Codes in Organizations[1]

Jacques Crémer
Université de Toulouse, IDEI-GREMAQ and CEPR

Luis Garicano
University of Chicago and CEPR

Andrea Prat
London School of Economics and CEPR

February 28, 2005

[1]We thank Phillippe Aghion, Karen Bernhardt, Wouter Dessein, Matthias Dewatripont, Bob Gibbons, Jerry Green, Daniel Hoffman, Augustin Landier, Jean Tirole, workshop participants at the CEPR/Toulouse Organizations workshop, the NBER Organizational Economics Workhop, and seminars at Harvard, MIT, UCLA and USC for useful comments and Pedro Vicente for outstanding research assistance.

Abstract

A code is a technical language that members of an organization learn in order to communicate among themselves and with members of other organizations. What are the features of an optimal code and how does it interact with the characteristics of the organization? This paper develops a theory of codes in organizations and studies the properties of optimal codes. There exists a fundamental tradeoff between choosing a specialized code, which simplifies communication within divisions, and a common code which facilitates communication between divisions. In turn, code specialization interacts with other endogenous features of the organization like its scope and the level of centralization of communication. We show that an exogenous decrease in information costs leads to larger, less centralized organizations which rely on common codes, rather than hierarchies, for communication. Our analysis illuminates some aspects of the impact of IT innovation on organizational structure described by the empirical literature. We show that the causal mechanism we propose is consistent with the evidence in some detailed case studies of decentralization and organization. We also study how conflicts of interests among organizations can lead to the adoption of biased codes or to excessive code variety.

1 Introduction

Accounting systems, human resource and other organizational data bases are particular types of codes.[1] In recent years, the management of these codes within firms has become more centralized, while communications have become less hierarchical and while, at the same time, decision making has become more decentralized. Robert J. Herbold, Chief Operating Officer for Microsoft from 1994 to 2001, described this apparent paradox as follows: “standardizing specific practices and centralizing certain systems also provided, perhaps surprisingly, benefits usually associated with decentralization.” In this paper, we develop a theory of codes and of their influence on the organization of firms, which enables us to analyze these links and provides a number of broader insights on how communication costs shape different aspects of organizational design and of the interaction between individual agents in these organizations.

Our theory builds on previous informal discussions in the economics literature, most notably by Arrow (1974), which emphasize the role of specialized codes in reducing the constraints on organizational performance imposed by individual bounded rationality.[2] The main focus of our analysis is to understand when subsets of agents who deal repeatedly with each other choose to design specialized codes, which are more efficient for their specific circumstances and when, instead, they prefer to use less precise common codes.

We start, in Section 2, by building a simple model of codes. Our agents are subject to two forms of bounded rationality. They have a limited ability to learn codes, but they also have a limited ability to solve problems. The use of a code facilitates communication, which in turn decreases the cost of incomplete information. We focus on the structure of optimal codes, and on their organizational consequences. We will not study the other obvious important issue in this framework, which is the optimal richness of codes and the trade-offs between investment in learning a new code and spending more resources on production activities.[3]

We identify some of the properties of efficient codes. Efficient codes use precise words for frequent events and vague words for more unusual ones. A more unequal distribution of events increases the value of the creation of a specialized

[1]For our purpuses, an organizational code is a technical language that an organization adopts to facilitate communication. It contains expressions that have a certain, well-specified meaning only within the set of users of that code. Those same expression convey a different meaning, or no meaning at all, to the rest of the population.

[2]Despite the existence of this pathbreaking initial analysis by Arrow, and the fact that specalied codes are widespread, there has been little formal analysis of the properties of codes, and their consequences for the organizations of firms have been nearly completely neglected. We review the brief literature in the conclusion.

[3]Garicano (2000) provides an example of the study of a similar tradeoff. The agents can learn how to process more tasks, but this has a cost.

code, since the precision of the words can be more tightly linked to the characteristics of the environment.

We show that when more than two agents communicate with one another, bounded rationality imposes sharply decreasing returns to the diversity of codes. Tailoring words to the needs of particular agents restrains the group of agents which can usefully learn them. As a consequence, agents will use either entirely separate codes or common codes: "dialects" cannot be optimal. This code commonality is a key determinant of the decreasing returns to scope in organizations, and shapes both organizational scope and their use of integrating mechanisms, which we study next.

When would a set of agents choose to use a common code over all its parts, so that all of them can communicate with each other? In Section 3, we argue that the choice would trade off the improved coordination between different services and the resulting degradation of communications within services. We identify the variables that determine the terms of these trade-offs: a common code is more likely to be adopted if the degree of synergy among services is high, the cost of imprecise communication is high and when the distribution of events in different services is similar.

We illustrate the limits that bounded rationality places on the richness of a common code, and thus the limits imposed by codes on organizational scope, with an example of a friendly fire incident between US airplanes and Army helicopters. A complex set of rules meant to facilitate between service communication was actually understood differently by the different services and resulted in a tragic communication failure which we explore.

Hierarchies provide an alternative method for coordinating two services. We represent a hierarchical superior as a translator, who enables services with different codes to cooperate. When communication costs are high, hierarchies are more efficient; when they are low, common codes and horizontal communications are more efficient. The reorganization of Microsoft under Robert J. Herbold in the 1990s, which we discuss in some detail in section 3.4.2 provides an illustration of these trade-offs and of the substitution of hierarchical communication by horizontal communication mediated by a common code.

Finally, section 4 studies how conflicts between organizations shape code adoption and organizational structure. We identify a first mover advantage: a shared code is suboptimally skewed towards the needs of early adopters. Moreover, there can exist too little code commonality, as, for each group of users, investing in a common code generates positive externalities towards other users. This implies that coordination within organizations will be better than coordination between organizations, which appears consistent with the evidence.[4] Some

[4] See Simester and Knez (2002), which compares coordination with internal and with external

evidence for these strategic effects and inefficiencies in adoption of common codes is provided by the interactions between the firms that engineered the B-2 stealth bomber, which we discuss in section 4.3. Thus taking into account these strategic considerations that arise when individuals can separately choose codes yields a number of insights on organizational dysfunctions and the way that organizations may deal with them. For instance, organizations must limit the flexibility of individual divisions (or individual agents) to choose their own codes, in order to avoid excessive code variety with the ensuing loss in between service communication..

In section 5, we discuss links with previous literature, and directions for future research. Appendix A presents the proof of proposition 6 whereas Appendix B presents an extension of the results which are presented in the body of the paper.

2 A Theory of Codes

In this section, we lay down a simple theory of the choice of codes, beginning by the case of two agents who need to communicate with each other, in subsections 2.1 to 2.3, and turning to the case of an agent who needs to communicate with several other agents in subsection 2.4.

2.1 Model

A "salesman" must communicate to an "engineer" information about the characteristics x of potential clients; characteristics are drawn with probability f_x from a finite set X.

A code $\mathcal{C}$ is a partition $\{W_1, W_2, \ldots, W_K\}$ of the set X. By uttering the word k, the salesman lets the engineer know that the characteristics x belong to the subset W_k. We will speak about the the *breadth* of word k, the number of events $n_k \equiv \sharp W_k$ that it contains, and about its frequency, $p_k \equiv \sum_{x \in W_k} f_x$.

The bounded rationality of the agents is represented by the maximum number of words, K, that they can learn.

Having received word k from the salesman, the engineer still must identify the precise characteristics x of the client. This *diagnosis* stage takes time and/or energy. The broader the word, the harder the search: the diagnosis cost is an increasing function d of the breadth n_k of k. The expected diagnosis cost of code

suppliers in the provision of similar parts by a high tech firm. They find that coordination with external suppliers involves slower reactions and less information exchange on the product design than coordination with internal suppliers on similar pieces, which is what the theory we develop would lead us to expect.

$\mathcal{C}$ is therefore

$$D(\mathcal{C}) = \sum_{k=1}^{K} p_k d(n_k). \tag{1}$$

If, as we shall assume under further notice, the value of serving a client is high enough that the organization would never want to exclude clients with certain values of x just for the sake of saving on diagnosis costs, the profit maximizing code is the code that minimizes the expected diagnosis cost.

The following simple example might make these definitions more concrete. The salesman is an interior decorator while the engineer is a craftsman who collaborates with the decorator. The characteristics x are the client's preference about some fixture. Through an informal conversation, the client conveys x to the decorator ("The door knob should have that self-ironic retro feel that is so hot in Paris these days"). The decorator then telephones the craftsman, who has access to a well-stocked warehouse, to ask him to install the appropriate fixture. Our key assumption is that the decorator cannot convey all the information he has acquired to the craftsman. Instead, they share a code to describe fixtures, and their bounded rationality necessitates a coarse code. After receiving the approximate description of the desired fixture, the craftsman collects all the fixtures in his warehouse that fit the description and transports them to the client's house for perusal. The diagnosis cost is given by the time spent putting together the possible fixtures and by the cost of transportation. It is natural to assume that the diagnosis cost increases in the coarseness of the description.[5]

2.2 Optimal codes

In this subsection, we derive some properties of the optimal code that two agents would use.

Proposition 1. *In an optimal code, broader words describe less frequent events: if $n_k > n_{k'}$, then $f_x \leq f_{x'}$ for any $x \in W_k$ and $x' \in W_{k'}$.*

Proof. Let k and k' be two words such that $n_k > n_{k'}$ in an optimal code $\mathcal{C}$.

