The First Time is Different: A Problem-Solving Approach to Innovation

Karen F Bernhardt-Walther*

This version: 04/29/2016

Abstract

I model innovation as a problem solving process and characterize an innovation firm's optimal organizational and managerial choices. Many of these choices differ from those optimal for other problem solving processes, e.g., at hospitals, law firms, or in manufacturing. These differences are due to a difference in the nature of the problems the firms face: Solving an innovation problem means figuring out how to do something for the first time. Solving a legal, managerial, or medical problem means matching the problem to the appropriate known solution. The results show that sustainable innovation may not only be a question of how to stimulate idea generation, how to assess which ideas to pursue, or how to incentivize workers to pursue reasonable ideas - but may also hinge on optimal organization and management of knowledge workers.

*University of Toronto. I thank Luis Garicano, Lars Stole, Emir Kamenica, Canice Prendergast, Robert Akerlof, Joseph Blasi, Alessandro Bonati, Eric Budish, John Conlon, Florian Ederer, Robert Gibbons, Gustave Manso, Bruce Weinberg, and Ben Cambpell for valuable discussions. I thankfully acknowledge the support through a Beryl Sprinkel Fellowship and a Beyster Fellowship. The paper previously circulated under the titel "A Tale of (Two) Teams: How Uncertainty and Complexity Drive Team Formation." Contact: bkaren@economics.utoronto.ca

If Ernest Hemingway, James Mitchener, Neil Simon, Franck Lloyd Wright, and Pablo Picaso could not get it right the first time, what makes you think that you will? - Paul Heckel[1]

[The "zero defects" program's] purpose is to communicate to all employees the literal meaning of the words "zero defects" and the thought that everyone should do things right the first time. - Philip Cosby[2]

1 Introduction

Innovation does not necessarily start with an idea. Instead, it often begins with a problem deemed worth solving. Yet, solving innovation problems does not share the characteristics of other problem solving processes we observe, for example, in law firms, manufacturing firms, or in hospitals: It rarely involves a knowldge-hierarchy, "management by exception," up-or-out contracts, or stark specialization. I propose that these differences arise naturally due to differences in the nature of the problem solving process: While solving a legal, managerial, or medical problem means matching the problem to the appropriate known solution, solving an innovation problem means figuring out how to do something for the first time. In other words, innovation problem solving is characterized by high task uncertainty, that is uncertainty about how and with what expertise to solve a problem. I present a framework that models both first-time and repeated problem solving processes and that allows both the task uncertainty and the complexity of problems the firm faces to vary. I show that although firms use a similar input (knowledge workers) to produce a similar output (problem solutions), they may differ substantially in terms of their organizational and managerial choices due to the underlying nature of the problem solving process. I show that organizationally team-like structures are optimal both when firms face very complex problems and when they face problems characterized by high task uncertainty, while hierarchies are optimal if some simple problems occur frequently. But managerially, firms facing repeated simple or complex problem will make similar choices, while firms facing problems for the first time will differ in their choice of contracts, decision rights, and optimal language: They are more likely to use low-powered incentive and/or relational contracts, are more likely to share decision rights with workers, and are more likely to use broad words to communicate compared to firms facing high task complexity.

Many innovations do not start with an idea but with a problem deemed worth solving. The Wright Brothers did not start with an idea of how to build a motorized heavier-than-air airplane. They started with the problem of making a better glider; they consecutively identified and solved the problems of stability, lateral control, lift, optimal wing design,

[1] As quoted in "'Sketching User Experiences"' by Bill Buxton, Elsevier Inc., 2007
[2] Quoted from "'Quality Is Free"' by Philip B. Crosby, McGraw-Hill Books, 1979

and propulsion. Eventually, they built a working airplane.[3] Thomas Edison did not invent the light bulb following a sudden revelation, but instead developed the first commercially practical incandescent light bulb by systematically resolving the problems of short life span, expensive production, and impractically high electric current requirements[4]. In the 1980s, engineers and designers at the Ford Corporation developed the then revolutionary Ford Taurus not by following particular ideas of design, form and engineering, but by successively solving problems of front wheel drive adoption, vibrations, dash board design, and marketing a car to a generation of Americans unfamiliar with the Ford brand along the way.[5]

Innovation in these examples was approached as a problem-solving process, yet the processes used were different from the problem-solving processes at law firms, in management hierarchies, or at hospitals: Edison did not use "up-or-out contracts" among the engineers at Menlo Park, the Wright brothers did not build a knowledge-hierarchy to design the various parts of their airplanes, and even though the designers and engineers at Ford brought different expertise into the project, they worked hard to cooperate and overcome their respective specializations.

Today, innovation intensive firms, such as IDEO and DesignContinuum, choose flat organizational structures based on project-focused teams. Team leaders "often emerge[d] on the basis of personal excitement" rather than expertise or tenure at the firm, and employees are assessed based on peer reviews.[6] While the problem solving firms in other industries recruit from professional schools that focus on training future problem-solvers, i.e., lawyers, managers, and doctors, innovation intensive schools hire from a variety of backgrounds: There are no professional schools for those who aspire to be innovators.

One possible explanation for these differences might be that the market place for solving innovation problems is different from solving legal, managerial, or medical problems. Demand for solutions to innovation problems may not be sufficient to sustain "specialized innovators" or, alternatively, problems worth solving may be scarce. But these are unlikely explanations: Valuable problems abound[7] and innovation remains a top priority among senior executives - in 2015 79% of senior executives named "Innovation" as their "top" or "one of their top three" strategic priorities.[8]

[3] Howard, Fred. "Wilbur and Orville: A Biography of the Wright Brothers," Knopf, 1987.

[4] Israle, Paul. "Edison: A Life of Invention." Wiley, 2000.

[5] Taub, Eric. "Taurus: The Making of the Car That Saved Ford." E. P. Dutton, 1991.

[6] Homke, Stefan. Nimgade, Ashok. "IDEO Product Development.", HBS 2007.

[7] There are many well-known valuable problems: Expanding our ability to use the body's immune system to fight cancer through immuno-therapy; finding an environmentally friendly replacement for small plastic credit/ hotel key/ ID cards that produce 75 million pounds of PVC each year that is compatible with existing infrastructure; getting better medical treatment for U.S. service members who often suffer catastrophic consequences from injuries incurred in the field; and so forth. There are also thousands of smaller scale "valuable problems:" Can an oven not warn the baker before the bread is starting to burn? Do my favorite pieces of clothing have to frazzle after repeated washing cycles? Does mechanical pencils have to leave part of the lead unused as it becomes to short to be held? And so on. Every time we are annoyed at something in our daily lives we are identifying something we'd be willing to pay for to see resolved and thus a potentially valuable problem.

[8] The Boston Consulting Groups tenth annual global survey of the state of innovation. For comparison,

In this paper I propose that the differences between innovation and other problem solving processes are instead due to the nature of the innovation process itself: Solving an innovation problem means figuring out *how* to do something for the first time. Solving a legal, managerial, or medical problem means *matching* the problem to the appropriate known solution.

In other words, general problem-solving firms (rightly) assume that the problems they face have a known solution.[9] The organization's goal is to most efficiently pass each problem to the person or team that has the appropriate training or expertise to solve it. In contrast, an innovation problem - by definition - is one that has never been solved before. No-one has seen this problem before or has been trained to solve it. Once a solution to an innovation problem has been found, it may be easy to recognize the solution as solving the problem, but finding the solution in the first place is difficult. While such solutions often combine existing technology and knowledge in novel ways[10], there is typically a high uncertainty about what a solution will look like or what expertise will eventually contribute to a solution. [11], [12]

I present a model that captures this task uncertainty underlying the innovation process and the recombination of knowledge to create new solutions. I analyze the optimal organizational and managerial response to varying degrees of task uncertainty. Critically, the model distinguishes between task uncertainty, i.e., not knowing which expertise is needed to solve a problem, from task complexity, i.e., the number of different expertise needed to solve a problem.

I show that the organizational responses to high task complexity and high task uncertainty are similar: a flat, non-hierarchical structure. But the managerial responses to task complexity and uncertainty differ in terms of optimal contracts and incentives, decision rights, and optimal language.

To capture how different combinations of expertise may be formed to solve a problem I model

during the great recession in 2008/2009 the percentage was at a ten-year low - 64%.

[9]On occasion, firms solving legal, managerial, or medical problems, too, face problems with unknown solution: special circumstance legal cases, rare medical conditions, once-in-a-lifetime managerial decisions. But in these cases the process of problem solving - who gets involved, how people involved communicate, how success is measured - is different in those firms, too.

[10]For 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. The Reebook Pump sport shoe used modified medical IV bags to implement "splint in a shoe" air bladders.

[11]Bill Buxton, a designer and principal researcher at Microsoft, summarized this: "almost anyone who has actually built a product [...] will tell you that they really didn't know enough to start until they had finished." in "Sketching User Experiences: Getting the design right and the right design", p.78

[12]Indeed it is not uncommon for such problems to be solved by someone with an unexpected background. Famously, the "longitude problem" was solved by watchmaker John Harrison, rather than - as many contemporaries expected - by astronomers or professional navigators. Dave Sobel, "Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time," 2007.
The "problem of flight" was solved by two bike mechanics, the Brothers Wright, not by an ornithologist. In hindsight we understand how these sets of expertise were useful for solving the respective problems. But before the solution was found, the usefulness of that expertise was far from clear. Similarly, there is a high uncertainty for today's valuable yet unsolved problems about which expertise will eventually solve it.

knowledge, skill, and experience as discrete units. I will use these terms interchangeably to describe human capital. The discrete units can be thought of as discrete bundles of human capital often distinguished in the real world: a college degree, being a tax lawyer or a certified accountant, or having twelve years of executive experience, being a surgeon, an anesthesiologist, or a bike mechanic.

In modeling knowledge in discrete units, I deviate from the previous literature. Typically, knowledge is modeled as a continuous parameter, making the firm's optimization problem tractable. But restrictions need to be placed on what knowledge workers can 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. Giving up continuity of the knowledge parameter keeps the model tractable without having to impose such restrictions. It allows me to consider (all) combinations of knowledge and solve the firm's optimization problem for the entire distribution space. In particular, task uncertainty and task complexity can be varied independently.

Beyond choosing a different representation of knowledge, I follow Garicano and Rossi-Hansberg [2006], Kremer [1993], and others, and in modeling a problem-solving knowledge economy: A firm is the owner of a task or problem distribution.[13] The firm has to hire workers who possess human capital in the form of knowledge in order to solve problems and to thus generate revenue. Problems are characterized by the skills needed to solve them. Different tasks may require different skill sets, and more knowledgeable workers are more expensive to hire. I assume that the firm knows its distribution of problems, but that it is uncertain about the exact skills required to solve any particular task. This uncertainty is not resolved until the task is successfully solved. The firm can attempt to solve a problem 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 the first half of the paper, I investigate the firm's organizational choice in response to its task distribution. The firm's optimization problem consists of choosing the maximal number of problem solving attempts it is willing to conduct and the combination of human capital the workers should have who perform each attempt. For some distributions the 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 solve some of the problems 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. However, for many other task distributions it is optimal for the firm to attempt all problems only

[13] Throughout the paper I use the terms "problems" and "tasks" interchangeably.

once using a flat, non-hierarchical organization where a very knowledgeable worker or a team of workers use a large knowledge set to tackle all problems the firm draws from the distribution.

I find that flat organizational structures are optimal both when the firm faces a distribution with high task uncertainty and when it faces a distribution with high task complexity. Multilayer hierarchies as in Garicano and Rossi-Hansberg [2006] are optimal when relatively simple problem occur relatively frequently.

In the second half of the paper, I analyze the firm's managerial choices in response to its task distribution. I look in turn at stylized models of contracts, decision rights, and optimal language. I find that firms facing high task uncertainty are more likely to use low-powered incentive and/or relational contracts, are more likely to share decision rights with workers, and are more likely to use broad words to communicate compared to firms facing high task complexity. In other words, even though both kinds of firms optimally choose the same organizational structure, their optimal managerial practices differ.

The paper's contribution is three-fold. The paper's main contribution is to the literature on innovation. The existing literature on innovation has mostly focused on innovation as a contractual problem: Firms find it difficult to contract on innovation activity, because inputs cannot be monitored and outputs exhibit a high variance. Manso [2007], Ederer [2008] and others show how firms can design long-term contracts to motivate innovation activity nonetheless. In contrast, this paper emphasizes the organizational and managerial choices a firm must make to focus on innovation activity. Demonstrating that the optimal structure differs between innovation and routine activities, this paper highlights the organizational challenge of implementing innovation.

Second, to the extent that a flat organizational structure can be interpreted as a team, this paper contributes to the team production literature. This strang of the literature has considered firm, industry, and individual characteristics to understand when team production is beneficial, see DeVaro and Kurtulus [2006] for a review. However, with the exception of Boning et al. [2003], problem characteristics have mostly been ignored. My results suggest that problem characteristics may indeed be crucial determinants of the benefits of team productions and that the optimal managerial choices, such as group incentives or shared decision making, likely depend on the firm's underlying task distribution. When empirically investigating the relationship between team productivity on one hand and organizational and managerial choices on the other hand, the underlying task distribution should to be taken into account.

Last but not least, the paper contributes conceptually to the organizational design literature in two ways. First, the model distinguishes between task uncertainty and task complexity and emphasizes their distinct impact on organizational and managerial choices. In contrast, the existing literature has either not distinguished between complexity and uncertainty or has assumed that they co-vary (Garicano and Rossi-Hansberg [2006], Cremer et al. [2007]). Second, the model allows for a simultaneous consideration of the entire distribution space and thus for comparative static on the prevalence of organizational forms.

The remainder of the paper is structured as follows. I formally introduce the model in section 2. In section 3 I present a complete solution for the special case of an economy with two units of expertise. In many ways, section 3 presents the core of the paper. It lays out the optimal organizational form for the entire distribution space and it develops the intuition for why a flat organizational structure is the optimal organizational response to both high task uncertainty and to high task complexity. Section 4 studies a firm's managerial choices and establishes that firms facing high task uncertainty and high task complexity, respectively, optimally choose different managerial practices. Section 5 analyzes further differences between complexity- and uncertainty-facing firms. Finally, section 6 concludes the paper.

2 The Model

2.1 Set-Up

A firm is the owner of a distribution of problems[14]. Revenue is generated by solving these problems. In order to do so the firm hires workers who possess knowledge. The firm's optimization problem is which workers to employ, which workers should work collaboratively,i.e., jointly on the same problems, and who should report to whom. For now, I consider the partial equilibrium model. In other words, I assume that the workers are endowed with a particular knowledge, that is, they have made their human capital investment decisions. I also assume that firms face equilibrium wages and no individual firm's hiring decisions affect these wages. For the entire paper, I set aside any conflicts of interest workers might have with the firm as well as any considerations of private benefits some workers might receive from having big teams or less knowledgeable workers report to them. I have no doubt that these factors affect organizational choices in firms, but with this work I want to highlight another factor: the distribution of tasks a firm faces.[15]

Knowledge, expertise, or skill in this economy is learned and used in discrete units, labeled A, B, C, and so forth. 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.

