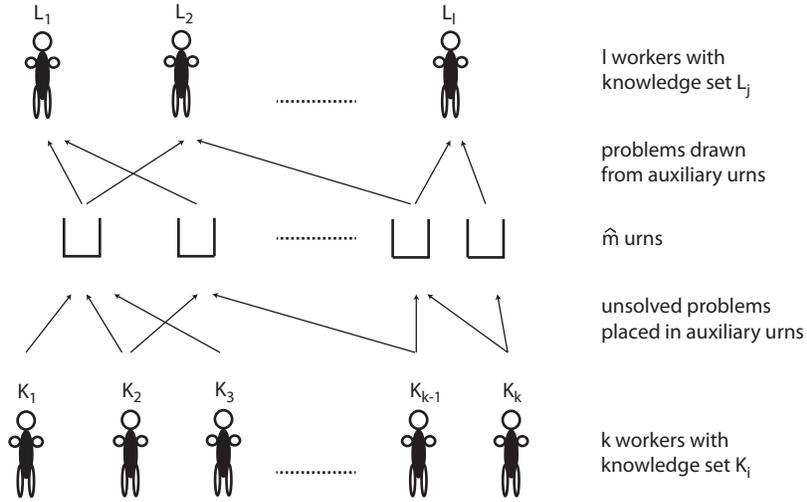


Figure 13: Notation for proof of claim 1.

fraction of all periods.

Thus, on average, worker j spends the fraction

$$x_{ij} = \sum_{a=1}^{\hat{m}} \hat{s}_{ia} \cdot t_{ja}$$

of all periods working on problems that worker i could not solve. In other words, in a fraction x_{ij} of all periods, worker i passes an unsolved problem to worker j .

It is useful to observe that if d_a denotes the share of problems discarded from urn a , then the number of problems put into urn a must equal the number drawn or discarded, so that

$$\sum_{i=1}^k s_{ia} = d_a + \sum_{j=1}^l t_{aj}.$$

I can now construct a new knowledge-set-urn mapping that reflects the problem passing between i and j directly: There are $l + 1$ urns in this new construction, one for each worker i and one for problems to be discarded. Every worker i passes x_{ij} of all problems it cannot solve to urn j , and $x_{il+1} = \sum_a \hat{s}_{ia} \cdot d_a$ to the $l + 1$ urn of discarded problems. Worker j

only draws problems from urn j . In particular, note that

$$\begin{aligned}
& \sum_{j=1}^{l+1} x_{ij} \\
&= \sum_{a=1}^{\hat{m}} \hat{s}_{ia} \cdot d_a + \sum_{j=1}^l \sum_{a=1}^{\hat{m}} \hat{s}_{ia} \cdot t_{ja} \\
&= \sum_{a=1}^{\hat{m}} \frac{s_{ia}}{\sum_{i=1}^k s_{ia}} \cdot d_a + \sum_{j=1}^l \sum_{a=1}^{\hat{m}} \frac{s_{ia}}{\sum_{i=1}^k s_{ia}} \cdot t_{ja} \\
&= \sum_{a=1}^{\hat{m}} \frac{s_{ia}}{\sum_{i=1}^k s_{ia}} \cdot \left(d_a + \sum_{j=1}^l t_{ja} \right) \\
&= \sum_{a=1}^{\hat{m}} s_{ia} \\
&= f_i
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^k x_{ij} &= \sum_{i=1}^k \sum_{a=1}^{\hat{m}} \hat{s}_{ia} \cdot t_{ja} \\
&= \sum_{i=1}^k \sum_{a=1}^{\hat{m}} \frac{s_{ia}}{\sum_{i=1}^k s_{ia}} \cdot t_{ja} \\
&= \sum_{a=1}^{\hat{m}} t_{ja} \cdot \sum_{i=1}^k \frac{s_{ia}}{\sum_{i=1}^k s_{ia}} \\
&= \sum_{a=1}^{\hat{m}} t_{ja} = g_j
\end{aligned}$$

In other words, for every arrangement of knowledge there exists an equivalent arrangement that passes the same fraction of problems from the same knowledge sets to urns addressed by the same knowledge sets. This establishes the claim. \square

As a consequence of claim 1, every arrangement of workers can be described as a directed graph, where vertices are labeled by knowledge set-urn pairs, and one vertex is labeled "problems to be discarded". Edges point from one vertex to another if problems not solved by the knowledge set of one vertex are put into the urn of problems to be addressed by the knowledge set of the other vertex. The vertex representing the urn of problems to be discarded is the only one that has no outgoing edge.