From $\mathcal{C}$, construct a new code $\tilde{\mathcal{C}}$ by moving event x from word k to word k'

[5] A further interpretation of this cost of receiving an imprecise message or word is the mispecification of the product that results when the engineer cannot fit precisely the product to the customer needs. This mispecification cost is, like the diagnosis cost, higher the broader the word.

and event x' from word k' to word k. We must have

$$\begin{aligned}
0 &\geq D(\mathcal{C}) - D\left(\tilde{\mathcal{C}}\right) \\
&= d(n_k)\, p_k + d(n_{k'})\, p_{k'} \\
&\qquad - d(n_k)\,(p_k + f_{x'} - f_x) - d(n_{k'})\,(p_{k'} + f_x - f_{x'}) \\
&= (d(n_k) - d(n_{k'}))\,(f_x - f_{x'}),
\end{aligned}$$

which proves the result. □

Proposition 1 implies that events of similar frequencies are grouped together in an optimal code: if $f_x < f_{x'} < f_{x''}$ and x and x'' belong to the same word, then x' also belongs to that word.

Proposition 1 relates word breadth to event frequency. The next result, Proposition 2, relates word length to word frequency; it requires the additional assumption that the function d is (weakly) convex.

Proposition 2. *Assume*

$$d(n+1) - d(n) \geq d(n'+1) - d(n') \tag{2}$$

whenever $n \geq n' \geq 1$. *Unless integer constraints make it impossible, in an optimal code broader words are used less frequently: if* $n_k - n_{k'} \geq 2$*, then* $p_{k'} \geq p_k$.

Proof. Let k and k' be two words such that $n_k - n_{k'} \geq 2$ in an optimal code $\mathcal{C}$.

From $\mathcal{C}$, construct a new code $\tilde{\mathcal{C}}$ by moving an event x from word k to word k'.

We have

$$\begin{aligned}
D(\mathcal{C}) - D\left(\tilde{\mathcal{C}}\right) &= d(n_k)\, p_k + d(n_{k'})\, p_{k'} - d(n_k - 1)\,(p_k - f_x) \\
&\qquad\qquad - d(n_{k'} + 1)\,(p_{k'} + f_s) \\
&= [d(n_k) - d(n_{k-1})]p_k - [d(n_{k'}+1) - d(n_{k'})]p_{k'} \\
&\qquad\qquad + f_x[d(n_{k-1}) - d(n_{k'+1})] \\
&\geq [d(n_k) - d(n_{k-1})]p_k - [d(n_{k'}+1) - d(n_{k'})]p_{k'} \\
&\qquad\qquad (d \text{ is increasing}) \\
&\geq [d(n_k) - d(n_{k-1})](p_k - p_{k'}) \quad \text{(by (2))}.
\end{aligned}$$

Because $D(\mathcal{C}) - D\left(\tilde{\mathcal{C}}\right) \leq 0$, we must have $p_k \leq p_{k'}$, which proves the result. □

We have assumed that events could be allocated between words arbitrarily. In some instances, however, the "meaning" of events imposes constraints on the languages which can be constructed. For example, if we are partitioning the color spectrum into discrete color words, most words will group contiguous points of the spectrum. In Appendix B we extend Propositions 1 and 2 to environments where events have a natural ordering: we show that for two contiguous words, the broader word is used less often and describes events with a lower average frequency.

2.3 The value of a code

Firms face environments with varying degrees of uncertainty, and this affects the value of the codes that they use. In this subsection, we show that the value of a minimal common language is greater when there is less uncertainty. On the other hand, more uncertainty increases the benefits of enriching a code, by incorporating new words.

We consider two distributions of the same set events, f and $\tilde{f}$. The distribution $\tilde{f}$ is more unequal than f if for any n any word that contains the n events with the smallest frequency according to f has a greater probability according to f than to $\tilde{f}$. More precisely:[6]

Definition 1. *The distribution $\tilde{f}$ is more unequal than the distribution f if $f_W \geq \tilde{f}_W$ for any word W such that $x \in W$ and $x' \notin W$ implies $f_x \leq f_{x'}$. If for such a word $f_W > \tilde{f}_{W'}$, then f is strictly more unequal than $\tilde{f}$.*

With this definition, we can show that communication cost is decreasing in inequality:[7]

Proposition 3. *If distribution $\tilde{f}$ is more unequal than distribution f, the minimal diagnosis cost associated with $\tilde{f}$ is (weakly) smaller than the minimal diagnosis cost associated with f.*

Proof. By proposition 1, we can name the words in the optimal code for distribution f in such a way that $k < k'$ implies that $f_x \leq f_{x'}$ for all $x \in W_k$ and $x' \notin W_{k'}$. This implies

$$d\left(n_{k-1}\right) - d\left(n_k\right) \geq 0 \text{ for all } k. \tag{3}$$

[6] Note that the inequality of distribution defines an order on distributions. f is more unequal than $\tilde{f}$ if and only if the distributions are equal up to a renaming of events. Furthermore, it is easy to prove that if f is strictly more unequal than $\tilde{f}$, then $\tilde{f}$ cannot be strictly more unequal than f.
Of course, these definitions are very close in spirit to first order stochastic dominance.

[7] The assumption that the diagnosis cost is convex is not needed.

Let $P_k = \sum_{k' \leq k} p_{k'}$ and $\tilde{P}_k = \sum_{k' \leq k} \tilde{p}_{k'}$; then

$$P_k > \tilde{P}_k \text{ for all } k. \tag{4}$$

We have

$$\begin{aligned}
\sum\nolimits_k p_k d\left(n_k\right) &= P_1 d\left(n_1\right) + \left(P_2 - P_1\right) d\left(n_2\right) + ... \\
&\quad + \left(P_{K-1} - P_{K-2}\right) d\left(n_{K-1}\right) + \left(1 - P_{K-1}\right) d\left(n_K\right) \\
&= P_1 \left(d\left(n_1\right) - d\left(n_2\right)\right) + ... \\
&\quad + P_{K-1} \left(d\left(n_{K-1}\right) - d\left(n_K\right)\right) + d\left(n_K\right) \\
&\geq \tilde{P}_1 \left(d\left(n_1\right) - d\left(n_2\right)\right) + ... \\
&\quad + \tilde{P}_{K-1} \left(d\left(n_{K-1}\right) - d\left(n_K\right)\right) + d\left(n_K\right) \text{ (by (3) and 4))} \\
&= \sum\nolimits_k \tilde{p}_k d\left(n_k\right).
\end{aligned}$$

This concludes the proof, as $\sum_k \tilde{p}_k d\left(n_k\right)$ is not smaller than the minimal diagnosis cost for $\tilde{p}$. □

In an unequal distribution, there are a few extremely likely events and a large number of rare events. Communication costs are low, because the optimal code assigns likely events to narrow words, and narrow words are very probable. The worst-case scenario occurs when all events are equiprobable: words will divide the event space into equiprobable sets, and this will impose a high communication cost.

An immediate consequence of proposition 3 is that increasing the number of words from 1 to $K > 1$ lowers communication costs more for more unequal distributions. On the other hand, moving from K words to a very large number of words (perfect communication) lowers communication less for a more unequal distribution.

2.4 Shared codes and dialects

The organizational analysis of section 3 relies heavily on the insight that communications between organizational units require the use of a common code. In this subsection, we provide some background for this analysis by studying the choice of code by an engineer who needs to communicate with two[8] salesmen, A and B, who face the same set of events X but have different distributions f_x^A and f_x^B. The stark model which we are using implies that there are no "dialects"; at the end of

[8]It should be clear that the assumption that there are two agents is made only for ease of exposition and is totally unnecessary for the results.

this subsection, we explain how it could be enriched in order to allow for their existence. We also argue that this should not affect our main insights.

There are three agents, the engineer and salesmen A and B, who face the same set of events X but have different distributions f_x^A and f_x^B. Per period, salesman A receives requests from m_A clients, and salesman B requests from m_B clients.

Each of the three agents can learn at most K words. Agent A uses code $\mathcal{C}_A$, agent B uses code $\mathcal{C}_B$, and the engineer, who must understand both agents, knows code $\mathcal{C}_A \cup \mathcal{C}_B$ (of course, he only uses the relevant part of this code when communicating with a salesman).

For instance, with $X = \{1, 2, 3, 4, 5, 6\}$, we could have

$$\mathcal{C}_A = \{\{1,4\}, \{2,5\}, \{3,6\}\}, \tag{5}$$
$$\mathcal{C}_B = \{\{1,2,3\}, \{4,5,6\}\}. \tag{6}$$

Then, the engineer must know five words, while A knows 3 and B knows 2.

Proposition 4 shows that the same code, which saturates the rationality constraints of all the agents, will be used in communicating with both salesmen.

Proposition 4. *The optimal codes contain K words and satisfy $\mathcal{C}_A = \mathcal{C}_B$.*

Proof. Clearly, an optimal code saturates the bounded rationality of the engineer, with $\mathcal{C}_A \cup \mathcal{C}_B$ containing K words. We must still prove $\mathcal{C}_A = \mathcal{C}_B$.

Suppose that $\mathcal{C}_A \neq \mathcal{C}_B$, which implies that both $\mathcal{C}_A$ and $\mathcal{C}_B$ contain at most $K - 1$ words. We call $k^{\min}$ be the narrowest noncommon word of these two codes,[9] and, without loss of generality, assume that it belongs to $\mathcal{C}_A$.