Agents posses knowledge, but due to bounded cognitive capability, that knowledge is limited.

[14] Throughout the paper I use the terms "tasks" and "problems" interchangeably.

[15] These assumptions reduce workers in this model to "packaging" that contains knowledge the firm needs. One might therefore think that the model is not specific to human capital and may apply to other assets the firm uses in production. However, I found that reasonable parameters for relative cost and relative value for non-human capital typically result in pathological results. Nonetheless, non-human-capital examples might exist where the model can be usefully applied. I would be very interested to hear about any such examples.

However, agents can work together as a team and thus combine their expertise. For example, if worker 1 possesses knowledge $\{A\}$ and worker 2 possesses $\{B\}$ then jointly they provide $\{A, B\}$. The cost of hiring agents depends on the knowledge $\mathcal{K}$ the agents have, and is denoted by $c_{\mathcal{K}}$. The firm only pays agents for periods in which they work for the firm, i.e., periods in which the agents attempt to solve a problem for the firm.[16] I assume that it is always more expensive to hire a more extensive knowledge set, i.e., $\mathcal{K}_1 \subset \mathcal{K}_2$ implies $c_{\mathcal{K}_1} \leq c_{\mathcal{K}_2}$. Unless otherwise stated, I do not impose any other restriction on the cost of hiring knowledge. In particular, in the presence of coordination cost, say, $c_{\mathcal{K}_1 \cup \mathcal{K}_2}$ may be larger than $c_{\mathcal{K}_1} + c_{\mathcal{K}_2}$, while in the case of synergies, $c_{\mathcal{K}_1 \cup \mathcal{K}_2}$ may be smaller than $c_{\mathcal{K}_1} + c_{\mathcal{K}_2}$ for disjoint knowledge sets $\mathcal{K}_1$ and $\mathcal{K}_2$.

The firm uses the workers' knowledge to solve tasks. Tasks may require any combination of knowledge, and are labeled accordingly. For example, $\mathfrak{t} = a$ is the task that only requires knowledge A, whereas $\mathfrak{t} = abc$ denotes the task that requires A, B, and C to be solved. 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 agent(s) with a knowledge set that contains the minimal necessary knowledge set can solve the problem equally well. For example, a team with the knowledge set $\{A, B\}$ can solve the three problems a, b, and ab, but an agent with knowledge set A can only solve a.

The firm knows the probability $p_{\mathfrak{t}}$ with which each possible task $\mathfrak{t}$ occurs. It chooses its optimal organizational form in response to given labor market wages and the distribution of tasks it faces. A main contribution of this paper is to show how the task distribution affects the organization (and management) of the firm. I will say a task is more *complex* if it requires more units of knowledge to solve. Also, a firm faces a *higher task uncertainty* if there is more uncertainty about which knowledge is required to solve a given problem. Uncertainty is highest for the uniform distribution.[17] Uncertainty is lowest if the firm faces a single-problem distribution, where $p_{\hat{\mathfrak{t}}} = 1$ for some task $\hat{\mathfrak{t}}$ and $p_{\mathfrak{t}} = 0$ for all other task $\mathfrak{t}$.[18]

Note that the possibility of different combinations of knowledge is what distinguishes this model from the previous literature. In a single-parameter model with a continuous knowledge parameter, we typically see either knowledge that correspond to disjoint knowledge sets, such as $\{A\}$, $\{B\}$, $\{C\}$, and so forth or to nested knowledge sets, such as $\{A\}$, $\{A, B\}$, $\{A, B, C\}$, see for example Garicano [2000], Garicano and Rossi-Hansberg [2006]. In contrast, this model allows for all of $\{A\}$, $\{B\}$, $\{C\}$, $\{A, B\}$, $\{A, C\}$, etcetera and the corresponding tasks a, b, c, ab, ac, and so on to be considered. The number of tasks an agent or

[16] This is consistent with a spot labor market, with allowing part time work, or with firms owning sufficiently many urns of the same distribution so that integer constraints do not bind.

[17] In an economy with N units of knowledge, the uniform distribution is $p_{\mathfrak{t}} = \frac{1}{2^N - 1}$.

[18] To define the uncertainty as the variance of the task distribution, we need to be able to "add" tasks, i.e., compute expressions like $\frac{1}{2}a + \frac{1}{2}ab$. One way to implement this is to consider the simplex in $\mathbb{R}^{2^N - 1}$ if there are N units of knowledge and assign each task to a vertex in the simplex. For example, if there are only $N = 2$ units of knowledge, the simplex in $\mathbb{R}^3$ is spanned by $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. Each of the three possible tasks a, b, and ab, can be assigned to one of these vertices. The uniform distribution $1/3(a, b, ab)$ has the expected value $1/3(1, 1, 1)$ and a variance of $2/3$. One can show that the uniform distribution maximizes the variance.

a team of agents with knowledge set $\mathcal{K}$ can solve grows exponentially with the size of the knowledge set $|\mathcal{K}|$.[19] This combinatorial growth of potential combinations of knowledge will make a flat organizational structure optimal for firms facing distributions with high task uncertainty. Combining "every kind of expertise with every other expertise" is relevant in practice, especially in the case of sustainable innovation.

A task is solved when agents who attempt solving the task have the necessary knowledge to solve it.[20] For any given problem, the firm generally does not know what knowledge is needed to solve the task until the task is solved. Attempting to solve a problem takes agents one period. Problems may be attempted more than once. I assume that failing to solve particular problem does not reveal information beyond the observation that agents with this particular knowledge set are not able to solve the problem. For example, if an agent with knowledge $\{A\}$ fails to solve a problem he does not learn whether the problem is of type b or ab. In practice, of course, failing to solve a problem often reveals additional information. However, I impose this simplifying assumption for now to separate the organizational impact such diagnostic information has on the organizational structure, which I discuss in 5.2.

If a problem is solved, revenue is generated. The revenue generated only depends on the type of problem $\mathfrak{t}$ solved. Accordingly, the revenue is denoted by $v_{\mathfrak{t}}$. I generally assume that more complex tasks are more valuable and that every task by itself is worth doing, i.e., $v_{\mathfrak{t}} > c_{\mathcal{K}}$, where $\mathcal{K}$ is the minimal knowledge set able to solve $\mathfrak{t}$. I do not otherwise impose any restrictions. In particular, there is no depreciation over time in the value generated by solving a problem.

The firm's optimization problem is to choose in which order which agents should attempt to solve problems drawn from the distribution $p_{\mathfrak{t}}$ in order to maximize profits. I call a solution to this optimization problem *organizational form*.[21,22]

The timing is as follows.

---

[19]In fact, any unit of knowledge in $\mathcal{K}$ may or may not be required to solve a particular task. Not counting the "empty" task, the number of problems a knowledge set $\mathcal{K}$ can address is $2^{|\mathcal{K}|} - 1$.

[20]I abstract away from the cognitive processes of actually finding a solution.

[21]Each 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.

[22]If the problems a firm solves do not depend on the firm's choice of organizational form, then maximizing profit per distribution 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

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

Time Period	Event
T = 1	Firm learns its problem distribution.
T = 2	Firm chooses its organizational form.
T = 3	The firm hires appropriate agents.
T = 4	Production takes place:
T = 4.1	The firm draws a problem, agents solve the problem they are facing if they can, passing on or disregarding those they can't.
⋮	The firm draws a problem, agents solve the problem they are facing if they can, passing on or disregarding those they can't.
	Revenue is generated in each of these sub-periods, and profits are realized.

To simplify notation, I write the knowledge sets without brackets from here on forward.

2.2 Structure of the Optimal Organizational Form

For the following discussion it is useful to formally establish the structure of a firm's optimal organizational form:

Lemma 1 An optimal organizational form always takes the form of $\mathcal{K}_1 \to \mathcal{K}_2 \to \mathcal{K}_3 \to \ldots \to \mathcal{K}_l$, where $\mathcal{K}_1$ represents knowledge set provided by the first agent(s) to address a problem drawn from the task distribution. Any problems not solved by the agent(s) with knowledge $\mathcal{K}_1$ are then attempted by agent(s) with knowledge $\mathcal{K}_2$ and so forth. Any task that remains unsolved after agents with knowledge $\mathcal{K}_l$ have attempted to solve it is discarded. Moreover, if $i > j$ then $\mathcal{K}_i \nsubseteq \mathcal{K}_j$ for all i, j.

The lemma essentially states that the optimal organizational form can be expressed as a sequence of knowledge sets that in turn attempt to solve tasks not previously solved. [23] The formal proof is relegated to the appendix. Because agents at every instance of the knowledge sequence solve some tasks, the share of tasks that is addressed at each stage is decreasing and fewer agent-time needs to be hired at each stage. Thus, each knowledge sequence larger than one has a natural interpretation as a hierarchy: Problems are only passed in one direction and there are fewer urns of tasks addressed by agents at each stage than in the preceding one.

[23] Note in particular, that there is no splitting and no joining of task distributions: In principle, it might be optimal for agent(s) to pass the problems they cannot solve to agent(s) with different knowledge sets. This, however, can only be singular optimal if the agent(s) learn something about the tasks they cannot solve. For now, I exclude this scenario, because I want to isolate how the task distribution affects the firm's organizational choice from the impact of diagnostic capabilities. I allow for diagnostic and discuss the resulting organizational impact in 5.2. Also, it might be possible that a firm is better off by combining the unsolved problems from agents with different knowledge sets. However, since firms only pay for marginal access to knowledge, the firm can duplicate the organizational structure for both remainder task distributions without reducing its profit.

Because a wide range of knowledge sequences can be optimal, hierarchies in this model sometimes contain teams at some level. This is not unheard of in the real world, where a team of engineers might report to a single manager or a partner in a consulting firm might supervise a team of associates. As a consequence, hierarchies and teams are not exclusive organizational choices in this model. To clarify the distinction, I will talk of single-layer and multi-layer organizational forms. In particular, I am interested in the question of when a single-layer all-knowing flat organizational structure is the optimal organizational form.

In the prior literature, the distribution of tasks had to be constrained to make the model tractable. For example, Garicano [2000] considers the two scenarios where, one, tasks are of nested complexity, i.e., of type "a", "ab", "abc", and so forth, or, alternatively, tasks are of constant complexity, i.e., "a", "b", "c", and so forth. In both cases, Garicano [2000] assumes that problems occur with exponentially decreasing frequency. One advantage of the model presented here is that it is tractable without constraints on the task distributions. Hence comparative statics on the prevalence of different organizational forms over the whole distribution space are feasible. The following lemma is a useful tool in that discussion.

Lemma 2 The region in the distribution space where a particular organizational form is optimal is convex.

2.3 Organizational Response to High Complexity and High Uncertainty

The model can be made more tractable by imposing a structure on the cost of knowledge or by constraining the collection of tasks considered. In this subsection I use the former and in the following section the latter approach to show that complexity and task uncertainty independently result in team formation.

Proposition 3 Assume that there are N units of knowledge in the economy and that the tasks are valuable enough such that firms facing any distribution will solve some tasks. Also, assume that the cost of accessing knowledge is linear in the number of units, i.e., $c_{\mathcal{K}} = C \times |\mathcal{K}|$. Then firms facing either

a. the single-problem distribution for the most complex problem, i.e., $p_{abcd...} = 1$ and $p_{\mathfrak{t}} = 0$ for all other $\mathfrak{t}$, or

b. the uniform distribution, i.e., $p_{\mathfrak{t}} = \frac{1}{2^N - 1}$,

optimally organize as an all-knowing flat organizational structure or do not produce at all.

The first part of this result is tautologically true - no other knowledge set can solve the most complex problem. Since the firm knows that this is the only problem it faces, it will either employ agents who jointly provide all knowledge in the economy or not produce at all. For the second part of the proposition I show that the all-knowing team is cheaper than any knowledge sequence of length bigger than one that ends in the complete knowledge set.

In the second case, complexity however is not the driver of the organizational choice: only a tiny fraction of the tasks the firm faces, namely $1/(2^N - 1)$, requires all knowledge. Instead, no smaller knowledge set is useful "enough" to first attempt to solve problems. The share of tasks that agents with a smaller knowledge set could solve is too small to make it optimal for the firm.[24]

One might argue that even though complex problems occur infrequently in the second result of proposition 3, they might still occur frequently enough to induce the all-knowing flat organizational structure to be optimal. The following result shows that even if the complexity is small and constant among all problems, the all-knowing flat organizational structure is still the optimal organizational form if a sufficient variety of tasks occur.

Proposition 4 Assume that there are N units of knowledge in the economy and that the tasks are valuable enough such that firms facing any distribution will solve some tasks. Also assume that the cost of accessing knowledge is linear in the number of units, i.e., $c_{\mathcal{K}} = C \times |\mathcal{K}|$. Then firms facing a uniform distribution of tasks with constant complexity $k \geq 2$ optimally organize as a single-layer all-knowing structure.

For example, if the only tasks the firm faces are of complexity 2, i.e., ab, ac, ad, ..., ef, ... but each of these tasks occurs with equal probability, then it is optimal for the firm to have agents work together such that every task is attempted with the complete knowledge set $A, B, .., E, F, ...$ In other words, even absent variations of complexity and absent any complex problems the firms might face, sufficient uncertainty about the expertise required to solve a task results in single-layer flat organizational structure.

The intuition for this result lies in the combinatorial growth of problems that can be solved with increasing knowledge. For example, A, B only solves one task, namely ab. Adding another unit of expertise means three tasks can be addressed: A, B, C can solve all of ab, ac, and bc. Four units of knowledge can solve 6 tasks and so on. This is the advantage of using discrete units of knowledge: To see the value of "combining expertise" when facing unknown tasks and its impact on the organizational choice of the firm.

[24] One may be tempted to draw a connection to the convex returns to knowledge - after all, a constant marginal increase in the cost of of a larger knowledge sets yields an increasing return in term of the tasks that can be solved. However, the relationship is not as straight forward, since the remainder distribution after a first few attempts to solve a task may well be concave, thus making a multi-stage organizational form optimal. Convexity is necessary but not sufficient. A sufficient condition for the all-knowing flat organizational structure to be optimal is as follows: If for every knowledge set $\mathcal{K}$ with $|\mathcal{K}| = k$ and every (sub)collection $\mathcal{S}$ of tasks solvable by $\mathcal{K}$

$$\begin{aligned} & Prob\,(\mathrm{t} \in \mathcal{S}) \cdot (N-k) \\ \leq \; & Prob\left(\mathrm{t} \text{ strictly more complex than some } \mathrm{t}' \in \mathcal{S} \text{ and } \mathrm{t} \notin \mathcal{S}\right) \cdot k. \end{aligned}$$

3 Optimal Organizational Response for $N = 2$

Next, I present a complete solution for the firm's optimal organizational choice for $N = 2$. While the case of $N = 2$ is no doubt a rather stylized representation, its graphical solution highlights some key intuitions.