While the set of such graphs is very large, the set of potentially optimal graphs is much smaller, as the following theorem states.

Theorem 2 If integer constraints do not bind, then the optimal knowledge arrangement can be described by a finite sequence of length l of the form $(1, K_1, \dots, r_l, K_l)$ where at the

first stage $m \times 1$ copies of the knowledge set K_1 attempt to solve problems from the original distribution. At the second stage $m \times r_2$ copies of the knowledge set K_2 solve problems not solved by K_1 , and so forth. Moreover, all K_i are distinct and $1 > r_2 > r_3 > \dots > r_l$.

The idea behind the proof of this theorem is a series of observations that limit the form of graphs that can be optimal knowledge arrangements and then show that the remaining graphs can be characterized in the above form. The key observations are: First, if a worker with a particular knowledge set has not been able to solve a particular problem then its units of knowledge do not suffice to solve it. It can therefore not be optimal for any problem to be addressed twice by the same knowledge set. Thus, all K_i must be distinct and, in particular, an optimal graph cannot contain any circles. Second, the homogeneity of the original urns, which all have the same distribution, implies that workers in the first stage have all the same knowledge set. This, in turn, results in all auxiliary urns at the second stage having the same distribution, so that knowledge sets at the second stage are homogeneous, and so on. Thus, it suffices to list the knowledge sets at each stage. Finally, each knowledge set in an optimal graph solves some problems it faces, because otherwise the graph could be improved by removing the knowledge set. Consequently, each subsequent stage requires a smaller number of knowledge sets. None of these observations is surprising.

Proof: Before discussing the observations below, it is useful to introduce some notation. I refer to figure 14 to illustrate the terminology. A *subgraph* is some subset of vertices and edges of the original graph that contains originating and ending vertex for every edge it contains. For example, the vertices B and C and the edge between them form a subgraph in figure 14. A subgraph is *closed* if it contains all edges outgoing from any vertex in it. The *closure* of a subgraph is the smallest closed graph that contains the subgraph. In figure 14, the vertices B through E and the edges between them form a closed subgraph. This subgraph is also the closure of any of the vertices B through E. The *organizational closure* of a subgraph is the smallest closed graph that contains the subgraph and for all vertices that are in the closure but not in the subgraph it also contains all incoming edges. An *organizational subgraph originating in some vertex* is the organizational closure of that vertex. Intuitively, the organizational subgraph contains all vertices and edges that one needs to know about to understand what happens to problems not solved at the originating vertex. In the example of figure 14, the organizational closure of C through E contains the entire graph. Only the organizational closure of the vertex B does not contain A and the edge between A and B.

Finally, a λ -*subgraph* of a graph G refers to the same organizational arrangement as G used for a fraction λ of all periods, and I will call two graphs *revenue equivalent* if they generate the same expected value.

Observation 1 Parallel processing of two distributions cannot do worse than single processing of a mixture of the two distributions.

Formally, let f and g denote two distinct distributions, and let G be the organizationally

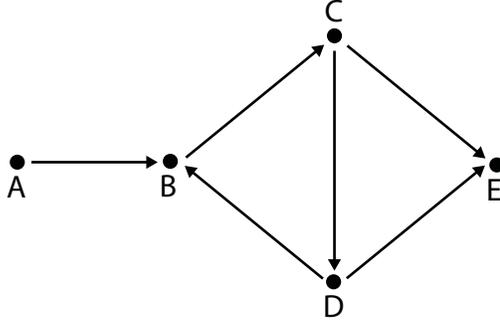


Figure 14: The vertices B through E together with the edges between them form a *closed* subgraph. This subgraph is also the *closure* of any of its vertices. Moreover, it is the *organizational subgraph* originating in B. However, the organizational subgraph originating in C, D, or E is the entire graph.