Transform $\mathcal{C}_B$ into $\tilde{\mathcal{C}}_B$ by adding $k^{\min}$ as follows: $k \in \tilde{\mathcal{C}}_B$ if and only if $k = k^{\min}$ or $W = W'/(W' \cap W_k)$ for some $W' \in \mathcal{C}_B$. Because $\#\tilde{\mathcal{C}}_B = \#\mathcal{C}_B + 1 \leq K$, the bounded rationality of agent B is still satisfied, and because $\#(\mathcal{C}_A \cup \tilde{\mathcal{C}}_B) = \#(\mathcal{C}_A \cup \mathcal{C}_B)$, the bounded rationality of the engineer is also satisfied

For every event $x \in X$, the length of the word in $\tilde{\mathcal{C}}_B$ that contains x is not larger than the length of the word in $\mathcal{C}_B$ that contains x. Moreover, as $\tilde{\mathcal{C}}_B$ contains one more word than $\mathcal{C}_B$, at least one event must be in a strictly narrower word in $\tilde{\mathcal{C}}_B$ than it was in $\mathcal{C}_B$. The new codes are strictly more efficient than the original ones, which proves the result. $\square$

Two examples can illustrate the proof of proposition 4. First, with $K = 5$, consider the codes of equations (5) and (6). The narrowest noncommon words[10] are $\{1,4\}$, $\{2,5\}$, and $\{3,6\}$; let us introduce $\{1,4\}$ into $\mathcal{C}_B$. Then, $\tilde{\mathcal{C}}_B =$

[9]That is $k \in \operatorname{argmin}_{\tilde{k}} n_{\tilde{k}}$ subject to $W_{\tilde{k}} \in \mathcal{C}_1 \cup \mathcal{C}_2$ and $W_{\tilde{k}} \notin \mathcal{C}_1 \cap \mathcal{C}_2$.

[10]We are taking some liberty with our terminology. Strictly speaking, we have defined a word to be the *name* of a set of events. In this discussion, a word is the set of events itself. This should create no confusion, and lighten considerably the exposition.

$\{\{1,4\},\{2,3\},\{5,6\}\}$; every event is now represented by a shorter word, and diagnosis cost must go down while the engineer must still learn five words, $\{1,4\}$, $\{2,3\}$, $\{5,6\}$, $\{2,5\}$ and $\{3,6\}$. Notice that $\mathcal{C}_A$ and $\tilde{\mathcal{C}}_B$ are still not efficient: if we add $\{2,5\}$ to $\tilde{\mathcal{C}}_B$, the engineer must still learn five words ($\{1,4\}$, $\{2,5\}$, $\{3\}$, $\{6\}$ and $\{3,6\}$), and the codes are more efficient.

A more complicated example starts from

$$\mathcal{C}_A = \{\{1,2,3\},\{4,5,6\},\{7,8,9,10\},\{11,12,13,14,15,16\}\},$$
$$\mathcal{C}_B = \{\{1,4,7,11\},\{2,5,8,12\},\{3,6,9,13\},\{10,14,15,16\}\}.$$

If we take $\{1,2,3\}$ as the narrowest noncommon word, the new code for B is

$$\tilde{\mathcal{C}}_B = \{\{1,2,3\},\{4,7,11\},\{5,8,12\},\{6,9,13\},\{10,14,15,16\}\}.$$

All events but 10, 14, 15 and 16 are now represented by shorter words.

We have shown that the engineer will use the same code to speak to both salesmen? Which code will be chosen? It is easy to see it will be the code which would have been chosen had the engineer faced only one salesman with a distribution of events equal to the expected distribution of events for the two salesmen. This is formalized in the next proposition.

Corollary 1. *Propositions 1 and 2 apply as stated to the common code if one defines*

$$f_x = \frac{m_A f_x^A + m_B f_x^B}{m_A + m_B}.$$

Of course, in reality, we would expect the engineer (and more generally hierarchical superiors) to know more words than salesmen, for two reasons. First, if bounded rationality imposed a cost on the acquisition of language rather than an absolute limit on the number of words that can be learned, it would be optimal for the engineer to learn more words than the salesmen. Second, if some agents have different abilities, it would be optimal to choose agent who is able to learn the most words as the engineer. Presumably, it would be optimal for the salesmen to share some words, while using specific words to communicate to the engineer events that they encounter much more often than the other salesman.

Garicano (2000) and Garicano and Rossi-Hansberg (2003) build theories of hierarchies where agents who have the ability to complete more tasks are chosen as "supervisors"; it may be possible to build similar theories, where the supervisors are able to learn more words. We take a small step in that direction in 3.3, where we assume that the firm can hire, at an additional cost, an agent with a larger K.

2.5 Two Word Codes

To facilitate the organizational analysis that follows, we modify the framework which we have used up to this point: we consider two word codes, a linear diagnosis cost, and assume a continuum of events.

Each salesman deals with consumers $x \in [0,1]$ with cumulative distribution functions

$$F(x;b) = (1-b)\,x + bx^2,$$

and density

$$f(x;b) = (1-b) + 2bx,$$

where $b \in [-1,1]$ measures the inequality of the distribution of events.

The diagnosis cost is linear: if the engineer knows that the client's characteristics fall in an interval, his diagnosis cost is $s\lambda$ times the size of that interval.

With an engineer and a single salesman, we can adapt proposition 1 to show that they will use a "right word" and a "left word", separated by $\hat{x}(b)$ which minimizes

$$s(x) = F\left(x;b\right)x + \left(1 - F\left(x;b\right)\right)\left(1-x\right), \tag{7}$$

which implies,[11]

$$\hat{x}(b) = \mathrm{argmin}_x\, s(x) = \frac{1}{6b}\left(3b - 2 + \sqrt{(3b^2+4)}\right). \tag{8}$$

The corresponding expected diagnosis cost is

$$\lambda D^*(b) = \lambda \times \frac{8 + 36\,b^2 - (4 + 3\,b^2)^{\frac{3}{2}}}{54\,b^2}. \tag{9}$$

The variations of D^* and $\hat{x}$ as a function of b are represented on figure 1. The diagnostic cost is convex in b, and is maximal for $b = 0$. The cutoff event $\hat{x}$ increases close to linearly in b.

3 Integration, separation and hierarchy

In the previous section, we studied the optimal code for exogenously given organizations. In this section, we optimize jointly on organizational structure and code: who should communicate with whom? what code should they use?

We develop a simple model with two services, each composed of one salesman and one engineer. We shall study communication and coordination among the two

[11] The function s is not convex in x. However, the quadratic function $s'(x)$ is convex with $s'(0) < 0$. Hence, there exists a unique $\hat{x}(b) \in (0,1)$ such that $s'(\hat{x}(b)) = 0$, $s(x)$ is decreasing on $[0, \hat{x}(b)]$ and increasing on $[\hat{x}(b), 1]$. Profits are single peaked, with a maximum at $\hat{x}(b)$.

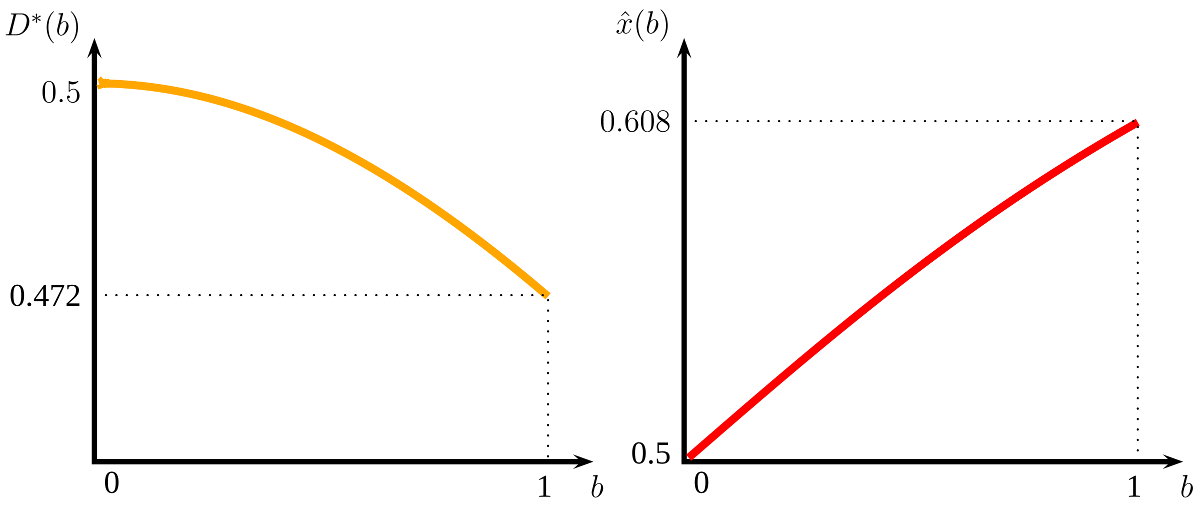

Figure 1: The variations D^* and $\hat{x}$ (equations (9) and (8)) as a function of b.

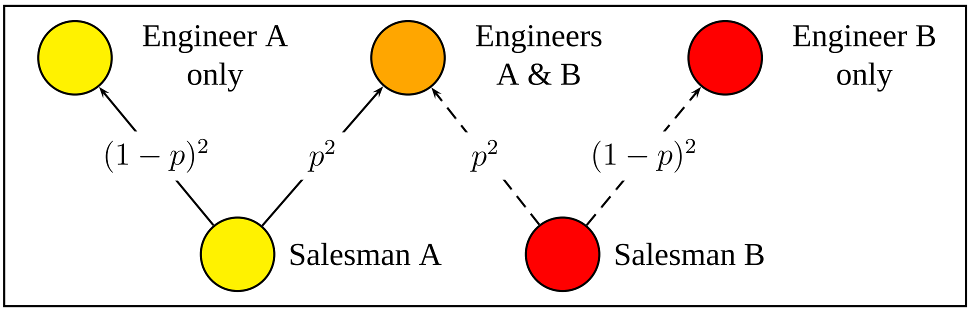

Figure 2: This figure illustrates the synergies between the two services discussed in section 3.1.

services, focussing on three possible organizational forms: *separation*, where the two services use different codes; *integration* where the two services share the same code; *translation*, where there exists a hierarchical structure supplying an interface between the services. We will identify the environments in which each of these organizational forms is optimal.

3.1 Synergies

In order to model the benefits of coordination between the services, we assume that they can help each other manage overflows of clients. A service has limited capacity: in each period, it can accommodate only one client. On the other hand, the number of clients who knock at its door is random, it can be zero, one or two. With positive probability, there is excessive load in one service and excessive capacity in the other. When this happens, the firm benefits from diverting some business from the overburdened service to the other.