There are now two simple tasks a and b, and one complex task ab. I assume that agents can only learn one unit of knowledge and so ab must be solved by an all-knowing team of agents. The cost of hiring the agents are denoted by c_A, c_B, and c_{AB} respectively, where $c_{AB} > max\{c_A, c_B\}$ but no other relationship is imposed. The revenue generated from solved tasks is denoted by v_a, v_b, and v_{ab}, respectively, where $v_{ab} > max\{v_a, v_b\}$.

3.1 Distribution Space and Organizational Forms

The distribution of problems a firm faces is characterized by the probabilities of these problems, p_a, p_b, and p_{ab}. The space of all such distributions 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.

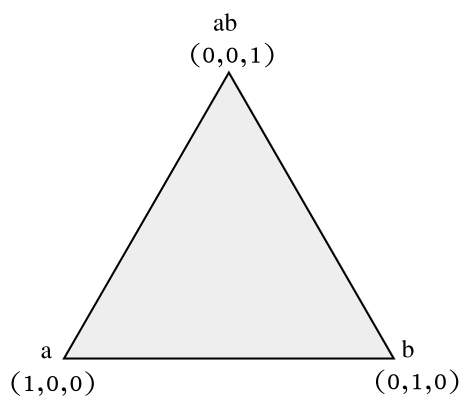


Figure 1: Each point in the triangle spanned by $p_a + p_b + p_{ab} = 1$ in $\mathbb{R}^3$ corresponds to a problem distribution (p_a, p_b, p_{ab}). The corners of the triangle correspond to the three (pure) distributions where only one of the three problems occurs. The simplex represents the space of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

Addressing a problem drawn from a distribution (p_a, p_b, p_ab), the firm has to choose in which order which agents should attempt to solve the problem. Lemma 1 implies that there are only ten potentially optimal knowledge sequences. Table 1 lists the ten knowledge sequences, together with their cost and value functions.

In sections 3.3 and 3.5 below I identify the firm's optimal organizational structure for each probability distribution (p_a, p_b, p_{ab}). To do so, I will compare all ten potentially optimal knowledge sequences and for each characterize the region within the distribution space where

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

it is optimal.

3.2 First-Best Outcome and Second-Best Trade-Off

The first-best outcome for a firm is to address every problem with the minimal knowledge set that can solve it. Problem a would be solved by an agent with knowledge A, generating a profit of $v_a - c_A$. Similarly, the problem b would be solved by an agent with knowledge B, and problem ab would be solved by two agents with the firm making a profit of $v_{ab} - c_{AB}$. A firm facing the distribution (p_a, p_b, p_{ab}) would generate the profit

$$\Pi_{FB}(p_a, p_b, p_{ab}) = p_a \cdot (v_a - c_A) \,+\, p_b \cdot (v_b - c_B) \,+\, p_{ab} \cdot (v_{ab} - c_{AB}).$$

This is the first-best benchmark. It corresponds to a case where a firm can costlessly diagnose each problem it draws and direct it to an agent or a team of agents with the appropriate knowledge. The extent to which the realized profit deviates from this first-bench mark reflects the additional cost to the firm due to the uncertainty about the type of the problem at hand.

When the firm only knows the probability with which each problem type occurs, it must choose which knowledge sets agents should have who are asked to attempt problems drawn from the distribution. Choosing a team of two agents with knowledge AB implies that all problems are solved, but it is costly to do so. Choosing an agent with one unit of knowledge is cheaper but may still be inefficient if the agent spends most of his time attempting problems he cannot solve. This tension between too much and too little knowledge determines the firm's second-best pay-off and organizational form.

3.3 Optimal Organizational Form when v_a, $v_b > c_{AB}$ and All Problems are Solved

Next, I solve the optimization problem for all distributions under the assumption $v_a, v_b > c_{AB}$. If the revenue from solving any problem is larger than the cost of the most expensive team, then addressing a distribution with the knowledge set AB generates a positive profit for all distributions. Therefore, production is profitable for each distribution and all problems get solved for each distribution.[25] In particular, the optimal organizational form is among the five listed in the lower half of table 1.

Comparing the profit generated by these knowledge sequences pairwise yields a set of inequalities which characterize the regions in which each of these five knowledge sequences is optimal. For example, a team and the knowledge sequence AB dominates the knowledge sequence $A \to AB$ if

$$\begin{aligned}
& \Pi(AB) > \Pi(A \to AB) \\
\Leftrightarrow \quad & p_a v_a + p_b v_b + p_{ab} v_{ab} - c_{AB} > p_a v_a + p_b v_b + p_{ab} v_{ab} - c_A - (1 - p_a) c_{AB} \\
\Leftrightarrow \quad & c_{AB} < c_A + (1 - p_a) c_{AB} \\
\Leftrightarrow \quad & p_a c_{AB} < c_A \\
\Leftrightarrow \quad & p_a < \frac{c_A}{c_{AB}}.
\end{aligned}$$

This is intuitive: For large p_a it would be wasteful to have knowledge B present while many a problems are solved. Thus, for large p_a the sequence $A \to AB$ is better. But for small p_a it would be inefficient for an agent with knowledge A to attempt to solve problems most of which he cannot solve by himself. In these cases, a team with knowledge AB is better.

A similar comparison with the other three potentially optimal knowledge sequences shows that a team and the knowledge sequence AB is optimal if and only if all of the following four inequalities are satisfied

$$\begin{aligned}
& \Pi(AB) > \Pi(A \to AB), \text{ i.e.,} \quad p_a < \frac{c_A}{c_{AB}} \\
& \Pi(AB) > \Pi(B \to AB), \text{ i.e.,} \quad p_b < \frac{c_B}{c_{AB}} \\
& \Pi(AB) > \Pi(A \to B \to AB), \text{ i.e.,} \quad p_a(c_B + c_{AB}) + p_b c_{AB} < c_A + c_B \\
& \Pi(AB) > \Pi(B \to A \to AB), \text{ i.e.,} \quad p_b(c_A + c_{AB}) + p_a c_{AB} < c_A + c_B
\end{aligned}$$

[25] To 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.

The inequalities that determine the region of optimality can be computed similarly for the other four knowledge sequences.[26]

Figure 2 a) shows the regions characterized by these inequalities for each of the five contingency sequences. Figure 2 b) shows the corresponding organizational forms that are optimal.

Most of figure 2 is intuitive. Near a vertex one particular problem occurs very frequently. It is efficient to solve all of these very frequent problems first. For example, all distributions in the proximity of $(1,0,0)$ are optimally first addressed by an agent with knowledge A. Along the edges, at most one type of problem remains unsolved after the first problem is solved, hence there is at most a two-stage knowledge sequence. Somewhere along the edge there must be a cut-off where the firm is indifferent between the two organizational forms optimal near the two respective vertices. For example, along the bottom edge a firm is indifferent between $A \to B$ and $B \to A$ at the point where

$$\begin{aligned} & \Pi(A \to B) \;=\; \Pi(B \to A) \\ \Leftrightarrow\quad & p_a v_a + p_b v_b - c_A - (1-p_a)c_B \;=\; p_a v_a + p_b v_b - c_B - (1-p_b)c_A \\ \Leftrightarrow\quad & p_a c_B = (1-p_a)c_A \\ \Leftrightarrow\quad & p_a = \frac{c_A}{c_A + c_B} \end{aligned}$$

Similarly, along the right hand side edge, a firm is indifferent between $B \to AB$ and AB at the point where

$$\begin{aligned} & \Pi(B \to AB) \;=\; \Pi(AB) \\ \Leftrightarrow\quad & p_b v_b + p_{ab} v_{ab} - c_B - (1-p_b)c_{AB} \;=\; p_b v_b + p_{ab} v_{ab} - c_{AB} \\ \Leftrightarrow\quad & p_b \;=\; \frac{c_B}{c_{AB}}. \end{aligned}$$

[26] In detail, $A \to AB$ is optimal if and only if the following four inequalities are satisfied

$$\begin{aligned} \Pi(A \to AB) \;&>\; \Pi(AB), \text{ i.e.,} \quad p_a > \frac{c_A}{c_{AB}} \\ \Pi(A \to AB) \;&>\; \Pi(B \to AB), \text{ i.e.,} \quad (p_a - p_b) > \frac{c_A - c_B}{c_{AB}} \\ \Pi(A \to AB) \;&>\; \Pi(A \to B \to AB), \text{ i.e.,} \quad \frac{p_b}{1-p_a} < \frac{c_B}{c_{AB}} \\ \Pi(A \to AB) \;&>\; \Pi(B \to A \to AB), \text{ i.e.,} \quad p_b < \frac{c_B}{c_{AB}} \end{aligned}$$

and $A \to B \to AB$ is optimal if and only if all of

$$\begin{aligned} \Pi(A \to B \to AB) \;&>\; \Pi(AB), \text{ i.e.,} \quad p_a(c_B + c_{AB}) + p_b c_{AB} > c_A + c_B \\ \Pi(A \to B \to AB) \;&>\; \Pi(A \to AB), \text{ i.e.,} \quad \frac{p_b}{1-p_a} > \frac{c_B}{c_{AB}} \\ \Pi(A \to B \to AB) \;&>\; \Pi(B \to AB), \text{ i.e.,} \quad \frac{p_a}{1-p_b} > \frac{c_A}{c_{AB}} \\ \Pi(A \to B \to AB) \;&>\; \Pi(B \to A \to AB), \text{ i.e.,} \quad \frac{p_a}{p_b} < \frac{c_A}{c_B} \end{aligned}$$

are satisfied. The inequalities for the remaining two sequences $B \to AB$ and $B \to A \to AB$ hold symmetrically.

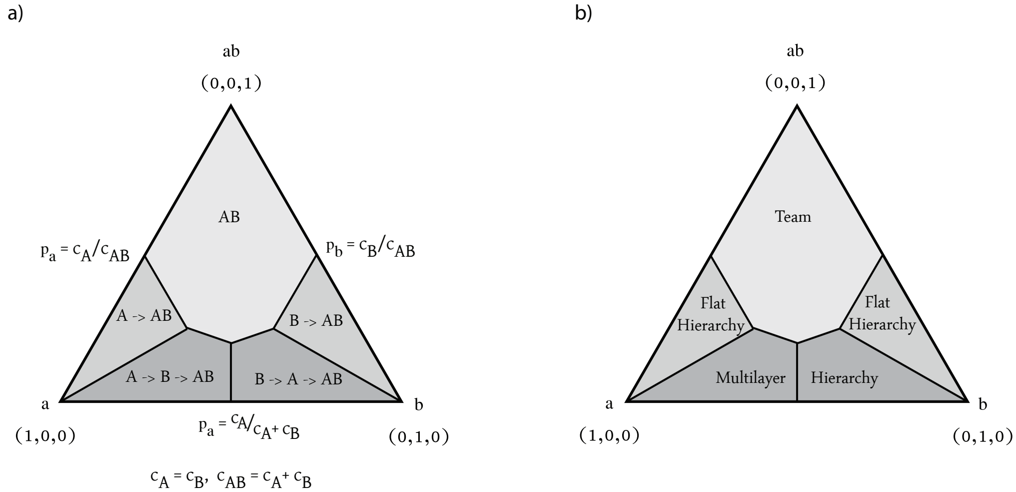


Figure 2: Points in the simplex represent distributions of problems p_a, p_b, p_{ab}. Problems are solved if agents with corresponding knowledge addresses the problem. The firm hires agents with knowledge sets. The firm decides which agents to hire, and in which order agents should attempt to solve problems. 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.

The key to understanding the solutions in the interior is to realize that problems that remain unsolved have a remainder distribution that corresponds to a point on the opposite edge of the vertex. For example, after all a problems of the original distribution $(2/3, 1/6, 1/6)$ have been solved, the remainder distribution is of the form $(0, 1/2, 1/2)$ and hence a point on the edge between $(0, 1, 0)$ and $(0, 0, 1)$. The areas where $A \to AB$ and $A \to B \to AB$ are optimal therefore correspond to the parts on the right hand edge where AB and $B \to AB$ are optimal, respectively.

The crucial insight from figure 2 is the wide prevalence of the team AB at the center of the distribution space. Consider, for example, the case where all problems occur with equal probability, i.e., one third. If $c_A = c_B = 1/2c_{AB}$ (as drawn in figure 2), this distribution is optimally addressed with a team AB. This is not due to the complexity of the problems: two thirds of all problems are simple and only require one unit of knowledge. If the simple problem was always the same, for example $(2/3, 0, 1/3)$, the distribution would optimally be addressed by a two-stage knowledge sequence, that is, a hierarchy.

Instead, no smaller knowledge set is useful "‘enough"’ to first attempt to solve problems: an agent with knowledge A (or B) could only solve one third of all problems. In this case, doubling the knowledge triples the number of problems that can be solve. Combinatorial complementarity translates into economic complementarity.

3.4 Extrapolating from the Prior Literature

To appreciate the contribution of figure 2 consider what one might have extrapolated from the existing literature.

Lemma 2 observed that the region where a particular knowledge sequence is optimal must be convex. Using this observation, the results from the prior literature can be extrapolated to the edges and to the interior of the distribution space.

First, Garicano [2000] addresses the case of nested complexity, i.e., problems of type "a", "ab", "abc", and so forth, as well as the case where tasks are of constant complexity, i.e., "a", "b", "c", and so forth. In both cases, Garicano [2000] assumes that problems occur with strictly decreasing frequency, and shows that hierarchies $A \to AB \to ABC \to ...$ and $A \to B \to C \to ...$ are optimal, respectively. For the $N = 2$ case, Garicano [2000] implies that $A \to B$ and $B \to A$ is optimal along the lower edge between the $(1, 0, 0)$ and $(0, 1, 0)$ vertices[27], and $A \to AB$ is optimal along the lower part of the other two edges of the simplex.

Next, the literature of optimal team formation focuses agents cooperating to solve a repeated complex task (Kremer [1993], Becker and Murphy [1992]). In this model, a repeated complex task corresponds to the distribution at the vertex $(0, 0, 01)$ where AB is optimal. Combining the team literature and the results from Garicano [2000] yields the optimal organizational form for both vertices connecting to $(1, 0, 0)$. In particular, AB must be optimal near the top vertex and $A \to AB$ near the bottom vertex $a = (1, 0, 0)$. The point of indifference is characterized by equality of both profit functions, i.e., by $p_a = \frac{c_A}{c_{AB}}$.

Thus, the prior literature determines the optimal organizational forms along the boundary of the distribution space, as shown in figure 3 a).