closed graph that optimally addresses $\lambda f + (1 - \lambda)g$. Then for any knowledge set K in G ,

$$\begin{aligned} & \text{Prob}(K \text{ solves problem } \mathbf{p} \mid \mathbf{p} \text{ drawn from } \lambda f + (1 - \lambda)g) \\ = & \lambda \cdot \text{Prob}(K \text{ solves problem } \mathbf{p} \mid \mathbf{p} \text{ drawn from } f) \\ & + (1 - \lambda) \text{Prob}(K \text{ solves problem } \mathbf{p} \mid \mathbf{p} \text{ drawn from } g) \end{aligned}$$

In other words, and not surprisingly, addressing the mixed distribution $\lambda f + (1 - \lambda)g$ with graph G is revenue equivalent to addressing λf with a λ -subgraph of G , and $(1 - \lambda)g$ with a $(1 - \lambda)$ -subgraph of G .

Observation 2 Extending the previous observation, if a graph contains a vertex that has two or more incoming edges with different conditional distributions corresponding to each of the edges, then this graph is not singular optimal.

Formally, let f and g denote two distinct distributions as before, and let $CV(f, G)$ and $CV(g, G)$ denote the continuation value generated by addressing these distributions with graph G . Let $CV^*(f)$ and $CV^*(g)$ denote the continuation value in an optimal graph.

Then the claim is that for all $\lambda \in (0, 1)$

$$CV^*(\lambda \cdot f + (1 - \lambda) \cdot g) \leq \lambda \cdot CV^*(f) + (1 - \lambda) \cdot CV^*(g).$$

To see that this is true, let G be the graph that is optimal for the mixed distribution, so that

$$CV^*(\lambda \cdot f + (1 - \lambda) \cdot g) = CV(\lambda \cdot f + (1 - \lambda) \cdot g, G).$$

From observation 1 it follows that

$$\begin{aligned} CV(\lambda \cdot f + (1 - \lambda) \cdot g, G) &= \lambda \cdot CV(f, G) + (1 - \lambda) \cdot CV(g, G) \\ &\leq \lambda \cdot CV^*(f) + (1 - \lambda) \cdot CV^*(g). \end{aligned}$$

The intuition is that the information which problem came from which conditional problem stream is lost when the distributions are mixed. No other value is gained. Instead of mixing

problems from f and g in one auxiliary urn that is addressed by one knowledge set, they can be put in different urns and be addressed by the same knowledge set for corresponding fractions of time. In this manner, if the problem is not solve, the information about “the problem’s past” is preserved.

This means that we can assume that edges that represent different conditional distributions point to different vertices. Equivalently, it suffices to consider graphs such that the incoming edges of any vertex carry the same conditional distribution.

Observation 3 It suffices to consider graphs where urns with the same distribution are addressed by the same knowledge set.

Let G be an optimal graph that contains two vertices where different knowledge sets K_1 and K_2 address problems from urns with the same distribution f . Moreover, let G_1 and G_2 be the organizational subgraph originating in K_1 and K_2 , respectively.

One can imagine that the continuation value of one of the knowledge sets addressing problems depends on the presence of the other knowledge set and the remainder problem stream it generates. However, from observation 2 we know that the overall value cannot be improved by mixing different distribution. Therefore, the continuation value for problems addressed by one knowledge set cannot depend on the remainder problem stream generated by the other knowledge set. Therefore, the two continuation values must be independent of one another.

If the continuation values are independent, then they must be equal, i.e., $CV(f, G_1) = CV(f, G_2)$. Then a graph, in which all problems from distributions of the form (f) are addressed by K_1 and, by extension, by G_1 , is revenue equivalent to G .

Observation 4 It suffices to consider graphs where outgoing edges from any vertex point to urns with the same distribution.

Since outgoing edges of any vertex carry the same distribution, this is a consequence of observation 3.

Observation 5 An optimal graph cannot contain any circles.

Proof by contradiction. Assume the graph G is an optimal arrangement of knowledge and contains a circle. In other words, there is a non-zero probability that a problem will pass through the same vertex twice. Then the graph can be improved by not considering the same problem twice.