More precisely, each engineer has the ability to attend to the needs of at most one client. The client arrival process is as follows (see Figure 2):

$$y = \begin{cases} 0 & \text{with probability } p, \\ 1 & \text{with probability } (1-2p), \\ 2 & \text{with probability } p, \end{cases}$$

where p belongs to the interval $[0, 1/2]$. This arrival process captures the effect of the variability in the expected number of clients of each type. If p is low, then each salesman is likely to find one client per period of each type. When p is high, although on average still 1 client is arriving, it is quite likely that either none or 2 will arrive. Thus p measures the importance of the synergy between the two services: a high p means that the services are likely to need to share clients, while a low p means that each service is likely to have its capacity fully utilized.

The two salesmen face different distributions of consumers, which, simplicity, are assumed to be symmetric to each other. Salesman A face consumers whose characteristics are drawn from the distribution $F(x; b)$, while salesman B faces distribution $F(x; 1-b)$. The parameter $b \in [0, 1]$ measures the inequality of the probability of different events for each salesman. Note that the distribution of characteristics of clients over all the clients who approach the firm is $F(x; 0)$.

The profit of the firm when it solves a client's problem is 1. The per-client diagnosis costs is λ: if the engineer knows that the client's characteristics fall in an interval of size s, his diagnosis cost is $s\lambda$. We assume that the diagnosis cost is sufficiently high to ensure that information must transit through a salesman before being sent to an engineer ($\lambda > 1$) but not so high that profit risks becoming negative ($\lambda < 2$).

Remark 1. In order to lighten the notation, in the main text we discuss a symmetric version of our model. Propositions[12] 5, 6 and 7 generalize basically as is to an "extended" version, where

a) the distributions of characteristics of the clients are any $F(x; b_A)$ and $F(x; b_B)$, without necessarily having $b_A = -b_B$;

b) the salesman of service $i \in \{A, B\}$, receives 0 client with probability p_0, 1 client with probability p_1, and 2 clients with probability p_2. Thus, there are two measures of the synergy between the two services: when p_0 is high, each service is more likely to be able to accommodate clients who arrived at the other service. When p_2 is high, each service is more likely to be able to use the help of the other service.

[12] A proposition analogous to 8 also holds.

We comment on the consequences of these more general assumptions in footnotes, and, in appendix A, we provide the proof of proposition 6 for the extended version.

3.2 Integration or separation?

In this section, we study the choice between segregation of services, where the salesman from service i only communicates with the engineers of his service, and integration where a salesman can communicate with either engineer.

In an *integrated organization*, a salesman must be able to communicate to the engineer from the other service the needs of his customer: indeed, as seen above, because $\lambda > 1$ sending the problem to the engineer without explanation is not profitable. The proof of proposition 4 can easily be adapted to show that the two services must use the same code, and the proof of corollary 1 can be adapted to prove that the optimal common language is $\hat{x}(0) = \frac{1}{2}$, the language which would be optimal with one engineer receiving messages from one salesman for which the density of characteristics is the average density of the two services, $F(x;0)$. Total profits are

$$\Pi^I = 2(1-p+p^2)\left(1-\lambda D^*(0)\right), \tag{10}$$

where $1-\lambda D^*(0) = 1-\frac{1}{2}\lambda$ is the expected total profit from serving one customer and $(1-2p)+p+p^2 = 1-p+p^2$ is the expected number of customers served by each service.

By (9), in isolation, each service has profits $(1-p)\left(1-\lambda D^*(b)\right)$, where $(1-p)$ is the expected number of customers it accommodates and $1-\lambda D^*(b)$ the profit from serving one customer (note that, by symmetry $D^*(b) = D^*(1-b)$). Therefore, the total profit in a *separated organization* is

$$\Pi^S = 2(1-p)\left(1-\lambda D^*(b)\right). \tag{11}$$

It is easy to check that the profits of a separated organization are positive.

The profits from an integrated organization are greater than the profits from a separated organization if[13,14]

$$\frac{p^2}{1-p} \geq \lambda \frac{D^*(0)-D^*(b)}{1-\lambda D^*(0)}. \tag{12}$$

[13]It is easy to check that there exist parameter values that lead to each one of these choices. For instance if $\lambda = 1.5$ and $p = 0.25$, then the difference between the two sides of (12) is a concave function of b, which is positive on $(-0.684, 0.684)$ and negative outside this interval.

[14]In the extended version of the model of Remark 1, equation (12) reads

$$\frac{p_0 p_2}{p_1+p_2} \geq \lambda \frac{D^*(\frac{b_A+b_B}{2}) - \frac{D^*(b_A)+D^*(b_B)}{2}}{1-\lambda D^*(\frac{b_A+b_B}{2})}.$$

The choice between these two organizational forms trade off the synergy gain with the loss in precision of communications due to the worsening of the code used.

The following proposition describes the comparative statics of the choice between an integrated and a separated organization.[15]

Proposition 5. *An integrated form becomes relatively more profitable compared to the segregated form when a) the diagnosis cost λ decreases, b) the synergy parameters p increases, or c) the heterogeneity of the two client distributions b decreases: let (λ^*, p^*, b^*) and (λ, p, b) be two sets of parameters such that $\lambda \leq \lambda^*$, $p \geq p^*$, $b \leq b^*$, with at least one of the inequalities strict, then the difference of profits between the integrated and the segregated form is larger under (λ, p, b).*

Proof. From (10) and (11), the difference in profits between the integrated and the separated organizational forms is (twice)

$$\left(1 - p + p^2\right)\left(1 - \lambda D^*\left(0\right)\right) - \left(1 - p\right)\left(1 - \lambda D^*(b)\right).$$

Its derivative with respect to λ

$$-(1 - p + p^2)D^*\left(0\right) + (1 - p)D^*(b)$$

is negative because $D^*\left(0\right) \geq D^*(b)$.

The derivative with respect to p is

$$2p\left(1 - \lambda D^*\left(0\right)\right) + \lambda\left(D^*\left(0\right) - D^*(b)\right) > 0.$$

Finally, the derivative with respect to b is

$$(1 - p)\lambda\frac{d}{db}D^*(b) < 0,$$

because of (9). □

Separate codes are preferable when synergies are relatively low, when the underlying probability distributions confronting the different units are sufficiently different, and when diagnosis costs are high so that there is a high premium on communicating precisely. Conversely, increases in synergies, in the equality of the distributions or decreases in diagnosis costs result in more code commonality.

[15]In the extended version of the model presented in Remark 1, an increase of either one of p_0 or p_2 is sufficient to make the integrated form relatively more profitable that the segregated form. The comparative statics with respect to the b_is assume that their mean stays constant, but that they become more dissimilar. Formally, the comparative statics read as follows: "If there are two sets of parameters such that $\lambda \geq \lambda^*$, $p_0 \geq p_0^*$, $p_2 \geq p_2^*$, $(b_A + b_B)/2 = (b_A^* + b_B^*)/2$, $|b_A - b_B| \leq |b_A^* - b_B^*|$, with at least one of the inequalities strict, then the difference of profits between the integrated and the segregated form is larger under $(\lambda, p_0, p_1, p_2, b_A, b_B)$."

The proof involves the same type of differentiation of (more complicated) functions as the proof of Proposition 5.

3.3 Hierarchy

We now consider a hierarchical organization where the two services use different codes but exploit the synergy by employing a fifth agent who provides translation. When inter-service communication is needed, the translator steps in. For instance, if salesman A has two customers, he communicates to the translator the type of the "extra" customer in the code used in service A. The translator will search for x, and then he will transmit the information to engineer B in the code used in service B.

Hiring a translator requires incurring a fixed cost μ, but since the translator is specialized in language, we assume that his diagnosis cost is lower than that of the engineers. In particular, we assume here that this cost is zero. The qualitative results of the analysis are valid if it is strictly positive, as long as it is lower than the engineers'.

The following proposition describes the variation of the optimal organization as a function of λ. The constraint on p ensures that the integrated organization is optimal for some λ.[16]

Proposition 6. *For any b, if p is high enough, there exists an interval (μ^{**}, μ^{*}) such for $\mu \in (\mu^{**}, \mu^{*})$, there exist $1 \leq \lambda_{\min} < \lambda_{\max} \leq 1/D^{*}(0)$ such that the unique optimal organization is*

$$\begin{array}{ll} \textit{integrated} & \textit{if } \lambda < \lambda_{\min} \\ \textit{hierarchical} & \textit{if } \lambda \in (\lambda_{\min}, \lambda_{\max}) \\ \textit{separated} & \textit{if } \lambda > \lambda_{\max} \end{array}$$

Proof. See appendix A. □

Figure 3 shows how the optimal organizational form varies as a function of the synergy parameter p and the diagnosis cost parameter λ. Let us first consider the choice between translation (*i.e.*, hierarchy) and separation. Translation incurs a fixed cost μ and increased diagnosis costs, but makes inter-service communication possible and thus allows the services to profit from the existing synergies. If the diagnosis cost λ is low, the extra cost due to translation is low and the net benefit is likely to be high. Thus, translation is more likely to be preferred to separation when λ is low.

Translation allows the two services to keep efficient service-specific codes – thus translation is likely to be preferred to integration when well adjusted codes are more important, that is when λ is large. Thus, as stated by the proposition, if the fixed cost μ of hiring a translator is low enough, for λ large enough the hierarchical structure is preferred to integration.

[16] Proposition 6 holds without change for the extended version of the model, except for the fact that it begins with "For any b_A and b_B, if p_0p_2 is high enough…".