To extrapolate to the interior of the distribution space, we can project remainder distributions onto the opposite edge as before: After an agent with knowledge A has solved all a problems, the remainder distribution only contains problems of type b and ab. So the remainder distribution is part of the edge between $(0, 1, 0)$ and $(0, 0, 1)$. This projection identifies the interior regions where knowledge sequences $A \to AB$ and $A \to B \to AB$ are optimal, respectively, according to whether the remainder distribution after problem a has been solved is optimally addressed by AB or $B \to AB$. The analogous argument identifies the regions where $B \to AB$ and $B \to A \to AB$ are optimal, respectively. If $c_A = c_B$, this symmetry determines the boundary between regions where A is the optimal first knowledge

[27] Near the vertex $(1, 0, 0)$, a knowledge sequence of the form $A \to B$ must be optimal, while in proximity of $(0, 1, 0)$ the knowledge sequence $B \to A$ is optimal. The point of indifference is characterized by equality of both profit functions, i.e.,

$$p_a v_a + p_b v_b \, - \, c_A - (1 - p_a) \cdot c_B \; = \; p_a v_a + p_b v_b \, - \, c_B - (1 - p_b) \cdot c_A$$

which implies

$$p_a = \frac{c_A}{c_A + c_B}.$$

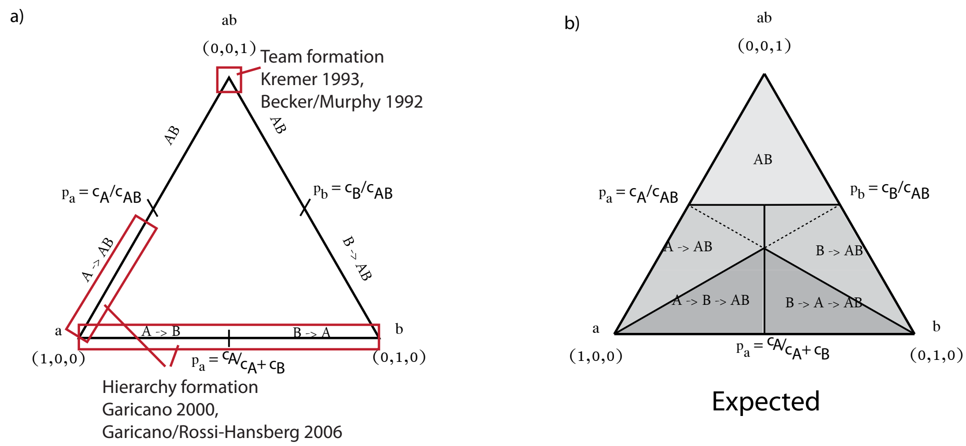


Figure 3: a) Optimal organizational forms along the edges of the simplex as derived from prior literature. b) Expected optimal organizational forms based on extrapolating the optimal organizational forms on the edge. The extrapolation uses (i) that regions where an organizational form is optimal must be convex, and (ii) projection of the remainder distribution onto the opposite edge. Compare with the actual organizational form as shown in figure 2

set and those where B is the optimal in the first stage of the knowledge sequence.

Finally, convexity implies that AB must be optimal for any distribution that is a convex combination of two distributions for which AB is optimal. So in particular, any convex combination of a distribution on the upper half of the right hand edge and a distribution on the upper half of the left hand edge of the distribution space must be optimally addressed by AB.

The resulting *extrapolated* optimal organization for each distribution in the simplex is shown in figure 3 b). Compare these forms derived from results in the prior literature with the actual optimal organizational forms as shown in figure 2. The main deviation is in the center of the distribution space where task uncertainty is large. As figure 2 shows: Teams optimally address these distributions.

3.5 General Solution for Optimal Organizational Form

Concluding the discussion of the $N = 2$ case, I present the general solution for optimal organizational forms, potentially allowing for $v_a < c_{AB}$ or $v_b < c_{AB}$. In this case, it is possible that not all problems get solved for every distribution. Therefore, all ten potentially optimal knowledge sequences have to be considered.The general solution is shown in figure 4. 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.

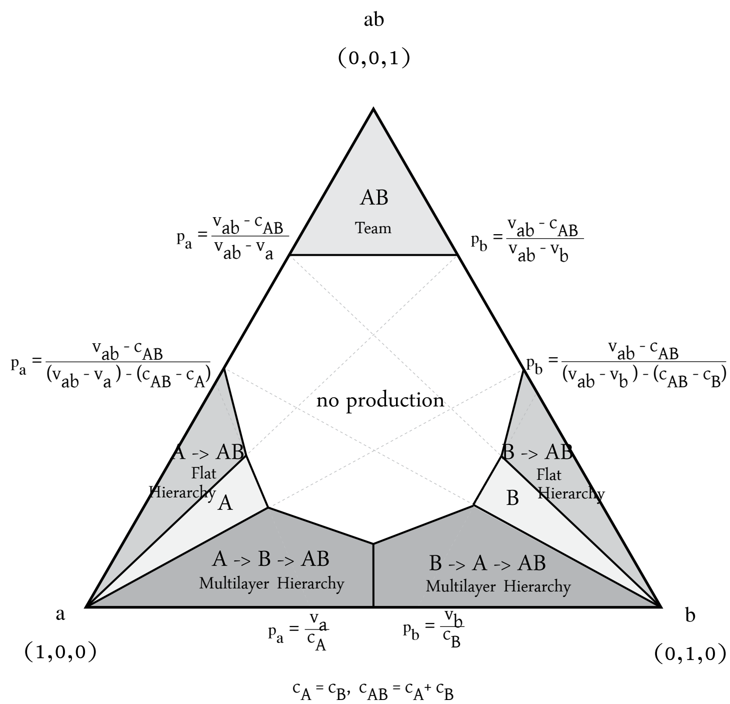


Figure 4: If $v_a, v_b < c_{AB}$, then not all firms solve all problems they face. All ten contingency sequences listed in table 1 are potentially optimal. The optimal knowledge sequence for each distribution is inscribed in the corresponding region.

The key take-away from figure 4 is that firms may find it optimal not to produce even though each task they face by itself would be profitable to solve. Indeed, the parameters assume that each of the three problems is valuable, i.e., $v_t - c_{\mathcal{K}_t} > 0$, 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.

4 Managerial Choices under Task Complexity and Task Uncertainty

In this section I study the management of workers in firms facing high task complexity and high task uncertainty, respectively. In the management of a firm, economists are most often concerned with (1) investment decisions and decision rights, (2) incentives and the structure of contracts, and (3) communication. Each of these three aspects is affected by the task distribution the firm faces.

In each of the following subsections I impose additional modeling assumptions in order to capture the interaction between the task distribution and the managerial choices the firm makes. It may seem at times that the set-up is too stylized or that assumptions to not apply equally well to firms facing task uncertainty and task complexity. The main objective of this section is not to contribute to the literature on incentives, decision rights, or optimal language, but to show that firms facing high task uncertainty differ from those facing high task complexity in the managerial choices they make. Identical assumptions that appear more suited for one type of firm than for the other underline the difference between the firms.

I find that we should expect uncertainty-driven teams to communicate using broader words, to rely less on high-powered incentives, to have more decision rights, and to use fewer task-specific tools and training. In other words, the driver of the team formation matters for the management of the team.

4.1 Investments and Decision Rights

Consider a scenario where a firm can increase the value of a solution (or reduce the cost) by investing in tools or training, for example, a particular software package, machine, or exposure to particular methodologies. Some of these investments are general in nature as they are valuable for a wide range of tasks. Other investments are task specific in that they are only valuable for a particular task. Moreover, the decision to make such an investment can be made ex-ante, before facing any particular problem, or ex-post, after having gained some information about the task at hand. If the investment decision is made ex-post, then either the workers who address the problem must have decision rights to make the investment or they must communicate what they have learned to a decision maker. I assume that for the firm to make reasonable ex-post decisions it needs to set-up appropriate communication channels or establish hiring practices such that employed workers can be endowed with decision rights. In either case, the firms needs to pay a fixed cost $F \geq 0$ to implement ex-post decision making. If the investment decision is made ex-ante, no information beyond the task distribution is available or needs to be communicated.

I assume that it is cheaper for the firm to make such investments ex-ante. The cost savings might reflect making the investment at an opportune moment, the firm having more time to negotiate or to find alternative suppliers.

Formally, assume for now that the fixed set-up cost for ex-post decision making is $F = 0$ and that the firm can choose from a menu of investments $(C_i, r_i(\mathfrak{t}))$. Each such investment increases the revenue generated when problem $\mathfrak{t}$ is solved to $(1+r(i,\mathfrak{t}))\cdot v_{\mathfrak{t}}$. An investment i *generic*, if $r_i(\mathfrak{t})$ is independent of $\mathfrak{t}$, and *specific* to a problem $\hat{\mathfrak{t}}$ if $r_i(\mathfrak{t}) = 0$ everywhere except for $\mathfrak{t} = \hat{\mathfrak{t}}$. If the firm makes investment i after a problem is drawn, the marginal cost are C_i. If investment i is made ex-ante, the marginal cost are δC_i, with $\delta < 1$. The discount of $1-\delta$ may reflect a discount in bulk ordering or the value of being able to make the investment at an opportune moment. 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.[28] I say that an investment i is *valuable* for problem $\mathfrak{t}$ if $r_i(\mathfrak{t}) \cdot v_{\mathfrak{t}} \geq C_i$.

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. If the firm makes the investment ex-post, the benefit is

$$\sum_{\mathfrak{t} \text{ s.t. } i \text{ is valuable}} p_{\mathfrak{t}} \left[r_i(\mathfrak{t}) \cdot v_{\mathfrak{t}} - C_i \right].$$

If the firm makes the investment ex-ant, the benefit to the firm is

$$\left[\sum_{\mathfrak{t} \text{ s.t. } i \text{ is valuable}} p_{\mathfrak{t}} r_i(\mathfrak{t}) \cdot v_{\mathfrak{t}} c \right] - \delta \cdot C_i.$$

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-ante if

$$\left[\sum_{\mathfrak{t} \text{ s.t. } i \text{ is valuable}} p_{\mathfrak{t}} r_i(\mathfrak{t}) \cdot v_{\mathfrak{t}} \right] - \delta \cdot C_i \; > \sum_{\mathfrak{t} \text{ s.t. } i \text{ is valuable}} p_{\mathfrak{t}} \left[r_i(\mathfrak{t}) \cdot v_{\mathfrak{t}} - C_i \right].$$

So the investment i is made ex-ante rather than ex-post

$$\delta \cdot C_i \; < \; C_i \cdot \sum_{\mathfrak{t} \text{ s.t. } i \text{ is valuable}} p_{\mathfrak{t}}.$$

which is equivalent to

$$(1 - \delta) \; > \; 1 - \sum_{\mathfrak{t} \text{ s.t. } i \text{ is valuable}} p_{\mathfrak{t}}.$$

The left hand side in this last inequality represents the savings that arise from making the investment ex-ante. The right-hand side reflects savings from not making unnecessary investments. Naturally, a firm prefers an ex-ante over an ex-post timing of investment i if the corresponding savings are larger.

We summarize this finding in the following proposition:

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

$$\delta \; \leq \sum_{\mathfrak{t} \text{ s.t. } i \text{ is valuable}} p_{\mathfrak{t}}.$$

If a valuable investment i is specific to problem $\mathfrak{t}$, then the investment is made ex-ante if and only if the probability of problem $\mathfrak{t}$ satisfies $p_{\mathfrak{t}} \geq \delta$.

[28] For 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.

The result is somewhat obvious, but it highlights the importance of the task distribution for the timing of investments. In particular, the proposition implies that (1) 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.

Now return to the specific setting where a firm faces the task distribution (p_a, p_b, p_{ab}) as before, and assume that the firm has the choice of three task-specific investments $(C_a, r_a), (C_b, r_b), (C_{ab}, r_{ab})$, respectively. Each of these investment will increase the value of the solving a corresponding task $\mathfrak{t}$ to $r_\mathfrak{t} \cdot v_\mathfrak{t}$ at cost $C_\mathfrak{t}$. Assume that all three investments are valuable, i.e, $r_\mathfrak{t} \cdot v_\mathfrak{t} > C_\mathfrak{t}$.

Then the firm prefers an ex-ante investment over an ex-post investment if and only if $\delta < p_\mathfrak{t}$. Figure 5 a) shows the regions in the distribution space for which a firm will make ex-ante investments for the most frequently occurring task (shown for $\delta < 0.5$). Whether or not a firm makes further ex-ante investments depends on the organizational structure of the firm: For a flat organizational structure, the firm will make the ex-ante investment for another task $\mathfrak{t}'$ if that task satisfies $p_{\mathfrak{t}'} > \delta$ as well. Figure 5 a) shows such distributions in the intersection of two shaded regions. For a hierarchy, a second level worker (or team) faces a residual distribution, e.g., if an A-worker solved all a problems, the firm's next layer faces the distribution $\big(0, p_b/(1-p_a), p_{ab}/(1-p_a)\big)$. The firm will make an ex-ante investment for problems solved by the second layer if and only if it would have made one for a first-layer worker addressing the residual distribution. Recall that we can find residual distributions on the opposite edge of the solved problem's vertex. We can thus deduce the ex-ante investments made for higher layers of hierarchies. These investments are shown in figure 5 b).

As long as the fixed set-up cost for ex-post decision making is $F = 0$, all investments not made ex-ante are made ex-post. If $F > 0$, ex-post investments depend on the task distribution the firm faces.

Assume that paying the set-up cost is worthwhile for all task distributions if no task-specific investment is made ex-ante, i.e.,

$$p_a\,(r_a v_a - C_a) \;+\; p_b\,(r_b v_b - C_b) \;+\; p_{ab}\,(r_{ab} v_{ab} - C_{ab}) \quad > \quad F \quad > \quad 0.$$

If the firm does not make any ex-ante investments, i.e., $p_a < \delta$, $p_b < \delta$, and $p_{ab} < \delta$, then the firm will invest in infrastructure to make ex-post investments. For example, in figure 5 a) all firms facing a distribution in the white triangle in the middle of the simplex will invest in the infrastructure and make all task-specific investments ex-post.

Firms that optimally make some ex-ante investments may or may not find it worthwhile to invest in an ex-post decision making infrastructure. For example, if the firm optimally invests in (C_a, r_a) ex-ante, the firm will invest in the set-up cost for ex-post decision making if and only if

$$p_b\,(r_b v_b - C_b) \;+\; p_{ab}\,(r_{ab} v_{ab} - C_{ab}) \quad > \quad F.$$

In particular, the more frequently one or more problems occur in the firm's task distribution,

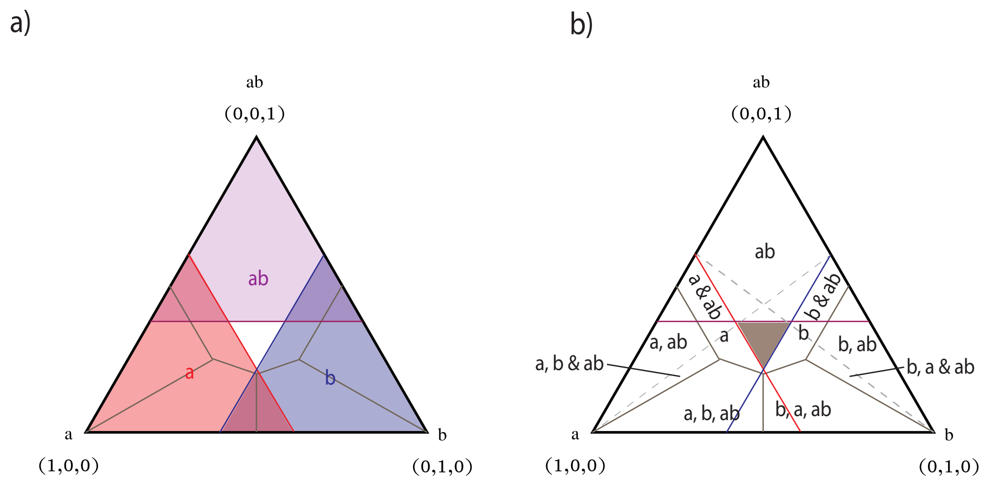


Figure 5: Assume all task-specific investments are valuable. a) The shaded regions in the simplex indicate the regions where an ex-ante investment in the respective task-specific tool or training is optimal. For the white area at the center of the simplex no task occurs frequently enough to warrant an ex-ante task-specific investment. In the specific case shown, the discount cost of the ex-ante investment are δC_t with $\delta < 0.5$. As δ increases, the regions where ex-ante investments are optimal decline in size. In particular, for $\delta > 0.5$ the colored regions would not overlap. b) Taking the optimal organizational form into account, we show all ex-ante task-specific investments. Small letters indicate the corresponding investments. Commas separate investments made for different layers of a hierarchy, ampersands indicated more than one task-specific investment is made. For hierarchical structures the investments corresponding to a layer's task(s) is made ex-ante if the task occurs frequently enough in the respective residual distribution. Firms facing distributions where no ex-ante investment is efficient will make all investments ex-post. The corresponding distributions are shaded grey. For sufficiently high set-up cost F only those firms will make ex-post investments.

the less likely the firm is to invest in ex-post decision making.