Formally, let K denote the knowledge set at the first vertex a problem may pass twice. Let K' be the vertex that an edge originating in K points to. From the previous observation we know that we can restrict attention to graphs where all edges outgoing from K point to vertices with knowledge set K' . It thus suffices to argue for one outgoing edge.

Let G' be the organizational subgraph originating at K' . Because of the circle in G , there is a vertex in G' that has an edge pointing to K , and therefore K is contained in G' . Now

assume that among all problems passed to K' the fraction of problems passed from K to K' is λ .

Next, construct a new overall graph \hat{G} as follows. It is a copy of the original graph G except that the subgraph G' is split into an λ -subgraph G'_λ and a $G'_{(1-\lambda)}$ -subgraph. The edge outgoing from K in \hat{G} points to K' in G'_λ . Then \hat{G} is revenue-equivalent to G .

However, \hat{G} can be improved as follows. Since the probability is zero that K can solve a problem it previously could not solve, attempting the same problem twice does not generate value. Thus, consider the graph \hat{G}' that is a copy of \hat{G} except that all edges from the subgraph G'_λ that point to K are redirected to the next vertex at which the probability of the problem being solved is non-zero (or allocated proportional if there is more than one such vertex) or to the “discard” urn if no such vertex exists. Then \hat{G}' generates the same revenue as \hat{G} and thus as G but at a lower cost.

Observation 6 In an optimal graph the same problem cannot be addressed by the same knowledge set twice.

The intuition for this is that the probability that the knowledge set can solve the problem is zero given that it previously could not solve the problem, thus the expected value generated is equal to the cost and thus negative.

Formally, observation 5 implies that any problem is addressed by a sequence of knowledge sets. Let K_1 be the first knowledge set that address an urn with problem distribution f . Then an optimal graph contains knowledge sets K_i such that K_i addresses a problem from f if and only if K_{i-1} attempted the problem and could not solve it. Let $a_i(f)$ denote the probability that K_i addresses a particular problem. If $K_i = K_j$ for some $i < j$, then

$$\text{Prob}(K_j \text{ solves problem} | K_j \text{ gets to address the problem}) = 0,$$

and $a_j(f, G) = a_{j+1}(f, G)$. Thus the graph can be improved by pointing the outgoing edges of the vertices with K_{j-1} to vertices with K_{j+1} . The resulting graph generates the same revenue but does so at lower cost than the original graph.

Observation 7 In an optimal graph there are finitely many vertices and every problem is attempted by finitely many knowledge sets.

This is an implication of the previous observation: Since there are finitely many units of knowledge in this economy, there are finitely many distinct knowledge sets as combinations of those units. \square

Theorem 2 establishes that an optimal organizational form can be expressed as $(1, K_1, \dots, r_l, K_l)$. Given a particular task distribution, the optimal coefficients r_i for $i > 1$ are determined by the probabilities with which problems occur and with which each knowledge set solves the problems it faces. For example, $r_2 = 1 - \text{Prob}(\mathbf{p} \in \mathcal{P}(K_1))$. Thus, optimal organizational forms can be characterized by the sequence of knowledge sets (K_1, \dots, K_l) only, and proposition 1 follows as a corollary from theorem 2.

A.2 Proof of Proposition 2

Proposition 2 The region in the space of problem distributions where one particular organizational form is optimal is convex.

Proof: Denote the profit a firm generates if it uses a contingency sequence $OF = \{K_1, \dots, K_l\}$ when facing distribution f by $\Pi(f; OF)$. Note that

$$\Pi(f; OF) = \sum_{\mathbf{p} \text{ solved by } OF} f(\mathbf{p})v_{pp} - \sum_{i=1}^l \text{Prob}(\mathbf{p} \text{ not solved by } K_1, \dots, K_{i-1}) \cdot c_{K_i}$$

is linear in f .