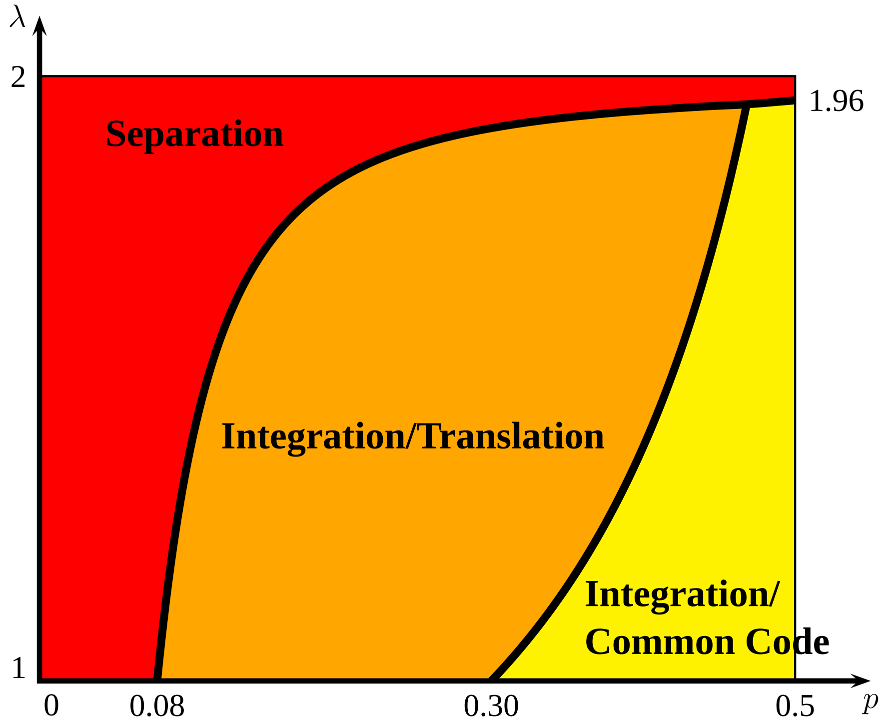

Figure 3: The choice between separation and integration with or without a hierarchy as a function of p and λ, when $b = 0.4$ and $\mu = 0.06$.

3.4 Implications and Evidence

3.4.1 Code Variety and Communication under Bounded Rationality: A Tragedy in the Desert

On April 14, 1994, two US Airforce F-15 fighters shot down two US Army Blackhawk helicopters over the Iraq no-flight zone, killing everyone inside. As documented in the excellent book-length account of the circumstances of the tragedy by Scott Snook (2000), the tragedy took place in broad daylight and involved highly experienced and qualified crews on both sides, who, furthermore, were monitored by a US Airforce AWACS (airborne warning and control system) flying above them.[17] In thorough subsequent investigations, no individual was found guilty; the tragedy was the result of grave organizational dysfunctions. This accident makes concrete some of the abstractions described above, common code, communication costs. It illustrates the role and consequences of code commonality in large organizations, and thus the role played by codes in limiting organizational scope.

The cultures of the different services of the US armed forces, understood in the

[17]The few paragraphs that follow cannot make justice to the insightful and complex causal chain that Snook establishes, and to the subtle discussion of the notion of causality in social sciences. The reader is encouraged to read this fascinating book.

broad sense of shared specific human capital (see Crémer, 1993), differ strongly. Whenever all services must come together in a military operation, a common plan, establishing the common set of rules and common code, is established. Operations Plan 91-7 described the rules governing Operation Provide Comfort (named this way since the no-fly zones were established to protect the Kurds in the North of Iraq and Shiites in the South). While this plan was quite detailed, it did not succeed in completely shaping the actual practice of the service's interactions. In particular, communication was hindered by a wide range of misunderstandings resulting from the incompatibility between the 'code' that governed Army operations with the code used by the Airforce. We illustrate this pattern with three key instances. First, the word 'aircraft' in the order forbidding the entry by aircraft into the No-Flight zone before the enemy radars were 'sanitized' by F-15s was understood by the Air Force to include helicopters, but by Army pilots to exclude helicopters (Snook, 2000:163). As a consequence, the Air Force pilots did not expect any American helicopter to be present in the no-fly zone, while the Army pilots did not think they were breaching the order. Furthermore, the AWACS crew thought is was responsible for planes, but not helicopters (2000: 163). Second, the two key acronyms concerning the no-fly zones, AOR (Area of Responsibility) and TAOR (Tactical Area of Responsibility) were understood differently by Army and Air Force: 'To the Army, AOR meant the area outside northern Iraq; to the Air-Force, it meant just the opposite.'(Snook, p. 2000:157) These different codes translated into crucially different interpretations concerning both the 'where' could aircraft be and the 'when' they could be there, and contributed to the misperception of the two Blackhawks as enemy helicopters by the fighter pilots. Third, the Air Force and the Army helicopters interpreted differently, for a long period of time up to the accident, the rules governing the electronic exchanges used to identify other aircraft as friendly or foe (the so-called IFF system, for 'Identify Friend or Foe'). As a result, the Army helicopters answered the Airforce pilots electronic query with the wrong code, and were identified as enemies (Snook, p. 2000:157). The Air Force pilots saw US helicopters where (they thought) they should not be, when they were not expected to be, using a wrong frequency of IFF, and shot them down as enemies.

Even this skeletal description of the accidents make two points clear. First, codes matter; absence of common codes greatly affects communication. Second, it is costly to impose a common code among a wide range of different services, given the limits on individual rationality. Taken together, these statements imply that code commonality places a bound on organizational scope.

3.4.2 Information Costs and Code Commonality: Evidence

In the past two decades, rapid advances in information and communication technology have caused a dramatic drop in the costs of information processing costs which, in the context of our model, can be interpreted as a reduction of the diagnosis cost parameter λ. Indeed, retrieving information has become cheaper. This is true for information contained in in-house databases and of the generally available information which can often be found on the World Wide Web. It is also the case that obtaining information from other individuals has become cheaper, both because of new means of communications, e-mail or SMS, and because older methods, such as telephone calls, have become less expensive. Making sense of a given message received from a third party, at least when this requires extra information, has therefore become less expensive.

Our theory predicts that these changes should imply an increase in 'integration' in the form of links across and within firms, through the use of hierarchies and common codes; moreover, within already integrated units, decreases in diagnosis costs reduce the 'translation' role of hierarchy, by facilitating 'horizontal' communication – the substitution of codes for hierarchies. In the rest of this section, we discuss some informal empirical evidence that supports the theory.

First, the reduction in information costs is correlated with increasing code commonality. Historically, the information generated by each business unit within a firm and by each function within each business unit has been coded and processed separately, according to the needs of that business unit or function; the different pieces of information were often defined in different ways and could not be easily aggregated.[18] As information costs have dropped, companies have sought ways to integrate this disperse information. In particular, this integration was obtained, between and within firms, through tools such as Enterprise Resource Planning (ERP) systems[19] and, earlier, Electronic Data Exchanges (EDI),[20]. This programs allow for the exchange of electronic data by standardizing the format of the data exchanged. Through these systems, firms have substituted flexible ways to code

[18]For example the database company Oracle had 70 incompatible databases for its human-resources department. This made it impossible to answer simple queries, such as how many employees were working at any time at the company. "If anyone wanted to find out the exact number of Oracle employees, it would take weeks of searching— and by the time the answer was found, it would already be out of date." ("Timely Technology," *The Economist*, January 31, 2002.)

[19]See for instance the products offered by SAP `http://sap.com` or Baan `http://www.baan.com`.

[20]We refer to EDI systems broadly, to include other related approaches such as CPFR ("Collaborative Planning, Forecasting and Replenishment") which involves deeper and more extensive electronic information sharing and has been installed, for example, by Nabisco and used with Webmans' Food markets ("Enterprise System," Financial Times, February 22, 1999); or web-based integrated value chains, such as the one introduced by Safeway in the UK ("You'll Never Walk Alone," The Economist, June 24, 1999).

their data for more rigid but unified central databases.[21]

Second, the reduction in information costs induces greater decentralization. Brynjolfsson and Hitt (2000) were the first to find evidence of this complementarity between IT and decentralization. Bresnahan, Brynjolfsson and Hitt (2002) find, using firm-level data, that greater use of information technology is associated with broader job responsibilities for line workers, and more decentralized decision-making. Caroli and Van Reenen (2001) also find, on entirely different data, evidence that the degree of decentralization of authority is complementary with the use of IT. Rajan and Wulf (2003), in a panel study of the hierarchical structure of firms, find that the span of control of the CEO is increasing over time, in particular, through the disappearance of the role of the COO. With more employees under his direct authority, the CEO can exert less control: decision making is more decentralized

Thus, the evidence does suggest that the drop in information costs led to (1) increasing commonality of codes in organizations and (2) increasing decentralization at the expense of hierarchy. This is not sufficient to show that these changes are causally linked in the same way as our model describes, that is, that the introduction of a common code allows for the substitution of the 'translation' role of hierarchy by direct horizontal communication between business units that otherwise would be 'speaking a different language.' Case study evidence, however, supports our theory.

Robert J. Herbold[22], Chief Operating Officer of Microsoft at the time, explains that in 1994 Microsoft had a completely decentralized set of information systems (Herbold (2000)). Each business unit used a different mapping of data to category: in the terminology of this paper, they all used different, specific, codes. The managers of the different units had chosen their own techniques of financial reporting, adapted to their own circumstances. In Herbold's words:

> "Some would develop financial information systems tailored to their particular needs. Others would analyze their financial performance in a way meant to reflect the environment of their country of operation. There was nothing seditious about this."

Similarly, there was no way to have a coherent overall image of human resources throughout the firm, with eighteen HR-related databases.

[21] In the words of a 'noted American e-commerce expert' cited by The Economist, ERP systems have replaced "fragmented unit silos with more integrated, but nonetheless restrictive enterprise silos"("Timely Technology," *The Economist*, January 31st, 2002).

[22] We rely heavily on Herbold personal account, in his *Harvard Business Review* article. All the quotes below are from his account.