Depending on the magnitude of F, the region in the simplex where firms make ex-post investments may extend to all investments not made ex-ante or only to those distributions where no ex-ante investments are made. Figure 5 b) shows which task-specific investments are made ex-ante, ex-post, or not at all taking the firm's optimal organizational structure as given.

To summarize, firms that face high task uncertainty distributions and therefore choose a flat organizational structure are more likely to engage in ex-post investments, share decision rights and/or have a more extensive communication structure than firms that choose a flat organizational structure because they face a task distribution with frequent complex problems.

4.2 Incentives and the Structure of Contracts

Assume that effort is costly for workers to provide and that the firm needs to either incentivize workers through performance-based contracts or invest in an infrastructure that ensures high input levels of effort. Such an infrastructure may involve explicit monitoring, extended hiring processes to ensure only intrinsically motivated individuals are hired, or investments in reputation to sustain relational contracts. In all of these cases, the cost associated with the practice is independent of the problem distribution the firm faces. Let F be the fixed cost investment the firm must make to ensure high (efficient) levels of effort are provided by its workers.

In contrast, implementing performance-based contracts does not require an infrastructure beyond the ability to observe the outcome each period and an enforcement mechanism. However, performance based contracts may be costly if hiring risk-averse workers requires the payment of a risk premium or if the performance pay has an overall damped impact effort.

I argue that the cost associated with performance-based contracts are higher for firms facing distributions with high task uncertainty, making them thus more likely to be the ones to adopt input-bases evaluations.

For specificity assume that firms face task distributions (p_a, p_b, p_{ab}) as before and performance-based contracts offered to worker i takes the form $\left(w_a^i, w_b^i, w_{ab}^i, w_0\right)$ where $w_{\mathfrak{t}}^i$ denotes the wage paid to worker i if after the problem has been solved the problem turns out to have been task $\mathfrak{t}$. The wage w_0 is paid if the problem is not solved. I normalize the worker's utility such that $U(w_0) = 0$. The probability $P_{\mathfrak{t}}$ that a worker with the appropriate expertise solves a problem $\mathfrak{t}$ depends on the effort the worker exerts. Let f_A be the effort provided by an A-worker and f_B be the effort provided by a B-worker, and let $C_i(f_i) = 1/2 f_i^2$ be the cost of effort. Given a particular contract, an A-worker working alone chooses which effort to exert to maximize expected utility

$$\max_{f_A} \ p_a P_A(f_A) U\left(w_a^A\right) \ - \ \frac{1}{2} f_A^2$$

and an A-worker working together with a B-worker chooses which effort to exert to maximize expected utility

$$\max_{f_A} \ p_a P_A(f_A) U\left(w_a^A\right) \ + \ p_b P_B(f_B) U\left(w_b^A\right) \ + \ (1-p_a-p_b) P_{ab}(f_A, f_B) U\left(w_{ab}^A\right) \ - \ \frac{1}{2} f_A^2.$$

I limit the following discussion to the organizational structure where an A-worker and a B-worker work together. The corresponding optimization problem exhibits both the risk-exposure and the damping of performance pay on the worker's effort. Both channels make performance-based pay less effective for firms facing high uncertainty distributions:

4.2.1 Risk Premium Depends on Task Distribution

Given a particular contract (w_a^A, w_b^A, w_{ab}^A) and equilibrium effort levels f_A^* and f_B^*, an A-worker's expected pay-off is

$$E\left[U\left(w^A\right)\right] = p_a P_A(f_A^*) U\left(w_a^A\right) + p_b P_B(f_B^*) U\left(w_b^A\right) + (1-p_a-p_b) P_{ab}(f_A^*, f_B^*) U\left(w_{ab}^A\right) - \frac{1}{2} f_A^{*2}.$$

Assume that the cost of effort is low enough such that each worker exerts sufficient effort to solve every problem he has the expertise to solve, i.e., $P_A(f_A^*) = P_B(f_B^*) = P_{ab}(f_A^*, f_B^*) = 1$. If the worker is risk-averse, the wages will reflect a risk-premium reflecting the variance in pay the worker is subject to. Given a particular distribution the firm faces (p_a, p_b, p_{ab}) the worker is exposed to the variance in pay equal to

$$\begin{aligned} Var(w^A) &= w_a^{A2} p_a(1-p_a) \ + \ w_b^{A2} p_b(1-p_b) \ + \ w_{ab}^{A2}(1-p_a-p_b)(p_a+p_b) \\ &\quad -2 w_a^A w_b^A p_a p_b \ - \ 2 w_a^A w_{ab}^A p_a(1-p_a-p_b) \ - \ 2 w_b^A w_{ab}^A p_b(1-p_a-p_b) \end{aligned}$$

Paying an A-worker when the solved problem turns out to have been a b problem reduces the variance in pay that the A worker is exposed to but it has no impact on the worker's effort provision. In particular, observe

$$\begin{aligned} \frac{dVar}{dw_b^A} &= 2 w_b^A p_b(1-p_b) \ - \ 2 w_a^A p_a p_b \ - \ 2 w_{ab} p_b(1-p_a-p_b) \\ &= 2 p_b \left(w_b^A \ - \ p_b w_b^A \ - \ p_a w_a^A \ - \ p_{ab} w_{ab}^A\right) \\ &= 2 p_b \left(w_b^A \ - \ \bar{w}^A\right) \end{aligned}$$

where $\bar{w}^A$ denotes the average pay the A-worker expects as well as the average pay the firm expects to spend. We see that the variance in pay will decrease as w_b^A increases as long as $w_b^A < \bar{w}^A$. Moreover, the difference between $w_b^A - \bar{w}^A$ is more salient the more frequently problem b occurs. In other words, the larger p_b the more expensive it is to incentivize a risk-averse A-worker. Figure 6 shows the distributions where it is most expensive to incentivize risk-averse A-worker and B-worker, respectively. Note that for larger values of p_b the A-worker does not need to be incentivized (in the first layer of the organization) due to the hierarchical structure optimally chosen for larger values of p_b.

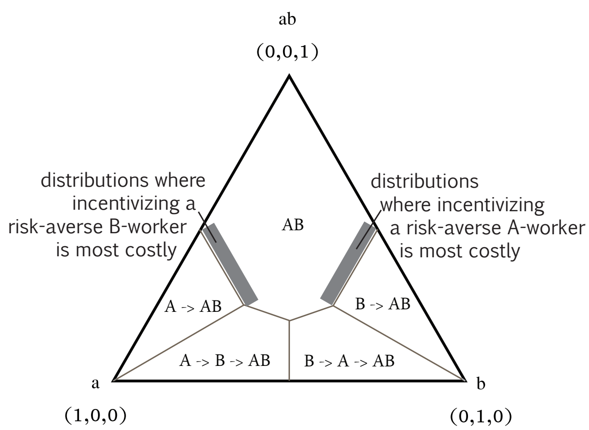


Figure 6: Worker i is offered a contract of the form $\left(w_a^i, w_b^i, w_{ab}^i\right)$. Paying an A-worker when the solved problem turns out to have been a b problem reduces the variance in pay that the A worker is exposed to but it has no impact on the worker's effort provision. The gray areas mark the distributions where it is most expensive to incentivize risk-averse A-worker and B-worker, respectively. Note that for larger values of p_b the A-worker does not need to be incentivized (in the first layer of the organization) due to the hierarchical structure optimally chosen for larger values of p_b.

As a consequence, firms facing distributions in the center of the simplex face relative high cost of both incentivizing A- and B-workers.

4.2.2 Impact on Effort Depends on Task Distribution

Now, in turn let us focus on the optimal effort provided. Assume utility is linear, i.e.

$$U_i\left(w_{\mathfrak{t}}^i\right) = w_{\mathfrak{t}}^i$$

and assume that the probability of solving a problem given f_A is exponential, i.e.,

$$P_A(f_A) = 1 - e^{-f_A} \quad P_B(f_B) = 1 - e^{-f_B} \quad P_{AB}(f_A, f_B) = \sqrt{\left(1 - e^{-f_A}\right)\left(1 - e^{-f_A}\right)}.$$

The probability P_{AB} is normalized to the same magnitude as P_A.

Then the A-worker's optimization problem simplifies to

$$\max_{f_A} \; p_a P_A(f_A) w_a^A \; + \; p_b P_B(f_B) w_b^A \; + \; (1 - p_a - p_b) P_{ab}(f_A, f_B) w_{ab}^A \; - \; \frac{1}{2} f_A^2.$$

The first-order-condition of this optimization problem is

$$f_A = p_a P'_A w_a^A \ + \ (1 - p_a - p_b) P'_{ab} w_{ab}^A.$$

We can rewrite the first-order condition as

$$\mathcal{F} \quad = \quad p_a w_a^A \ + \ (1 - p_a - p_b) \cdot \frac{\sqrt{(1 - e^{-f_B})}}{\sqrt{(1 - e^{-f_A})}} w_{ab}^A \quad - \quad f_A e^{f_A}.$$

Observe that keeping effort levels of the B level worker fixed, the best effort response of the A-worker to a given level of effort f_B of the B-worker decreases in p_b:

$$\begin{aligned} \frac{\partial f_A}{\partial p_b} &= -\frac{\frac{\partial \mathcal{F}}{\partial p_b}}{\frac{\partial \mathcal{F}}{\partial f_A}} \\ &= \frac{\frac{\sqrt{(1-e^{-f_B})}}{\sqrt{(1-e^{-f_A})}} w_{ab}^A}{-0.5(1 - p_a - p_b) \cdot \frac{\sqrt{(1-e^{-f_B})}}{(1-e^{-f_A})^{3/2}} \cdot e^{f_A} w_{ab}^A - e^{f_A} - f_A e^{f_A}} \\ &< 0 \end{aligned}$$

The B-worker faces an analogous optimization problem. The two workers' respective first order conditions implicitly define best response functions

$$\begin{aligned} f_A e^{f_A} &= p_a w_a^A \ + \ (1 - p_a - p_b) \cdot \frac{\sqrt{(1 - e^{-f_B})}}{\sqrt{(1 - e^{-f_A})}} w_{ab}^A \\ f_B e^{f_B} &= p_b w_b^B \ + \ (1 - p_a - p_b) \cdot \frac{\sqrt{(1 - e^{-f_A})}}{\sqrt{(1 - e^{-f_B})}} w_{ab}^A \end{aligned}$$

based on which we can numerically determine the equilibrium effort levels provided.

Figure 7 b) shows the equilibrium effort provided by an A-level worker for different distributions, keeping the contracts (w_a^A, w_b^A, w_{ab}^A) and (w_a^B, w_b^B, w_{ab}^B) fixed. We see that indeed the equilibrium effort of the A-worker decreases in probability p_b. In contrast, panel a) shows the equilibrium effort provided by an A-level worker who is working alone and whose effort can only affect the solution of a problems. Observe that for a A-worker who works alone the equilibrium effort increases in p_a. Given the firm's task distribution and the equilibrium effort provided by A- and B-workers but ignoring the firm's organizational structure, panel c) shows iso-profit lines across the distribution space, with profit increasing toward the ab vertex. In panel d) the boundaries of optimal organizational forms are overlaid. we see that in particular firms that optimally organize in a flat organizational structures due to high task uncertainty reap less benefits from performance-based contracts than firms that optimally organize in a flat organizational structures due to high task complexity.

4.3 Communication

Finally, consider communication and the optimal language for firms facing high task uncertainty and high task complexity.

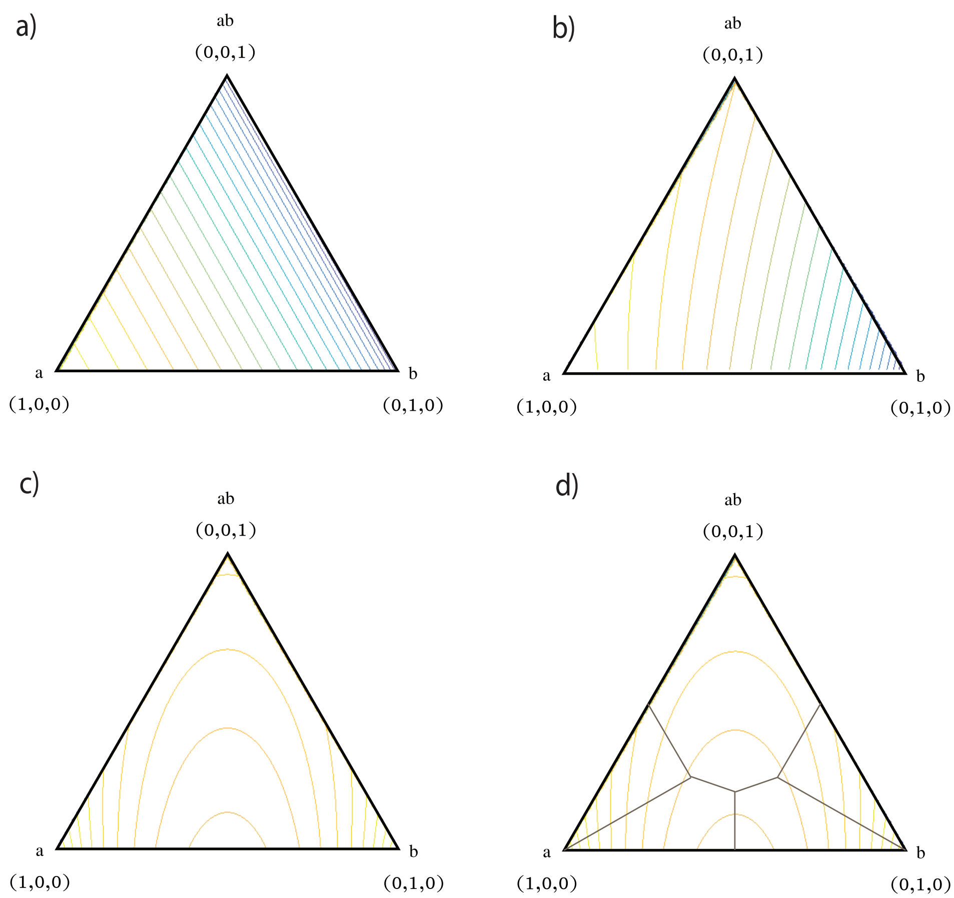


Figure 7: Worker i is offered a contract of the form $\left(w_a^i, w_b^i, w_{ab}^i\right)$. Panel a) shows iso-effort lines of an A-worker if that A-worker is working alone. Effort levels increase as p_a increases. Panel b) shows the iso-effort lines for an A-worker if the A-worker works jointly with a B-worker. Effort levels increase as p_b decreases. Panel c) shows iso-profit lines across the distribution space, with profit levels increasing toward the ab vertex. In panel d) the boundaries of optimal organizational forms are overlaid.