Assume that some organizational form \hat{OF} is optimal at two distributions f_1 and f_2 . Then for $i = 1, 2$

$$\Pi(f_i; \hat{OF}) \geq \Pi(f_i; OF) \quad \forall OF.$$

Therefore

$$\begin{aligned} \Pi(\lambda \cdot f_1 + \lambda \cdot f_2; \hat{OF}) &= \lambda \cdot \Pi(f_1; \hat{OF}) + (1 - \lambda) \cdot \Pi(f_2; \hat{OF}) \\ &\geq \lambda \cdot \Pi(f_1; OF) + (1 - \lambda) \cdot \Pi(f_2; OF) \\ &= \Pi(\lambda \cdot f_1 + \lambda \cdot f_2; OF) \end{aligned}$$

In other words, \hat{OF} is optimal for any convex combination of f_1 and f_2 . This establishes the claim. \square

A.3 Proof of Proposition 3

Proposition 3 Sufficient Criterion for Optimality of Hierarchies

Assume that the marginal cost of knowledge is constant. If there exists a knowledge set K such that

$$\text{Prob}(\mathbf{p} \in \mathcal{P}(K)) > \frac{|K|}{N},$$

then a team with complete knowledge is not optimal. In particular, if the firm solves all problems, a multistage contingency sequence is optimal.

Proof: Let c denote the marginal cost of knowledge.

Proof by contradiction. Assume that the firm addresses the task distribution with a single-stage contingency sequence with complete knowledge $K_N = \{A, B, \dots, Z_N\}$. Then the firm solves all problems it confronts and generates profits of:

$$\Pi(f; \{K_N\}) = \sum_{\mathbf{p}} f(\mathbf{p})v_{\mathbf{p}} - N \cdot c.$$

Now, instead consider the organizational form $OF' = \{K, K_N\}$. The firm still solves all problems, and generates profits of

$$\Pi(f, \{K, K_N\}) = \sum_{\mathbf{p}} f(\mathbf{p})v_{\mathbf{p}} - |K| \cdot c - (1 - \text{Prob}(\mathbf{p} \in \mathcal{P}(K))) \cdot N \cdot c.$$

By assumption,

$$\text{Prob}(\mathbf{p} \in \mathcal{P}(K)) > |K|/N.$$

This implies that

$$|K| \cdot c + (1 - \text{Prob}(\mathbf{p} \in \mathcal{P}(K))) \cdot N \cdot c < |K| \cdot c + \frac{N - |K|}{N} \cdot N \cdot c = N \cdot c.$$

In other words, the cost of implementing a two-stage hierarchy are less than forming an all-knowing team. This establishes the claim. \square

A.4 Proof of Theorem 1

Theorem 1 Sufficient Criterion for Optimality of Teams

Assume that the marginal cost of knowledge is constant and that a firm solves all problems it faces. For any collection of problems \mathcal{S} let \mathcal{S}^X denote the collection of problems that are strictly more complex than some problem in \mathcal{S} . Then a team with complete knowledge is the optimal organizational form, if for every knowledge set K and every subset $\mathcal{S} \subset \mathcal{P}(K)$ of problems solvable by K

$$\frac{\text{Prob}(\mathbf{p} \in \mathcal{S})}{\text{Prob}(\mathbf{p} \in \mathcal{S}^X \setminus \mathcal{S})} \leq \frac{|K|}{(N - |K|)},$$

that is, if the share of problems solvable by K relative to the correspondingly more complex problems not solvable by K is small relative to the number of units in K .

Proof: Proof by contradiction. Assume a multistage contingency sequence $\{K_1, K_2, \dots, K_{l-1}, K_l\}$ is optimal with $l \geq 2$. Since the firm solves all problems, $|K_l| = N$ has to be the knowledge set with all knowledge. Denote $|K_{l-1}| = k$. Further, denote the collection of problems that an agents with knowledge K_{l-1} solves in this knowledge sequence with \mathcal{S}_{l-1} . Let p_{l-1} denote the probability, that a problem could not be solved by the $l - 2$ first knowledge sets and that an agent with knowledge set K_{l-1} confronts this problem. Define p_l similarly. Note that $p_l = p_{l-1} - \text{Prob}(\mathbf{p} \in \mathcal{S}_{l-1})$.

I will show that the firm is better off with the contingency sequence $\{K_1, K_2, \dots, K_{l-1} \cup K_l\}$ which contradicts the optimality of $\{K_1, K_2, \dots, K_{l-1}, K_l\}$. Since K_l is the all-encompassing knowledge set, $K_{l-1} \cup K_l = K_l$.