> “When asked about head counts, managers answers usually were, to put it charitably, poetic.”

The tailoring of the information to the specific needs of the different business units compromised communications between them, as different measures needed to be reconciled.

Taking advantage of the drop in information costs, Microsoft introduced ‘common codes’ in these two areas; now, according to Herbold, all managers could easily make sense of the information produced by any business unit. Paradoxically, and as our model predicts, this centralizing move provided “benefits usually associated with decentralization” as managers had easy access to relevant information and could use it directly.

Even though the adoption of a common code appears to have been beneficial to Microsoft, the German Country Manager refused initially to go along with the common code, least his unit lost the unique fit of its own code to the German problems. In the words of that country manager:

> “We put years into the development of our own information systems because those systems uniquely capture the nuances of the German Business. Those nuances are important.”[23]

Even if adoption of a common code is in the interest of the company it may not be in the interest of all the agents who are involved. Within a firm, the “center” can presumably impose a common code on the different business units. In some cases, agents must decide independently whether to move to a common code. In the next section, we study conflicts of interest in the choice of organizational codes. They are particularly important when separate firms, necessarily involving separate decision makers, are involved.[24]

4 Strategic Code Adoption

So far, we have assumed that all agents share the goal of maximizing social surplus. In this section, we study the adoption of codes in the presence of conflicts of interest.

[23]Obviously, these complaints only show that the center thought the codes were inefficiently different while the country managers thought that the codes were just appropriately adapted to their different environments. On the other hand, the center presumably cares both about coordination between countries and the profits within each country, whereas the country managers care mostly about local conditions. There is therefore at least some presumption that the center’s objective function is better aligned with the interests of the firm as a whole.

[24]This is not to say such concerns are non-existent within firms. Herbold points out that a previous similar effort in Procter and Gamble failed when the CEO refused to overrule a recalcitrant division manager who wanted to preserve the previous, non-integrated, systems.

With complete contracts, agents with diverging objectives could agree to select the surplus-maximizing code and, if necessary, make appropriate side payments. However, it is reasonable to assume that code adoption is non-contractible, as firms cannot sign contracts that commit them to adopting a particular code, for several reasons. Outsiders cannot verify the actual code used in internal communications unless they are given, at prohibitive cost, full access to the firm: even if a particular organization formally adopts a code that its members may not actually use it. As a matter of fact, a common organizational dysfunction is for agents to create shortcuts, codes that are suitable for their own, private, communications, different from the official codes, with the result that miscommunications along the chain of command are common.

In this section, we first examine sequential code adoption: two firms must in turn choose a code, and we study the incentives for the firm that chooses first. Knowing that its decision affects the decision of the other firm, what code will it choose? We then analyze the consequences of free-riding for code adoption. We start from a situation in which firms have different codes but can adopt a common code, at some (fixed) cost. The presence of externalities may inhibit the adoption of an efficient common code.

4.1 First-Mover Bias

As in section 3, there are two services, A and B with the distribution functions $F(x;b)$ and $F(x;-b)$. However, these two services are now two separate firms: salesman A and engineer A belong to firm A while salesman B and engineer B belong to firm B. When salesman i has one customer, he communicates only with his engineer. When he has two customers, he will offer the second to engineer j, who accepts if he has not received a customer from her salesman — the profit thus generated is allocated in proportion σ to the salesman's firm and $1-\sigma$ to the engineer's firm.[25]

Timing is sequential. First, firm A adopts a code. After observing this code, firm B chooses its own, which can be either the same or different. Once the codes have been chosen, customers arrive and the services behave as in the previous section.

Because $\lambda > 1$, with separate codes, customers can only be served by the firm they first approach, and each firm has profits

$$\pi_i^S(p,b) = (1-p)(1-\lambda D^*(b)),$$

as in section 3.

[25]This is equivalent to assuming that the salesman (the engineer) makes a take-it-or-leave-it offer with probability σ $(1-\sigma)$.

With the same code, firms can "trade" customers. Suppose firms adopt a common code with some $x \in (0,1)$. The expected diagnosis costs for engineer A when he receives a client from salesman A is $s(x)$ (see equation 7); when he receives a client from salesman B it is

$$\begin{aligned}\tilde{s}(x) &= xF(x;-b) + (1-x)(1-F(x;-b)) \\ &= (1-x)F(x;b) + x(1-F(x;b)) = 1 - s(x)\end{aligned}$$

It is also easy to see that the expected diagnosis costs for engineer B who receives a client from salesman A or salesman B are also respectively $s(x)$ and $\tilde{s}(x)$. The profit of firm A is

$$\begin{cases} 1-\lambda s(x) & \text{with probability } 1-2p+p(1-p) \\ (1-\lambda s(x)) + \sigma(1-\lambda s(x)) & \text{with probability } p^2, \\ (1-\sigma)(1-\lambda\tilde{s}(x)) & \text{with probability } p^2, \\ 0 & \text{with probability } p(1-p). \end{cases}$$

The first line corresponds to all these cases where firm A serves one of its own customers and does not trade with the other firm: this happens when it has exactly one customer, or when it has two customers and the other firm has at least one. The second line corresponds to the case where firm A has two clients, one that it services and one that is serviced by the other firm who has no customer. The third line corresponds to the case where firm A serves a client from the other firm, while the fourth line corresponds to the case where it serves no clients.

Similarly, the profit of firm B is

$$\begin{cases} 1-\lambda\tilde{s}(x) & \text{with probability } 1-2p+p(1-p) \\ (1-\lambda\tilde{s}(x)) + \sigma(1-\lambda\tilde{s}(x)) & \text{with probability } p^2, \\ (1-\sigma)(1-\lambda s(x)) & \text{with probability } p^2, \\ 0 & \text{with probability } p(1-p). \end{cases}$$

From the expressions above we can compute the expected profits of the firms under a common code with x:

$$\begin{aligned}\pi_A^C(p,b,\sigma \mid x) &= 1-p+p^2-\lambda\left(\left(1-p+\sigma p^2\right)s(x) + (1-\sigma)p^2\tilde{s}(x)\right); \\ \pi_B^C(p,b,\sigma \mid x) &= 1-p+p^2-\lambda\left(\left(1-p+\sigma p^2\right)\tilde{s}(x) + (1-\sigma)p^2 s(x)\right).\end{aligned}$$

From the viewpoint of firm A, the best common code that will be accepted by B is the solution of

$$\max_x \quad \pi_A^C(p,b,\sigma \mid x) \tag{13}$$

$$\text{subject to } \pi_B^C(p,b,\sigma \mid x) \geq \pi^S(p,b). \tag{14}$$

The term in x in π_A^C is equal to a negative number multiplied by

$$(1-p-\sigma p^2)s(x)+(1-\sigma)p^2(1-s(x)) \\ =(1-\sigma)p^2+[1-p+(1-2\sigma)p^2]s(x). \quad (15)$$

Because $p \leq 1/2$,

$$1-p+(1-2\sigma)p^2 \geq 1-p-p^2 \geq 0,$$

therefore the coefficient of $s(x)$ in (15) is positive. The reasoning of footnote 11 can be applied to show that the profit of the first mover is single peaked in x.

Similarly, π_B^C is single peaked, and therefore the set of xs that satisfy the constraint in problem 14 is an interval. Firm A will choose in this interval the x which is the closest to its favorite code.

This proves the following proposition.[26]

Proposition 7. *If there exists a joint code such that adopting this joint code is Pareto-superior to using separate codes (in the sense that the profits of both firms are greater), in equilibrium a joint code is adopted. However, this code is biased to the benefit of the first-mover: it will be the code closest to its preferred codes among all Pareto-superior codes.*

Despite contractual incompleteness, firms do adopt a common code if and only if it is efficient to do so. However, when a common code is adopted, it is skewed towards the needs of the first mover.

The firm that chooses first takes only into account its expected profit. This includes the cost of its internal communications and part of the cost of inter-firm communication; but it does not take into account the cost of internal communication for the other firm. The first mover minimizes its communication cost by selecting a code that fits the distribution of characteristics of its customers, whereas the efficient code fits the "average" distribution of characteristics of the customers of both firms. The 'selfishness' of the first-mover is limited only by the

[26] The same result holds true in the extended version of the model defined in Remark 1. We have to distinguish between the functions s_A and s_B, and the functions $\tilde{s}_A$ and $\tilde{s}_B$. We have $\tilde{s}_i = 1 - s_i$.

The term in x in the profit of firm i is equal to a negative number multiplied by

$$(p_1+p_2)s_i(x)+p_0p_2(1-\sigma)\tilde{s}_i(x)-\sigma p_2p_0(\tilde{s}_j(x)) \\ =(1-p)(p_0p_2)+((p_1+p_2)-p_0p_2(1-\sigma))s_i(x)+\sigma p_2p_0s_j(x).$$

Because $p_0p_2(1-\sigma)<p_2$, the coefficient of $s_i(x)$ is positive and this function is the sum of two functions which satisfies the conditions described in footnote 11. It therefore also satisfies these conditions. This proves the single peakedness of the profit functions, and the result.

participation constraint of the follower: given that a common code is efficient, the first mover must make sure that the follower has sufficient incentive to adopt the common code.

4.2 Excessive code variety

With sequential adoption, a common code is adopted whenever it is efficient, however the code is biased in favor of the first mover. On the other hand, if changing codes induce switching costs, a common code could not be adopted, even if it could increase aggregate profits.

We will suppose that we start from a situation where firms have, for some reason, each chosen their optimal separate codes, and assume that switching to a different code induces a cost c. The firms become aware of each other, and it would be more efficient, in the sense of increasing aggregate profits, for one firm to switch to the code of the other. Because switching codes is non-contractible, and because it brings positive externalities to the other firm, there exist circumstances in which this switching does not take place.