Cremer et al. [2007] formally model the optimal vocabulary in a team depending on the frequency with which the team encounters particular events. They consider a set of problems that members of a team need to communicate about. Since the team members are subject to bounded rationality, they can only learn a limited number of words. Therefore, more than one problem might optimally be assigned to the same word. Cremer et al. [2007] argue that the marginal benefit of an additional word is high at first and then decreasing for firms that face a narrow task distribution, i.e., one with a few problems occurring very frequently. In contrast, the marginal benefit of an additional word is constant or only slowly decreasing for a broad task distribution, i.e. one with many problems occurring infrequently. Consequently, the implication the task distribution has on the size of the optimal language is ambiguous.

However, given a fixed number of words to work with, the optimal language for a firm with a narrow distribution will always contain more words referring to a few number of events, while the optimal language for a firm with a wide distribution will mostly consist of words refering to a large number of events.

Applied to the setting of this paper, the optimal language for flat organizational structures due to high task complexity consists of words with narrow and specific meanings. In contrast, the optimal language for firms facing high task uncertainty will be more general and consist of words with broad meaning.

As an unexplored consequence, firms with complexity-driven teams may be more easily able "codify" the necessary language, since it is more specific and narrow. For example, the firm may be able to write down a relevant "vocabulary" list for an A worker. The firm may hence be easily able to substitute one "A"-worker with another "A"-worker. However, within uncertainty-driven teams, the language is likely to be more fluid and history-dependent, e.g., "This part of the problem looks similar to the problem we worked on two periods ago." In that case, changing team-members would result in a (partial) loss of the language developed.

4.4 Managing Real-World Teams facing High Task Uncertainty and High Task Complexity

To summarize the previous subsections, compared to a complexity driven team an uncertainty driven team is expected to be more communication intensive, communicate using broader words, rely on less high-powered incentives, have more decision rights, less task-specific tools and training, be less likely to be monitored, and use reputation or other alternative mechanisms to overcome free-riding problems.

As a case study I here contrast a surgery team and a team working at the industrial design company IDEO. Figure 8 shows a surgery team and a team at IDEO at work. [29] Whereas

[29] Sources: a) The Pediatric Cardiac Surgery Inquest Report, Chpt. 3, `http://www.pediatriccardiacinquest.mb.ca/ch03/diagram3_2.html`, Palestine Children Relief Fund, `http://palestinenote.com/cfs-filesystemfile.ashx/__key/CommunityServer.Blogs.Components.WeblogFiles/`

the former consist of prescribed number of nurses, anesthesiologist, surgeon, and assistants with specific responsibilities, the latter consist of ”four to eight people with different skill sets and experiences” and no particular roles.

The comparison between the complexity-driven surgery team and the uncertainty-driven IDEO team shows many differences between these teams. The surgery team uses a precise, narrow language to conduct an operation, while the IDEO team does not employ any specialized language. Roles and responsibilities are clearly prescribed for the members of the surgery team but vary or are undetermined for the IDEO team. Almost all of the equipment and tools the surgery team uses during the procedure were purchased by the hospital before the patient was even admitted. In contrast, IDEO relies on generic tools such as “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, [and] duct tape” for their project work and purchase additional items as needed after a project has started. These differences align with the different managerial choice we expect based on the discussion of the previous subsections - even though both surgery and brainstorming at IDEO is done in “teams.”

5 The Task Distribution Matters: Profits and Longevity

In section 3 I showed that both complex tasks and uncertain tasks result in the same organizational response, namely a flat organizational structure. In section 4 I argued that none the less firms facing uncertain and complex tasks will make different managerial choices. In this section I argue that firms facing distributions characterized by high task complexity or high task uncertainty-driven differ in other ways, too. First, I establish that firms facing high task uncertainty are less likely to achieve first-best profits. Second, the difference between realized and first-best profits makes these firms more likely to invest in diagnosis technology. As a consequence, firms facing high task-uncertainty are less likely to maintain their organizational and managerial choices over time compared to firms facing high task complexity.

5.1 The Cost of Uncertainty

As discussed in section 3.2 the first-best benchmark is the profit firms would realize if they were able to solve every problem with the minimal knowledge set. In the world of $N = 2$, this first-best profit is given by

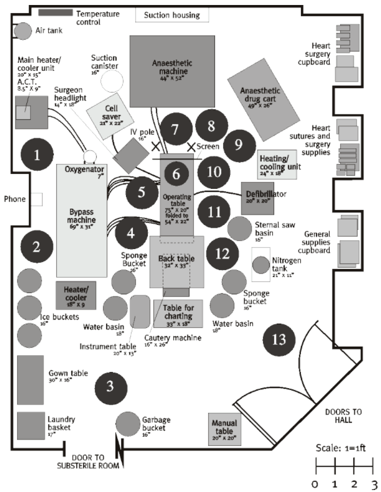
$$\Pi_{FB}(p_a, p_b, p_{ab}) = p_a \cdot (v_a - c_A) \; + \; p_b \cdot (v_b - c_B) \; + \; p_{ab} \cdot (v_{ab} - c_{AB}).$$

The cost of uncertainty is the cost incurred by the firm through hiring excessive expertise, because it does not know which expertise is needed for a given task. Given the optimal

news/0576.Steve-Sosebee-_2D00_-Italian-doctors-1.jpg.
b) http://blog.gentry.io/ideo-seen-through-a-leica-m3.

Team Work

a)



Operating Team at Work

Operating Room Layout Example

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

b)

IDEO Team at Work.

Figure 8: a) An operating team at work and the layout for 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 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."

organizational form for every task distribution, we can compute the actual profit firms realize given the respective optimal organizational choice. The cost of uncertainty is the difference between the first-best profit and the actual realized profit.

The result is shown in figure 9. The first panel shows the the first-best benchmark for a scenario where $v_{ab} - c_{AB} > v_a - c_A = v_b - c_B$. Naturally, the potential profit is proportional to p_{ab}. The middle panel shows the second-best profit firms can realize, and the last panel shows the difference.

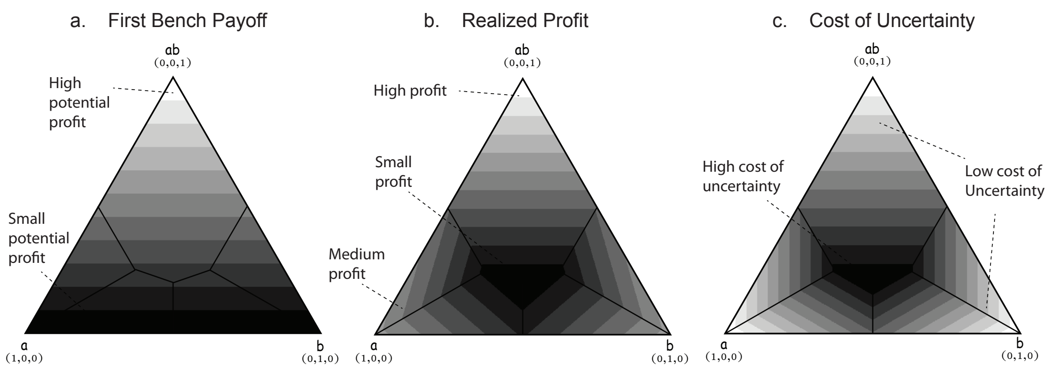


Figure 9: Equi-pay-off lines in the distribution space with two units of knowledge. **a.** First-best pay-off. Every task is solved with the minimal necessary knowledge set. Parameters shown satisfy $v_{ab} - c_{AB} > v_a - c_A = v_b - c_B$. **b.** Actual profit firms realize if every distribution is addressed with the respective optimal organizational form. Distributions at the vertices are the most profitable ones, distributions at the center of the simplex are the least profitable to address. **c.** Cost of uncertainty, i.e., the difference between the first-best and the second-best pay-off. This reflects cost the firm incurs when attempting to solve a problem with insufficient or too extensive knowledge, i.e., cost due to not knowing which precise knowledge is necessary to solve a given task.

Figure 9 also answers the question of which task distribution a firm would want to choose, if it could. For example, firms entering a new market first might have first choice. Figure 9 shows that firms prefer distributions close to the vertices and close to the edges. Indeed, if firms can choose how to position themselves as they enter the marketplace, distributions with high task uncertainty are chosen last.

Note also that figure 4 implies that distributions with high task uncertainty may become profitable later as cost of accessing knowledge and expertise decreases.

Both of these two observations suggest that a recent rise in the prevalence of teams may be due to a rise in the prevalence of uncertainty-driven teams rather than complexity-driven teams.

5.2 Firms Facing High Task Uncertainty are Not Forever - The Organizational Impact of Diagnostics

Figure 9 suggests that firms might be willing to pay for mitigation of task uncertainty, especially firms facing distributions with high task uncertainty. To mitigate task uncertainty means to gain information about the knowledge that is necessary to solve a particular task before attempting to solve the task, i.e., to diagnose a problem.

In the context of this model, diagnosing a problem means to split a distribution into two (or more) conditional distributions. For example, a diagnosis may be able to detect whether a or b are part of the problem. A problem of type ab might be sorted into either group. Then problems drawn, for example, from a distribution (1/4, 1/2, 1/4) might be sorted into the two (conditional) distributions (3/4, 0, 1/4) and (0, 3/4, 1/4).[30] A perfect diagnosis would be a device that can perfectly identify the task at hand. Figure 10 illustrates the split of the firm's task distribution into two (conditional) distributions.

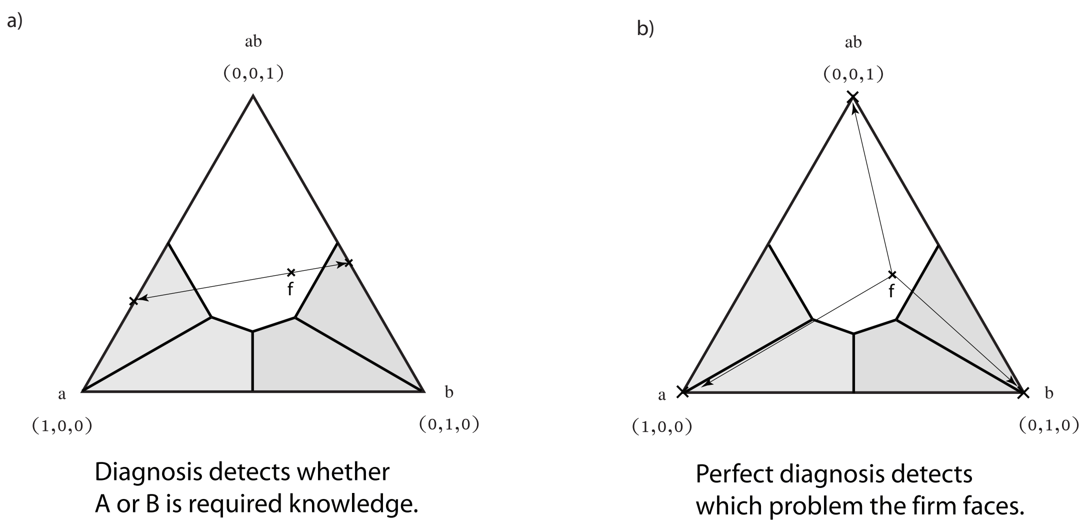


Figure 10: The two panels show two different forms of diagnosis technology. Panel a. shows a case where firms can detect whether knowledge A or B is required to solve a given problem. For problems that require both types of knowledge the technology randomly detects either knowledge. The resulting two distributions are on the edges of the distribution space. Panel b. illustrates the effect of perfect diagnosis technology. When use of the technology perfectly identifies all three types of problems, each problem can then be solved by the optimal first-best knowledge set. Every task distribution is split into distributions at the vertices.

If the conditional distributions after a diagnosis can be addressed more profitably than the

[30]In other words, the diagnosis splits the distribution into two or more distributions on the edge of the simplex such that the original distribution is a convex combination of the resulting distributions.

original distribution, then firms are be better off by investing into such a diagnosis device. The larger the cost of uncertainty, and the larger the reduction in uncertainty due to the diagnosis technology, the more firms are willing to pay for the diagnosis device.

The cost of uncertainty shown in figure 9 panel c. represents the firm's willingness to pay for a perfect diagnosis technology: Firms facing distributions with high task-uncertainty are willing to pay the most. Indeed, as figure 11 shows for perfect diagnosis technology made available at non-zero cost, these firms are the first to adopt a triage technology.

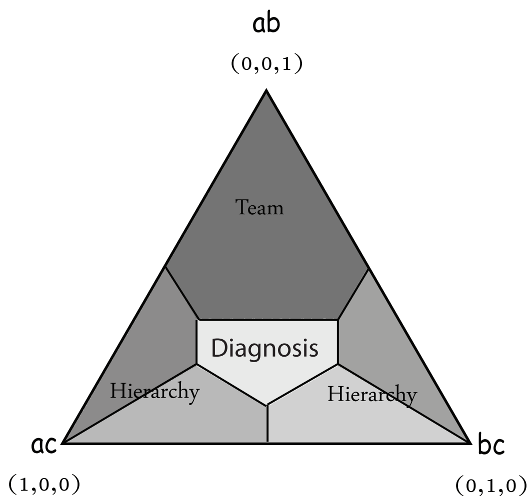


Figure 11: Firms in this economy have the choice to adopt a perfect diagnosis technology at cost C. When the technology perfectly identifies all three types of problems each problem can then be solved by the optimal first-best knowledge set. For $C = 0$, of course, all firms would adopt the technology. For $C > 0$, not all firms will do so. For large C, firms facing distributions with high task-uncertainty are the first ones to adopt the technology.

Many start-ups can be thought of as having a high-uncertainty as to the distribution of problems they are addressing. In other words, in expectation many start-ups are starting out near the center of the distribution space. Figure 11 shows that such firms have an incentive to learn about the distribution they are facing, and then to learn to diagnose the problems they face. In other words, few firms that start out addressing distributions with high task uncertainty will continue to do so in the long-run.[31]

Throughout this paper, I have assumed that failing to solve a problem is not informative. In reality, of course, failure often is informative: An agent with knowledge A might not be able to solve a task at hand, but the agent might be able to identify whether the task requires his input at all, i.e., whether the task is b or ab. Figure 11 and Panel c. in figure

[31] The exception to this observation are those firms that specialize in problems with high task uncertainty. Examples for such firms include product development and industrial design firms, as well as some customization and some consultancy firms.