If the firm organizes according to

$$OF = \{K_1, K_2, \dots, K_{l-1} \cup K_l\},$$

then it will not solve fewer problems and hence will not generate less revenue than if it organizes according to

$$OF' = \{K_1, K_2, \dots, K_{l-1}, K_l\}.$$

It thus suffices to show that the cost of implementing the shorter sequence are smaller than the cost of implementing the longer sequence. The difference in cost between the two sequences is given by

$$\begin{aligned} \Delta C &= (p_{l-1}c_{K_{l-1}} + p_l c_{K_l}) - (p_{l-1}c_{K_{l-1} \cup K_l}) \\ &= p_l c_{K_l} - p_{l-1} (c_{K_{l-1} \cup K_l} - c_{K_{l-1}}) \\ &= p_l (c_{K_l} + c_{K_{l-1}} - c_{K_{l-1} \cup K_l}) - (p_{l-1} - p_l) (c_{K_{l-1} \cup K_l} - c_{K_{l-1}}) \end{aligned}$$

The shorter sequence is optimal if and only if $\Delta C > 0$. So far, the description of ΔC is always valid. The first term describes the change in cost for problems solved by an agent with knowledge set K_l , and the second term describes the cost change for problems solved by an agent with knowledge set K_{l-1} , keeping the first $l-2$ stages of the contingency sequence fixed. In particular, if K_l and K_{l-1} are disjoint, then the first term reduces to zero, and $\Delta C < 0$. In contrast, if $K_{l-1} \subset K_l$, as is the case in this proposition, then the first term reduces to $p_l c_{K_{l-1}}$.

To see that criterion (2) implies $\Delta C > 0$, observe that in the setting of this proposition

$$\begin{aligned} p_l &\geq \text{Prob}(\mathbf{p} \in \mathcal{S}_{l-1}^X \setminus \mathcal{S}_{l-1}) \\ p_{l-1} &= \text{Prob}(\mathbf{p} \in \mathcal{S}_{l-1}) + p_l \\ c_{K_{l-1}} &= k_{l-1} \cdot c \\ c_{K_l} &= N \cdot c \end{aligned}$$

where c is the constant marginal cost of knowledge. Thus (2) implies

$$\begin{aligned} p_l c_{K_{l-1}} &\geq \text{Prob}(\mathbf{p} \in \mathcal{S}_{l-1}^X \setminus \mathcal{S}_{l-1}) \cdot |K| \cdot c \\ &> \text{Prob}(\mathbf{p} \in \mathcal{S}_{l-1}) \cdot (N - |K|) \cdot c \\ &= (p_{l-1} - p_l) (c_{K_{l-1} \cup K_l} - c_{K_{l-1}}). \end{aligned}$$

In other words, the shorter contingency sequence is less expensive to implement. This contradicts the optimality of the original sequence and establishes the claim. \square

A.5 Proof of Proposition 4

Proposition 4 Assume that the marginal cost of knowledge is constant. Assume that a change in uncertainty does not result in fewer firms producing. Then

1. An increase in the uncertainty about the necessary knowledge set increases the prevalence of teams.

2. An increase in the uncertainty about the value generated by a particular effort does not affect the optimal organization of a firm.

Proof: The first part of the proof argues that shifting weight in the distribution space toward the center region, where task uncertainty is highest, increases the prevalence of teams.

Formally, let f_c denote the distribution at the center of the distribution space where all $2^N - 1$ problems occur with equal probability. Theorem 1 implies that a team with complete knowledge is the optimal organizational form addressing distribution f_c . Now consider a change to the distribution space that sends every distribution f to some convex combination $\hat{f} = \alpha_f f + (1 - \alpha_f) f_c$. If f is optimally addressed by a team with complete knowledge, then so is \hat{f} . This follows from proposition 2. In other words, the prevalence of teams cannot decrease.

For some distributions close to the indifference line between teams with complete knowledge and other organizational forms, the shift from f to \hat{f} , however, implies a change in organizational form toward a team with complete knowledge. Thus, the overall prevalence of teams will increase.