To choose the case the most favorable to switching, we assume that the two firms are symmetric. Let π^S be the profit of any of the firms if it conserves its optimal code (S stands for '*s*eparate'), and π^J be the profit of each firm if they use the optimal common code (J stands for '*j*oint'). Changing to the common code increases aggregate profits if $2\pi^J - 2c > 2\pi^S$; if this inequality holds, both firms (as $\pi^J - c > \pi^S$) will have incentives to change code and there exists an equilibrium in which they do.

Let us now turn to the case where aggregate profits would increase if one of the firm adopted the code of the other, but not when they both switched to the optimal joint code. We call π_A^A be the profits of firm A and π_B^A the profits of firm B after firm B has adopted the code of firm A;[27] this is the case when

$$2\pi^S > 2\pi^J - 2c \tag{16}$$

and

$$2\pi^S < \pi_A^A + \pi_B^A - c. \tag{17}$$

As Proposition 8 states, these two inequalities are compatible with

$$\pi_B^A - c < \pi^S; \tag{18}$$

under some circumstances: in equilibrium no firm would change code, even though aggregate profits would be increased if they did.

[27] Because of the symmetry of the model, the analysis would be the same if firm A adopted the code of firm B.

Proposition 8. *In the symmetric version of our model, in the presence of switching costs, for λ small enough there exists an interval $[c^*, c^{**}]$ such that for any c in this interval, at equilibrium the two firms will keep separate codes when aggregate profits would be increased if one firm switched to the code of the other.*

Proof. Equations (16) and (17) are equivalent to

$$\pi^J - \pi^S < c < \pi_A^A + \pi_B^A - 2\pi^S.$$

If firm B adopts the code of firm A, firm A extracts the same profits from its clients that it serves than under separate codes, and furthermore extracts some profits from trading clients with firm B; therefore $\pi_A^A > \pi_S$ and

$$\pi_A^A + \pi_B^A - 2\pi^S > \pi_B^A - \pi_S.$$

Equations (16), (17) and (18) are therefore equivalent to

$$\max\{\pi^J - \pi^S, \pi_B^A - \pi_S\} < c < \pi_A^A + \pi_B^A - 2\pi^S,$$

or, because[28] $\pi_B^A < \pi^J$,

$$\pi^J - \pi^S < c < \pi_A^A + \pi_B^A - 2\pi^S. \tag{19}$$

The aggregate profits under the optimal joint code are greater than under any other joint code only because of lower diagnostic costs. Hence, when λ becomes small, the difference between $\pi_A^A + \pi_B^A - 2\pi^S$ and $2 \times (\pi^J - \pi^S)$ goes to zero, which implies that the right hand side of (19) becomes strictly greater than the left hand side; letting c^* be equal to the left hand side and c^{**} to the right and side, we have proved the result. $\square$

4.3 Evidence on Strategic Code Adoption: The design of the B-2 Bomber

The adoption of a common code for the design of the B-2 bomber by four independent firms provides some evidence of the 'strategic' aspects of the adoption process discussed in the previous two subsections. It also provides further evidence on the relationship between technology, code adoption and decentralization.

[28]We have $\pi_A^A + \pi_B^A < 2\pi^J$, and $\pi_A^A > \pi_B^A$, as when firm B adopts the code of firm A, firm A uses its optimal code for a greater proportion of the clients its serves than firm B.

The development of the B-2 bomber began in 1981.[29] Advances in information technology made it possible for Northrop, Boeing, Vaught (a division of LTV) and General Electric, the four companies in charge of the design, to create a centralized database to facilitate the design of the bomber. A key element of the construction of that database was the 'B-2 Product Definition System', essentially a common code, a "technical 'grammar' by which engineers and others conveyed information to each other."[30] In this case, the adoption of a common code was largely contractible, as it was embedded in software, although its use was still individually, and not jointly, determined. As a result, the strategic considerations analyzed above still played a role in the complicated process of adoption of a common code.

First, Boeing and Vaught were unenthusiastic about the adoption of a common code; as a Boeing engineer explained: 'we were developing our own system CATIA [...] We knew we wouldn't be using CATIA if we had to be compatible with this huge, monolithic database' (Argyres, 1999:166). In order to overcome this resistance, the Air Force accepted to subsidize the training costs incurred by Boeing and Vaught (Argyres, 1999: 166), presumably because a centralized code was efficient. Both the contractibility and the presence of a central player with ability to coerce the parties avoided the 'excessive code variety' of Proposition 8. Second, in the spirit of Proposition 7 , the optimal code was biased towards the needs of one of the parties, Northrop, whose system was adopted by all the parties (Argyres, 1999: 167).

The development of the B-2 was the 'first major aerospace program to rely on a single engineering database to coordinate the activities of the major subcontractors on a large-scale design and development project' (Argyres, 1999:163). The use of this database had two consequences. First, designers from different companies could participate jointly in the design. In previous projects, the difficulty of cross-company communication had meant all designers, with the exception of those of the motors (which are a relatively stand-alone component requiring little coordination) had belonged to the same firm.[31] Thus, the existence of a common code allowed integration of several teams were before there was none possible, an effect illuminated by section 3.2. Moreover, this integration reduced the hi-

[29]The account that follows draws heavily on a detailed cased study by Argyres (1999). For background information on the B-2 and links to other sources of information, see the site of the Federation of American Scientists: `http://www.fas.org/nuke/guide/usa/bomber/b-2.htm`.

[30]"This grammar was established through a highly-developed and highly standardized data formation and modeling procedures of the system, which laid down well-defined rules for communicating complex information inherent in the part design" (Argyres, 1999: 171). These rules included tight definition of 14 part families and "agreed upon modeling rules for defining lines, arcs, surfaces etc." (Argyres 1999:169).

[31]Argyres, personal communication to the authors.

erarchical coordination, since among the main consequences of the creation of a relatively rigid, unifying codes was an increase in decentralized decision making: "the technical grammar defined by the B-2 systems established a social convention which limited the need for a single hierarchical authority."(Argyres 1999: 173). This is consistent with our description of the substitution of hierarchies by codes in 3.3.

5 Related literature and conclusions

There are substantial and growing literatures on codes and on the consequences for organizational forms of the limits imposed by bounded rationality on the communication capabilities of agents. We begin this concluding section by reviewing these literatures and explaining their relationships to this paper, and end by a discussion of some directions for future research.

First, an area outside of economics, information theory (Shannon, 1948) studies optimal codes. However, because the questions posed are very different, so is the analysis. In particular, information theory is concerned with issues such as: representing the messages with sequences of binary digits (bits) that are as short as possible, what is the mean length of a message?; what is the capacity of different types of channels? what is the error in decoding, and what does it depend on?; can one design a code that allows as close to perfect communication as possible even in a noisy channel? In general, the theory assumes that the sender must transmit all of the information, and chooses the code that minimizes transmission cost. In our setting, which is concerned with the organizational implications of agents' bounded rationality, the transmission cost is given, but the sender is prevented from transmitting all the information. The optimal code maximizes the value of information transmitted. Moreover, we also study code adoptions by agents who can independently consider their own individual costs and benefits in adopting a code – this is obviously also out of the realm of information theory.

The idea that there is a trade-off between generality and specialization of codes, explored for instance in sections 2 and 3 was already informally explored in Arrow's celebrated *The limits of organization* (1974), where, after discussing the endogenous development of codes within organizations, he identifies the trade-off between general codes that allow for wide communication and specialized codes tailored to the needs of particular organizations.

Other antecedents of our work include Crémer (1993), who presents a bounded rationality analysis of corporate culture. He argues that 'corporate culture is the stock of knowledge shared by members of the corporation, but not by the general population from which they are drawn', and suggests that this knowledge stock is formed by three pieces: a shared knowledge of facts, a common code, and a shared knowledge of rules of behavior. He then goes on to study, within a team theoretic framework, the benefits of shared knowledge.[32] In the same context, Prat (2002) explores the connection between the optimal extent of information homogeneity within a team and the kind of complementarities that exist among different team members. However, neither of these two papers examine the design of codes or their consequences for the design of organizations. Battigalli and Maggi (2002) construct a sophisticated model of language, which they then use to develop a theory of contract incompleteness. Their language is a code with the purpose of legal verification which is built by combining primitive sentences and logical connectives (AND, OR, NOT, etc...). A contract uses the available language to partition the set of events and associate it to the parties' obligations. Like Battigalli and Maggi's we take into explicit account the cost of using language to partition the set of events. However, our focus on organizations is radically different from their focus on contracts.[33] Like us, Wernerfelt (2003) considers codes that minimize communication costs within an organization, but with a different focus. In his model, agents have common interests but decisions are decentralized, and he studies the existence of symmetric or asymmetric equilibria with one or multiple codes. His analysis does illuminate the existence of equilibria with different codes when a common codes would be optimal, but does not extend to organizational implications or to the strategic aspects of the choice of codes. Finally, Dewatripont and Tirole (2003)'s study of communication emphasizes the strategic interactions between the communication efforts made by different agents, not the language they use.

Building on Marschak and Radner's (1972) *team theory* a number of authors have studied how the limits on communications imposed by affect organizational structure. Crémer (1980) studies the optimal allocation of tasks into divisions, whereas other authors have been more interested in developing a theory of hierarchies. Radner (1993) and others (see Van Zandt, 1999 for a survey) stress the limited computation capacity of agents. Closer to our work is Bolton and Dewatripont (1994), who consider a more general communication cost structure; this leads to a theory which builds on the trade-off between communication costs and returns to specialization. In Garicano (2000), the bounded rationality of agents prevents them from learning the solution to all the problems potentially faced by the organization, but they can request help from other agents when they do not know how to solve one of them. He shows that the firm will organize itself in a

[32] Chowdhry and Garmaise (2004) build on this work by studying organizational capital and corporate culture in a dynamic model of learning in the firm and analyze the consequences of such capital for financial magnitudes.