9 highlight the importance of agents being able to do so, and how a firm's organizational choice depends on whether or not agents can do so.

Last but not least, figure 11 shows the differential impact diagnosis technology has on complexity and uncertainty driven teams: While complexity driven teams are - naturally - not affected by the diagnosis technology, uncertainty driven teams are made obsolete by it.

6 Conclusion

In this paper I use a novel modeling approach of firms as problem solvers to shed light on the challenges firms may face when striving to be repeatedly innovative. Within the model, innovation problems correspond to problem distributions with high task uncertainty that are solved by combining existing knowledge. Knowledge is represented in discrete units rather than as a continuous parameter. As a consequence, considering the entire distribution space of all problems becomes tractable. We can simultaneously consider distributions characterized by some relatively simple problems occurring relatively frequently, distributions characterized by high task uncertainty, and those characterized by high task complexity.

I show that firms facing distributions with high task complexity and those facing distributions with high task uncertainty optimally choose the same flat non-hierarchical organizational structure. But I argue that the two firms choose different contract structures, adopt different decision rights, and implement different languages.

I believe that the distinction between complexity- and uncertainty-driven teams is meaningful in the real world and is, for example, reflected in the apparently contradictory advise given on "how to build a better team." Recommendations range from setting clear leadership, defining clear role expectations, and hiring "for competency first" to distributing leadership, flexible role assignments, and hiring "for character first." While the former set of suggestions seem more appropriate for complexity-driven teams, the latter set appears more suited for uncertainty-driven teams.

The results in this paper are consistent with teams being the optimal organizational form to perform a hip replacement surgery (complex and predictable) as well as develop a new product with certain specifications (unknown solution). But the paper also suggests that the management of teams should be matched to the underlying task distribution.

References

Peter Augsdorfer. *Forbidden Fruit: An Analysis of Bootlegging, Uncertainty and Learning in Corporate R&D*. Averbury, 1996.

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.

Brent Boning, Casey Ichniowski, and Kathryn Shaw. Opportunity counts: Teams and effectiveness of production incentives. *NBER Working Paper*, 2003.

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.

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.

Jacques Cremer, Luis Garicano, and Andrea Prat. Codes in organizations. *QJE*, 2007.

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.

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.

Bengt Holmstrom. Agency cost and innovation. *Journal of Economic Behavior and Organization*, pages 305–327, 1989.

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.

Eric Taub. *Taurus: The Making of The Car that Saved Ford*. A Dutton Book, 1991.

Gianmario Verona and Davide Ravasi. Unbundling dynamic capabilities: an exploratory study of continuous product innovation. *Industrial and Corporate Change*, 12:577–606, 2000.

A Extensions: Uncertainty and General Equilibrium Outcomes

A.1 Uncertainty Does Drive Team Formation

In section 3 I showed that teams are the optimal organizational form to address a distribution with high task uncertainty, even if complex problems are rare. Yet, it might be the case that the rare complex problems are driving the team formation rather than the task uncertainty. To support the claim that indeed task uncertainty is the driver, I here consider the simplex spanned by three problems of constant complexity: ab, bc, ac. There are now four relevant knowledge sets: AB, BC, AC, and ABC. Let c_{AB}, c_{BC}, c_{AC}, and c_{ABC} denote the corresponding cost of hiring agents with the respective knowledge. I want to ask for which distributions the "all-knowing" team which provides the complete knowledge ABC is optimal. Consistent with the previous notation, denote the revenue generated by v_{ab}, v_{bc}, and v_{ab}, respectively.

As before, distributions near a vertex are optimally addressed with a knowledge sequence that has the knowledge set at its first stage that can solve the most frequent problem. This determines the optimal organizational form along the boundary of the distribution space as shown in figure 12 a). As before, we can extrapolate the optimal knowledge sequences from the boundaries of the distribution space to the interior: Distributions in the proximity of a vertex are optimally addressed with a knowledge sequence that starts with the knowledge sequence that can solve the most frequent problem. The remainder distribution of all unsolved problems corresponds to a distribution on the edge opposite the original vertex. If the cost of the respective knowledge sets is symmetric, then we would expect the symmetric partitioning shown in figure 12 b) to represent the regions where each of the six knowledge sequences is optimal.

Instead, at the center of the distribution space a team providing all three units of knowledge is optimal, as shown in figure 13.

As before, distributions with high task uncertainty are optimally addressed by a team providing knowledge ABC. In this example, task complexity can not be a driver of team formation, because all tasks have the same complexity.

One might object that team formation at the center of the distribution space is not due to task uncertainty but due to "'synergy."' But if synergy were the sole driver then we would expect teams to be singular optimal in the middle range of each of the three edges, too. For example, a firm addressing the distribution ($p_{ac} = 2/3$, $p_{bc} = 1/3$, $p_{ab} = 0$), could capture 'synergy' by using a team providing all three units of knowledge rather than paying for four units of knowledge one third of the time. But it is not optimal for the firm to do so.[32] Thus, 'synergy' alone cannot explain the optimality of teams at the center of the

[32] If the cost of knowledge is linear in the number of units of knowledge, then the expected cost of ABC is $3c$. But the expected cost of $AC \rightarrow BC$ is only $2c + (1 - p_{ac}) \cdot 2c = 22/3c < 3c$.

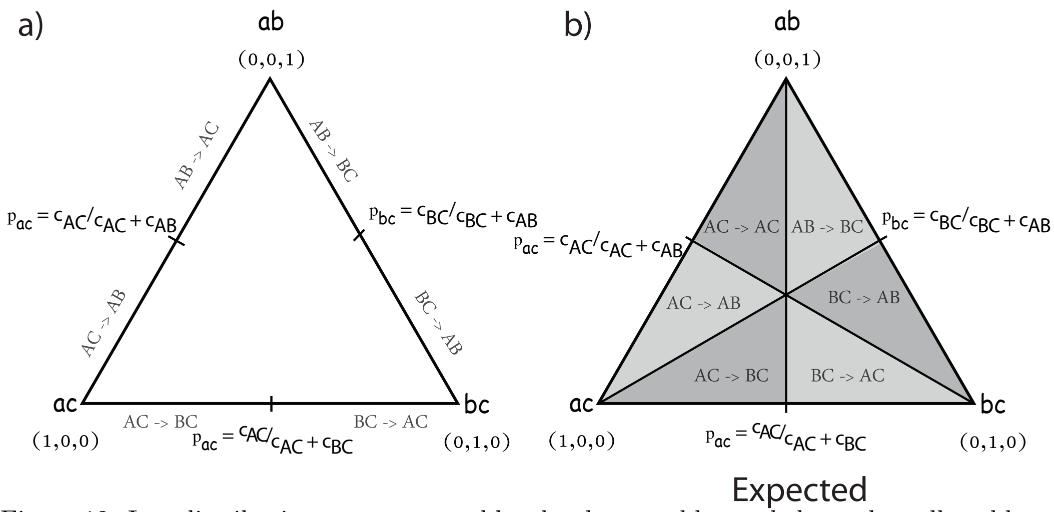


Figure 12: In a distribution space spanned by the three problems ab, bc, and ac all problems have the same complexity. a) The optimal organizational sequence along the boundaries contains a knowledge set that can solve the most frequent problem first and the less frequent problem second. b) If the cost of the respective knowledge sets is symmetric, projecting remainder distributions onto the opposite edge of the simplex implies we should expect these regions to describe the distributions for which the respective knowledge sequences are optimal.

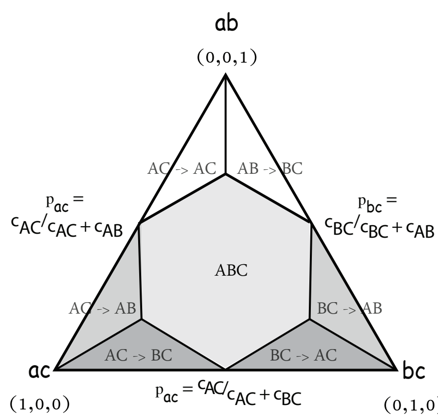


Figure 13: In a distribution space spanned by the three problems ab, bc, and ac all problems have the same complexity. But distributions with high task uncertainty at the center of the distribution space are optimally addressed by a team providing knowledge ABC. The figure drawn here assumes the cost of knowledge is linear in the number of units of knowledge.

distribution space.

A.2 Concavity, Convexity, and Team Formation

Given the discussion in the main sections of this paper, one might suspect that returns to knowledge are driving the optimal organizational choice. However, this is only approximately true and only for constant marginal cost of knowledge.

First, this is not true for economic returns to knowledge

$$V(\mathcal{K}) = \sum_{\mathfrak{t} \text{ solvable by } \mathcal{K}} p_{\mathfrak{t}} \cdot v_{\mathfrak{t}} \; - \; c_{\mathcal{K}}.$$

This function $V(\mathcal{K})$ can be convex or concave depending on how valuable particular solutions are. But as long as it is optimal to solve all problems, changing the value parameters $v_{\mathfrak{t}}$ does not affect the optimal organizational form. For a specific example, consider the distribution $p_a = 3/4$, $p_{ab} = 1/4$, and $p_b = 0$. Assume that $c_A = 1$ and $c_{AB} = 2$. Table 2 shows how varying v_{ab} changes the shape of $V(\mathcal{K})$, even though figure 2 implies that the organizational form $A \to AB$ is optimal in all three cases.

$v_a = 2$, $v_{ab} = 3$

$\mathcal{K}$	$\lvert\mathcal{K}\rvert$	$V(\mathcal{K})$
A	1	.50
AB	2	.25

decreasing in $\lvert\mathcal{K}\rvert$

$v_a = 2$, $v_{ab} = 5$

$\mathcal{K}$	$\lvert\mathcal{K}\rvert$	$V(\mathcal{K})$
A	1	.50
AB	2	.75

concave in $\lvert\mathcal{K}\rvert$

$v_a = 2$, $v_{ab} = 10$

$\mathcal{K}$	$\lvert\mathcal{K}\rvert$	$V(\mathcal{K})$
A	1	.50
AB	2	2

convex in $\lvert\mathcal{K}\rvert$

Table 2: The shape of economic returns to knowledge is not a predictor of the optimal organizational choice. Example: Distribution $p_a = 3/4$, $p_{ab} = 1/4$, and $p_b = 0$; cost $c_A = 1$, $c_{AB} = 2$. Varying the value of the complex problem v_{ab} changes the shape of returns to knowledge. But since both problems are always worth solving, figure 2 implies that the organizational form $A \to AB$ is optimal in all three cases.

Alternatively, one might use "returns to knowledge" in terms of the probability that a random problem is solved with an (optimal) increasing knowledge set,

$$\text{RtK}(k) = \max_{|\mathcal{K}|=k} \text{Prob}(\mathcal{K} \text{ solves}) = \max_{|\mathcal{K}|=k} \sum_{\mathfrak{t} \text{ solvable by } \mathcal{K}} p_{\mathfrak{t}},$$

This expression ignores the revenue that solving problems generate. Instead it reflects upon the cost saved in a second stage: The larger the probability that the first-stage knowledge set solves the task at hand, the less often the knowledge set at the second stage needs to be engaged. Moreover, this expression captures the combinatorial growth, i.e., how many more problems can be solved by increasing the knowledge set.

However, the shape of Rtk is also not predictive of the optimal organizational form. In the economy with $N = 2$, Rtk is concave if and only if the (optimal) one-unit knowledge set can solve at least half of all problems. The right panel in figure 14 shows the regions where Rtk is concave and convex, respectively. Comparing this with figure 2, which is replicated in the left panel, shows that convexity and concavity do not predict the optimal organizational form.

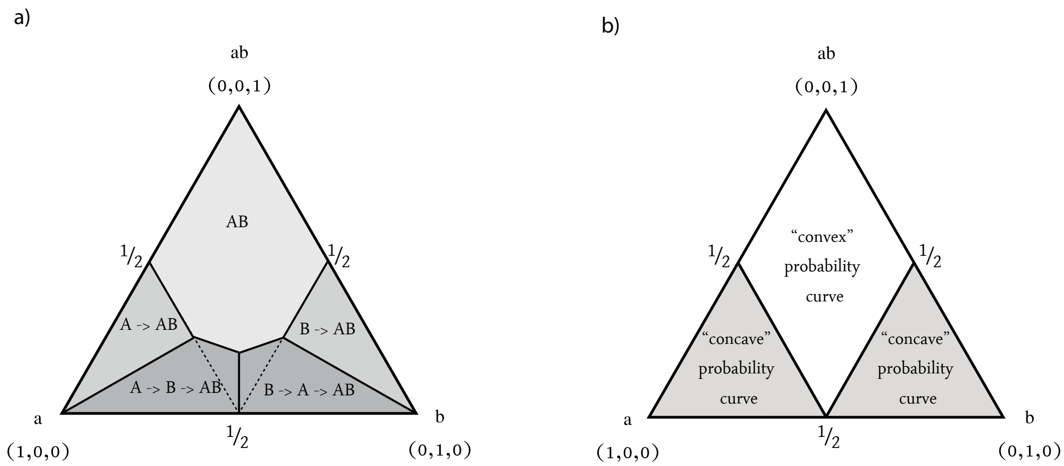


Figure 14: The left figure shows the prevalence of optimal organizational forms assuming that all problems are solved for all distributions and the marginal rate of knowledge is constant, i.e., $c_{AB} = 2c_A = 2c_B \leq v_A, v_B, v_{AB}$. The right figure shows the regions in which the probability curve is "convex" and "concave", respectively. In a world with 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.

Nonetheless, figure 14 suggests that convexity is necessary though not sufficient for team formation to be optimal. This is true in general as long as units of knowledge can be accessed at constant marginal cost:

Proposition 6 Assume that the marginal cost of knowledge is constant. Let $\{\mathcal{K}_k\}_{k\geq 1}$ be a collection of knowledge sets, such that $\mathcal{K}_k$ maximizes the revenue among all knowledge sets of cardinality k. If the probability of solving a problem, $Prob(\mathfrak{t} \text{ solvable by } \mathcal{K}_k)$ is concave in k, then a team with complete knowledge is not the optimal organizational choice. In particular, if the firm solves all problems, a multistage contingency sequence is optimal.

The reason that convexity is not sufficient is that the conditional distribution of unsolved problems is not necessarily convex, even if the original distribution was convex.

For example, consider the distribution

$$p_a = .45 \quad p_b = .35 \quad p_{ab} = .2.$$

Returns to knowledge RtK are convex for this distribution - Rtk (1) = .45 and Rtk(2) = 1. However, after problem a is solved, the remainder distribution

$$\tilde{p}_a = 0 \quad \tilde{p}_b = .64 \quad \tilde{p}_{ab} = .36$$

is concave. The gains from organizing in two stages after problem a has been solved are sufficient to make a multi-stage organizational form the overall optimal choice. Indeed, if c denotes the constant marginal cost of knowledge, then in this particular case, the expected cost from $A \to B \to AB$ are $1.95c$ while the cost of a team AB are $2c$.