For the second part of proposition 4 it suffices to observe that the uncertainty about the values generated does not interact with the organizational design of the firm:

$$\begin{aligned}
& \max_{l, \{K_1, \dots, K_l\}} E \left[\sum_{\mathbf{p} \text{ s.t. } \exists K_i: \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p}) \varrho v_{\mathbf{p}} \right] - \sum_{i=1}^l \left(1 - \sum_{\mathbf{p} \text{ s.t. } \exists K_j: j < i \text{ and } \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p}) \right) c \cdot |K_i| \\
& \cong \max_{l, \{K_1, \dots, K_l\}} E[\varrho] \cdot \left[\sum_{\mathbf{p} \text{ s.t. } \exists K_i: \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p}) v_{\mathbf{p}} \right] - \sum_{i=1}^l \left(1 - \sum_{\mathbf{p} \text{ s.t. } \exists K_j: j < i \text{ and } \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p}) \right) c \cdot |K_i| \\
& \cong \max_{l, \{K_1, \dots, K_l\}} 1 \cdot \left[\sum_{\mathbf{p} \text{ s.t. } \exists K_i: \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p}) v_{\mathbf{p}} \right] - \sum_{i=1}^l \left(1 - \sum_{\mathbf{p} \text{ s.t. } \exists K_j: j < i \text{ and } \mathbf{p} \in \mathcal{P}(K_i)} f(\mathbf{p}) \right) c \cdot |K_i|.
\end{aligned}$$

This completes the proof. \square

A.6 Proof of Proposition 5

Proposition 5 Assume $\delta < 1$. If an investment i is made, then the investment i is made ex-ante if

$$\delta \leq \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}).$$

If a valuable investment i is specific to problem \mathbf{p} , then the investment is made ex-ante if and only if the probability of problem \mathbf{p} satisfies $f(\mathbf{p}) \geq \delta$.

Proof: The condition of the statement implies

$$-\delta C_i \geq - \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}) C_i,$$

which in turn means

$$\begin{aligned}
& \sum_{\mathbf{p}} f(\mathbf{p}) r_i(\mathbf{p}) \cdot v_{\mathbf{p}} - \delta \cdot C_i \\
& \geq \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}) r_i(\mathbf{p}) \cdot v_{\mathbf{p}} - \delta \cdot C_i \\
& \geq \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}) r_i(\mathbf{p}) \cdot v_{\mathbf{p}} - \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}) C_i \\
& = \sum_{\mathbf{p} \text{ s.t. } i \text{ is valuable}} f(\mathbf{p}) [r_i(\mathbf{p}) \cdot v_{\mathbf{p}} - C_i].
\end{aligned}$$

This establishes the claim. \square

A.7 Proof of Proposition 6

Proposition 6 Assume that every distribution is profitable to address. Assume that both communication cost τ and coordination cost $coord$ depend on a technological parameter s such that $coord'(s), \tau'(s) < 0$. If $coord(s)/\tau(s)$ is increasing in s , then flat hierarchies become more prevalent. In particular, firms may optimally switch from either multi-layer hierarchies or teams to flat hierarchies.

Proof: Since every distribution is profitably addressed, every firm solves all problems it confronts. So the value generated by a firm facing the task distribution f is given by $V = \sum_{\mathbf{p}} f(\mathbf{p}) \cdot v_{\mathbf{p}} = p_a v_a + p_b v_b + p_{ab} v_{ab}$.

Without loss of generality, I restrict focus to those distributions for which B is not the optimal first-stage knowledge set. I compare the three possible organizational forms in turn.

First, the multilayer hierarchy $A- > B- > AB$ yields a higher profit than the flat hierarchy $A- > AB$ if and only if

$$c_A + \tau(1 - p_a)c_B + \tau p_{ab}(c_A + c_B + coord) < c_A + \tau(1 - p_a)(c_A + c_B + coord).$$

This is equivalent to

$$(1 - p_a)c_B + p_{ab}(c_A + c_B + coord) < (1 - p_a)(c_A + c_B + coord).$$

A change in τ has no effect on the choice between these two organizational forms. A decrease in the coordination cost by $\Delta coord$, however, decreases the left-hand side by $p_{ab}\Delta coord$ and the right-hand side by $(p_{ab} + p_b)\Delta coord$. Thus, a decrease in $coord$ is more beneficial for the flat hierarchy than for the multilayer hierarchy.