It is an open theoretical question whether the share of distributions for which an all-knowing team is optimal increases as the number of units of knowledge increases.

A.3 Organization and Labor Markets

So far, I have assumed that the wages, i.e., the cost of accessing knowledge, are given exogenously. In this subsection, I release this assumption and endogenize the cost of accessing knowledge. Wages are now general equilibrium outcomes of the labor market and determined simultaneously with the firms' organizational form.

To describe labor demand, I assume that there is a unit-firm for each distribution. For given wages, each firm determines whether to produce and if so with which organizational form. The organizational form determines the firm's labor demand. Aggregating all firms' demand over the entire distribution space yields the market demand. Wages are determined by the market clearing equilibrium.

Assume that a share of the population is endowed with knowledge A and the rest can provide knowledge B. A market-clearing wage is a pair c_A, c_B such that the demand for A-workers and for B-workers is equal to their supply. Figure 15 shows the outcome for an equal split within the population(a), and a market with a majority of A-workers. Note that the wage of A workers does not decrease proportionally to the increase in the share of of A workers in the population: The organizational adaptation of firms increases demand for A workers and mitigates the decrease in wages.

This result is consistent with the observation that higher-ranked agents 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.

Similarly, it is optimal for many firms to organize the workers as a team only if market wages for different knowledge workers is fairly homogeneous. To see this point, assume for a moment that $c_A = 0$. Then a worker with knowledge set A would solve all a-problems

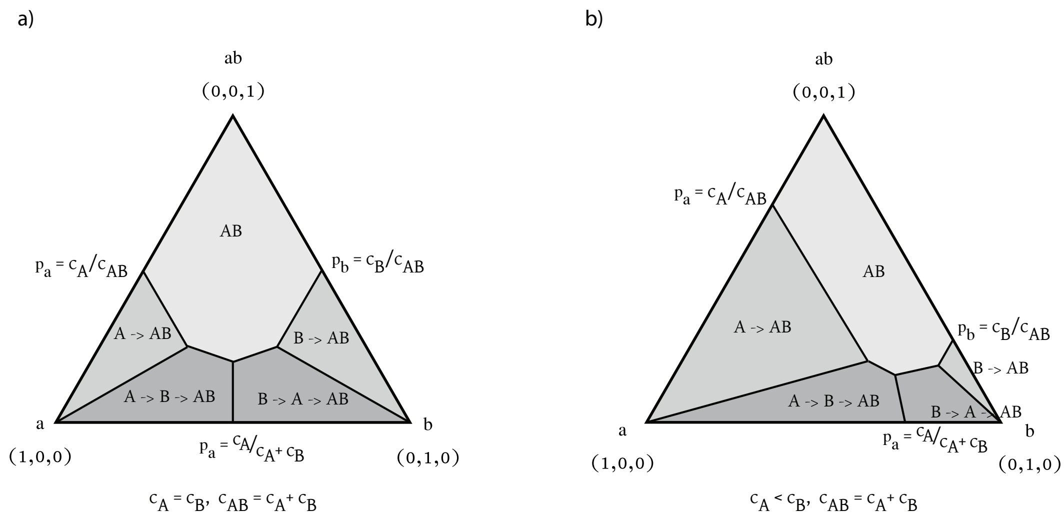


Figure 15: As the share of A-workers in the population increases, the relative cost of accessing A versus accessing B decreases, and knowledge sequences with A as the first stage knowledge set become more prevalent. The figure assumes that $c_{AB} = c_A + c_B$.

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. Thus, the observation of relative homogenous pay among members of a team may be an equilibrium outcome between the labor market and the organization of the firm.

A.4 Extension: Complementarity and Substitution in an Organizational General Equilibrium of Skilled and Unskilled Workers

Using a different interpretation of the model, it can fruitfully be employed to shed light on which extent skilled and unskilled workers are complements or substitutes.

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 those that learn one unit of knowledge 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}.

If every distribution occurred infinitely often and there is a limited supply of labor, then production only occurs at the most profitable vertices. In this case, every organizational form is a contingency sequence of length one, and every problem that is solved is addressed by the first-best knowledge set. To rule out this scenario, 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 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 3.5 allows us to compute the optimal organizational forms for all distributions in the distribution simplex. 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 kind of workers in the graph in figure 16, where the demand for A and for B workers have 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 16. 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 a 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 16 shows the corresponding partition of the distribution simplex into the regions where different organizational forms are optimal.

The constant demand curves in figure 16 show the subtle interaction between more and less skilled workers: They are partially substitutes, partially complements depending on the relative supply of both kind 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

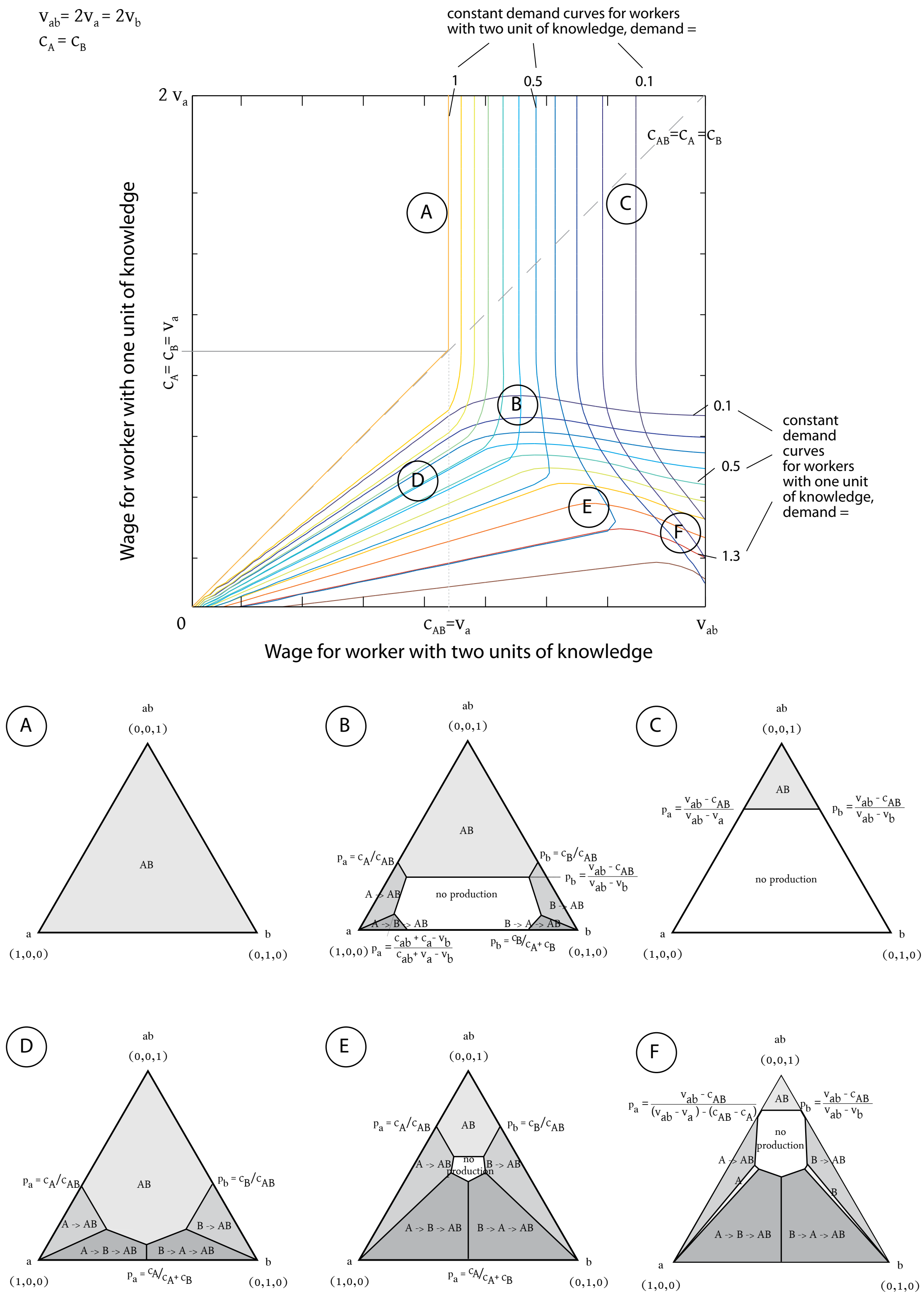


Figure 16: The top graph shows constant demand curves as the labor cost for both kinds of workers vary. Conversely, given the supply of the two kinds of workers we can consider the two demand curves where demand equals the labor supply. The intersection of these two curves corresponds to the equilibrium wage given the labor supply. The lower half of the

kind 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 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.

B Proofs

B.1 Proofs of Section 2.2

Lemma 1 An optimal organizational form always takes the form of $\mathcal{K}_1 \to \mathcal{K}_2 \to \mathcal{K}_3 \to ... \to \mathcal{K}_l$, where $\mathcal{K}_1$ represents knowledge set provided by the first agent(s) to address a problem drawn from the task distribution. Any problems not solved by the agent(s) with knowlegde $\mathcal{K}_1$ are then attempted agent(s) with knowledge $\mathcal{K}_2$ and so forth. Any task that remains unsolved after agents with knowledge $\mathcal{K}_l$ have attempted to solve it is discarded. Moreover, $i > j$ implies that $\mathcal{K}_i \nsubseteq \mathcal{K}_j$ for all i, j.

Proof: I prove the first part of the statement using induction by the number t of tasks $\mathfrak{t}$ in the support of the probability distribution p the firm faces:

$$t = \left|\{\mathfrak{t} \,|\, p(\mathfrak{t}) \neq 0\}\right|.$$

Step 1: $t = 1$.
Let $\mathfrak{t}$ be the single task that occurs with probability 1. Also, let $\mathcal{K}$ be the smallest knowledge set that can solve $\mathfrak{t}$, e.g., if $\mathfrak{t} = abc$ then $\mathcal{K} = \{A, B, C\}$. Then the general assumption that more extensive knowledge is more expensive, i.e., $\mathcal{K} \subset \mathcal{K}'$ implies $c_\mathcal{K} \leq c_{\mathcal{K}'}$, means that $\mathcal{K}$ is also the cheapest knowledge set that can solve $\mathfrak{t}$.

If the value of solving the task exceeds the cost of doing so, i.e., if $v_\mathfrak{t} \geq c_\mathcal{K}$, then the firm will optimally organize by hiring agents with $\mathcal{K}$. If it is not worth solving the task, i.e., if $v_\mathfrak{t} < c_\mathcal{K}$, then the firm will optimally discard all problems it faces.

In either case, the firm optimally organizes either as knowledge sequence with either one or none knowledge set.

Step 2: $t > 1$. Assume the statement is true for all distributions with a support of size less than t.
If the firm does not operate at all, i.e., if it discards all tasks it faces, then the statement is pathologically true for $l = 0$.

If the firm does operate, then agent(s) with some knowledge set $\mathcal{K}_1$ must be the first to attempt a task drawn from the distribution. If that knowledge set is part of an optimal organizational form, then the agent(s) must solve some tasks. Otherwise, the firm would be better off not hiring these agents. Once the agent(s) have solved some tasks, the distribution of all remaining, i.e., unsolved, tasks has a support of size less than t. By assumption, there is a knowledge sequence of the form $\mathcal{K}_2 \to \mathcal{K}_3 \to ... \to \mathcal{K}_l$ that optimally addresses the remainder distribution. Thus, the firm's otpimal organizational form takes the form of $\mathcal{K}_1 \to \mathcal{K}_2 \to \mathcal{K}_3 \to ... \to \mathcal{K}_l$.

Since the number of potential tasks in any economy with N units of knowledge is limited, this establishes the first part of the claim.

I prove the second part of the statement by proof by contradiction: Assume that some firm optimally organizes by

$$OF = \mathcal{K}_1 \to \mathcal{K}_2 \to \mathcal{K}_3 \to ... \to \mathcal{K}_i \to ... \to \mathcal{K}_l$$

with $\mathcal{K}_i \subseteq \mathcal{K}_j$ form some $i > j$. Consider now instead the organizational from

$$OF' = \mathcal{K}_1 \to \mathcal{K}_2 \to \mathcal{K}_3 \to ... \to \mathcal{K}_{i-1} \to \mathcal{K}_{i+1} \to ... \to \mathcal{K}_l.$$

Then OF' solves all tasks that OF solves, since all tasks solved by $\mathcal{K}_i$ are also solved by $\mathcal{K}_j$. Moreoever, OF' is cheaper to maintain, since tasks that remain unsolved after agent(s) with $\mathcal{K}_{i-1}$ have attempted to solve them are not attempted by agents with knowledge $\mathcal{K}_i$, and instead passed on directly to agent(s) with knowledge $\mathcal{K}_{i+1}$. Thus, OF' generates more profit than OF, and hence OF cannot be optimal.

This establishes the second part of the lemma and completes the prove. □

Lemma 2 The region in the distribution space where a particular organizational form is optimal is convex.

Proof: Denote the profit a firm generates if it uses a contingency sequence $OF = \{\mathcal{K}_1, ..., \mathcal{K}_l\}$ when facing distribution $p(\mathfrak{t}$ by $\Pi(p; OF)$. The revenue the firm generates is given by the value of all problems solved:

$$\text{Revenue}(p; OF) = \sum_{\mathfrak{t} \text{ solved by } OF} p(\mathfrak{t}) v_{tt}.$$

The cost the firm incurres is the cost of the knowledge sets hired for the time periods they are hired. At every level, agents with the corresponding knowledge set are only needed and only paid if agents in the previous stages were not able to solve the task, i.e.,

$$\text{Cost}(p; OF) = \sum_{i=1}^{l} Prob\big(\mathfrak{t} \text{ not solved by } \mathcal{K}_1, ..., \mathcal{K}_{i-1}\big) \cdot c_{\mathcal{K}_i}.$$

Therefore, the overall profit

$$\Pi(p; OF) = \sum_{\mathfrak{t} \text{ solved by } OF} p(\mathfrak{t}) v_{tt} - \sum_{i=1}^{l} Prob\big(\mathfrak{t} \text{ not solved by } \mathcal{K}_1, ..., \mathcal{K}_{i-1}\big) \cdot c_{\mathcal{K}_i}$$

is linear in p.

Now assume that some organizational form $\hat{OF}$ is optimal at two probability distributions p_1 and p_2. Then for $i = 1, 2$

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

Therefore

$$\begin{aligned}
\Pi(\lambda \cdot p_1 + \lambda \cdot p_2; \hat{OF}) &= \lambda \cdot \Pi(p_1; \hat{OF}) + (1 - \lambda) \cdot \Pi(p_2; \hat{OF}) \\
&\geq \lambda \cdot \Pi(p_1; OF) + (1 - \lambda) \cdot \Pi(p_2; OF) \\
&= \Pi(\lambda \cdot p_1 + \lambda \cdot p_2; OF)
\end{aligned}$$

In other words, $\hat{OF}$ is optimal for any convex combination of p_1 and p_2. This establishes the claim. □