Second, the team organization AB yields a higher profit than the flat hierarchy $A- > AB$ if and only if

$$c_A + c_B + coord < c_A + \tau(1 - p_a)(c_A + c_B + coord).$$

This is equivalent to

$$c_B + \frac{coord}{\tau} < (1 - p_a)(c_A + c_B + coord).$$

If $coord(s)/\tau(s)$ is increasing in s , then so is the left-hand side of this inequality. The right-hand side, on the other hand, is decreasing in s . Hence, the threshold for p_a is decreasing in s , above which a flat hierarchy is optimal, and flat hierarchies become more prevalent.

Finally, a similar argument shows that team organization AB yields a higher profit than the multilayer hierarchy $A \rightarrow B \rightarrow AB$ if and only if

$$c_B + \frac{coord}{\tau} < (1 - p_a)c_B + p_{ab}(c_A + c_B + coord).$$

As before, if $coord(s)/\tau(s)$ is increasing in s , then so is the left-hand side of this inequality. The right-hand side, on the other hand, is decreasing in s . Hence, teams become less prevalent. This completes the proof. \square

A.8 Proof of Proposition 7

Proposition 7 Assume A is verifiable, non-specific knowledge and B is firm-specific expertise that an agent can only learn once he has joined the firm. Assume that

$$v_b, v_{ab} < v_a + c_{AB} \left(\frac{c_{AB} - c_A}{c_A} \right).$$

Then it is never optimal for a firm to distort its team organization into a two-tier hierarchy in order to use promotion as an incentive device for A workers to invest in learning B . It can be optimal for a firm to distort its team organization into a three-tier hierarchy, where workers in the “ AB ” job are shielded from solving problems that only require firm specific knowledge.

Proof: The condition of the proposition implies

$$\frac{c_A}{c_{AB}} < \frac{c_{AB} - c_A}{\bar{v}_{ab} - v_a}.$$

Therefore,

$$(*) \quad p_a < \frac{c_A}{c_{AB}} \quad \text{and} \quad (**) \quad \frac{c_{AB} - c_A}{\bar{v}_{ab} - v_a} < p_a$$

cannot be simultaneously true.

In other words, if the first and second-stage jobs in a two-tier hierarchy are sufficiently distinct to make promotion a feasible commitment device, then the underlying task distribution is such that the firm optimally organizes as a hierarchy absent of moral hazard concerns.

Next, consider promotion in a three-tier hierarchy. If a worker in an “ A ”-job invests in learning B , he generates the profit $p_a v_a + p_b v_b + p_{ab} v_{ab} - c_A$. If the worker is promoted to work at the top level of the hierarchy, then $(1 - p_a)/p_{ab}$ workers who only solve b problems

have to be employed to generate a full-time workload for a worker with an “ AB ”-job. This is worth for the firm to do, if and only if

$$v_{ab} - c_{AB} + \frac{1 - p_a}{p_{ab}} \left(\frac{p_b}{1 - p_a} v_b - c_B \right) > p_a v_a + p_b v_b + p_{ab} v_{ab} - c_A. \quad (5)$$

Recall that a team dominates the three-tier hierarchy if

$$c_{AB} < c_A + (1 - p_a)c_B + p_{ab}c_{AB}. \quad (6)$$

Both inequalities can be simultaneously true, even if the values v_{ab} and v_b are bounded by the above inequality. For example, consider the setting where $v_{ab} = 2v_a = 2v_b = 2v$ and $c_{AB} = 2c_A = 2c_B = 2c$. Then the above bound on v_{ab} and v_b reduces to

$$v < 2c.$$

Inequality (5) simplifies to

$$\frac{(p_b + p_a)p_{ab} + p_b}{p_a b} v > \frac{2p_{ab} + p_b}{p_a b} c,$$

and inequality (6) reduces to

$$2p_{ab} > p_a.$$

So for example, for the distribution $(p_a, p_b, p_c) = (1/3, 1/3, 1/3)$ a firm would optimally organize as a team. However, for this distribution promotion would offer a credible commitment device as long as

$$9c < 5v.$$

If the loss in profit from distorting the organizational form toward a hierarchy is exceeded by the benefit from having promotion as a commitment device, then the firm facing this distribution would distort its organizational form. \square