[33] Similarly in their focus on contracting and not in organization, Chatterji and Filipovich (2004) describe a language as a partition of the set of possible histories of the game generated by contracts and study the effect of language ambiguity on judicial interpretation.

knowledge hierarchy, in which agents closer to the production floor deal with the most common problems while higher rank agents deal with less frequent problems.

Thus, to the best of our knowledge, none of the previous literature studies the relationship between the organizational code and the organizational choices of the firm. Focusing on this relationship has allowed us to build a theory that tightly links an important component of bounded rationality to the theory of hierarchies, to the span of control of managers, to the strategic advantage that first movers have in the design of projects. Furthermore, we have obtained some testable hypotheses from the model that seem in accordance with the evidence uncovered by economists concerning decentralization and information technology; we have shown that the causal mechanism we propose is consistent with the evidence in some detailed case studies of decentralization and organization.

Specifically, our analysis illuminates at least three sets of issues absent from the literature. First, and at an immediately applied level, it provides an explicit link between centralized communication and code commonality. In particular, we have shown that codes and hierarchy are likely to be substitutes; that is, for a given firm scope, increasing code commonality will reduce the reliance on hierarchies. Second, more generally, we have provided a simple way to analyze and formalize an elusive idea, the idea of a code, and shown how this formalization could be used to study organization issues. Third, we have characterized some of the main strategic issues that are likely to bias code adoption.

Also, understanding the loss in precision imposed by the need for common codes offers a fresh, bounded-rationality based, perspective on organizational boundaries. Modern theories of the firm scope have emphasized the role played by different allocations of asset ownership in providing incentives to agents (e.g. Hart and Moore 1986). Our theory suggest that, if allowing agents to efficiently communicate with one another requires establishing a common code, then agents' bounded rationality limits the scope of the firm.[34] By including broader, more diverse business inside a common code, communication within each service deteriorates. Thus, the organization should group related services or products and segregate unrelated ones, since the latter cause larger decreases in the efficiency of communication with smaller gains in synergies.[35]

[34]In fact, the original Coase (1937) paper on the firm's boundaries emphasized the importance of organization costs in setting a limit on firm size, and argued that a reason these organization costs exist is the limiteds on the manager's resources. Building on that insight, Oliver Williamson (1967) starts a tradition, which we discuss in the conclusions, of modeling the source and effects of managerial bounded rationality on firm scope. That literature, however, does not allow to consider the possibility of transacting 'outside' the organization and limits its attention to inside the hierarchy interactions.

[35]This logic provides one way to understand discussions in the business press of the need to

Our analysis suggests several interesting avenues for new research. First, our model yields testable hypotheses, and the availability of large databases of business texts and their ease of access may allow for a study of the commonality of the language used across different services of different firms or across different firms in an industry. Beyond testing the relation between integration of codes and characteristics of the environment, such research would allow for a direct test our hypothesis on the 'centralized information, decentralized decision-making'– the substitution of codes for hierarchies. In particular, one should observe more 'delayering' (less hierarchy) and more horizontal communication as codes become more common.

Second, on a theoretical level, our analysis can be used to provide some structure to the concept of 'hard' and 'soft' information, which is increasingly used in the contracting literature (see Aghion and Tirole, 1997; Stein, 2002). These concepts can be given a precise meaning in our model. A word within a code is hard information; it can easily be passed far down the chain of command or in space. The exact meaning of the word, that is clarifying which of the possible events within it are referred to, is soft information – that is, it is not very costly to eventually figure out this meaning in one to one communication, as one can continue talking until the other party understands, but is very costly to do so when communication is long distance or along a chain of command.

Third, also on a theoretical level, it would be interesting to explore code adoption in a dynamic setting. We conjecture that there exists a U-shaped relationship between the persistence of the environment in which the organization operates and the persistence of the code that the organization uses. Codes are stable over time if the environment is either very immobile (a specialized code needs not be modified) or it is highly unpredictable (a constant non-specialized code is the best solution).

A final promising research avenue concerns the interaction between organizational codes and labor market dynamics. A worker who learns an organizational code acquires organization-specific human capital. How portable is such capital between organizations? In turn, how does portability affect equilibrium wages and job turnover? Finally, how does the optimal code policy change once the organization realizes that the code it adopts affects the career prospects of its employees and, therefore, its hiring success? Anecdotal evidence suggests that organizations choose very different policies. Some, like Southwest Airlines, strive to imbue their employees with a strong corporate culture that set them apart from the rest of the industry. Other organizations (like university departments and research centers) have an incentive structure that puts a large premium on code portability (to publish, one must communicate with the rest of the profession not just with di-

'focus on the core business,' outsource the 'non-core' assets, etc.

rect colleagues). Still, others create a distinctive code even though they have high employee turnover (like McKinsey), possibly because they are exploiting a recognized first-mover advantage: their employees are highly valued on the market exactly because they have acquired that particular code.

References

[1] Aghion, Philippe and Jean Tirole. “Formal and Real Authority in Organizations” *Journal of Political Economy*, 105 (1): 1—29, 1999.

[2] Argyres, Nicholas S. “The Impact of Information Technology on Coordination: Evidence from the B-2 ”Stealth” Bomber.” *Organization Science*, 10 (2): 162—180, 1999.

[3] Arrow, Kenneth J. . *The Limits of Organization*. Norton, New York, 1974.

[4] Battigalli, Pierpaolo and Giovanni Maggi. “Rigidity, discretion, and the costs of writing contracts.” *American Economic Review* 92(4): 798–817, 2002.

[5] Blankevoort, P. J. “Contradictory effects of a project management dictionary.” *International Journal of Project Management*, 4 (4): 236—238, 1986.

[6] Bolton, Patrick and Mathias Dewatripont. “The firm as a communication network.” *Quarterly Journal of Economics*, 104(4): 809–839, 1994.

[7] Bresnahan Timothy F., Erik Brynjolfsson, and Lorin M. Hitt, “Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence.” *Quarterly Journal of Economics*, 117 (1): 339–376, 2002.

[8] Brynjolfsson, Eric, and Lorin Hitt. “Beyond Computation: Information Technology, Organizational Transformation and Business Performance.” *Journal of Economic Perspectives*, 14 (4): 23—48, 2000.

[9] Browning, Larry D., Janice M. Beyer, and Judy C. Shetler. “Building Cooperation in a Competitive Industry: SEMATECH and the Semiconductor Industry.” *The Academy of Management Journal*, 38(1): 113—151, 1995.

[10] Caroli, Eve and John Van Reenen, “Skill Biased Organizational Change? Evidence from a Panel of British and French Establishments.” *Quarterly Journal of Economics*, 116(4): 1449—1492, November 2001.

[11] Chatterji, Shurojit and Gragan Filipovich. “Incomplete Contracting due to Ambiguity: Natural Language and Judicial Interpretation.” mimeo, Colegio de Mexico, 2004.

[12] Chowdry, Bhagwan and Garmaise, Mark J. “Organization Capital and Intrafirm Communication” mimeo, UCLA, 2004.

[13] Crémer, Jacques. “Corporate Culture: Cognitive Aspects.” *Industrial and Corporate Change*, 3(2): 351—386, 1993.

[14] Crémer, Jacques. “A Partial Theory of the Optimal Organization of a Bureaucracy.” *Bell Journal of Economics*, 11(2), 683-693, 1980.

[15] Garicano, Luis. “Hierarchies and the Organization of Knowledge in Production.” *Journal of Political Economy*, 108(5): 874—904, 2000.

[16] Garicano, Luis and Esteban Rossi-Hansberg. “Organization and Inequality in a Knowledge Economy.” Mimeo, 2003.

[17] Hart, Paul and Carol Saunders, “Power and Trust: Critical Factors in the Adoption and Use of Electronic Data Interchange.” *Organization Science*, 8 (1), 23—42, 1997.

[18] Herbold, Robert J., “Inside Microsoft: Balancing Creativity and Discipline.” *Harvard Business Review*, January 1, 2002.

[19] Jakob Marschak and Roy Radner. *Economic Theory of Teams*. Yale University Press, New Haven, Connecticut, 1972.

[20] Prat, Andrea. “Should a Team Be Homogeneous?” *European Economic Review* (46)7: 1187–1207, 2002.

[21] Rajan, Raghuram G. and Julie Wulf. “The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies.”, Working Paper, Wharton School University of Pennsylvania, 2002.

[22] Roy Radner. “The organization of decentralized information processing.” *Econometrica* 61(5): 1109–1146, 1993.

[23] Claude E. Shannon. “A mathematical theory of communication.” *Bell System Technology Journal* 27: 379–423, 1948.

[24] Simester, Duncan and Knez, Marc. “Direct and Indirect Bargaining Costs and the Scope of the Firm.” *Journal of Business*, 75(2): 283-304, 2002.

[25] Stein, Jeremy. “Information Production and Capital Allocation: Decentralized vs. Hierarchical Firms” *Journal of Finance*, 62(5): 1891-1921, 2002.

[26] Timothy P. Van Zandt. “Decentralized information processing in the theory of organizations. ” In *Contemporary Economic Issues*, Vol. 4: Economic Design and Behavior (ed. Murat Sertel), MacMillan, London, 1999.

[27] Wernerfelt, Birger. “Organizational Languages”, mimeo. Available at `http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=26163`, 2003.

Appendix A
Proof of Proposition 6.

We present the proof for the extended version of the model described in Remark 1, because, although the two proofs are very similar, this might not be clear when the reader is faced with the proof of the simple case.

In a hierarchy, engineer A will receive with probability p_1+p_2 a client coming from salesman A and with probability p_0p_2 a client coming from salesman B. Therefore, the distribution of the characteristics x of the customers that engineer A serves is $F\left(x; b'_A\right)$,where

$$b'_A = \frac{p_1+p_2}{p_1+p_2+p_0p_2} b_A + \frac{p_0p_2}{p_1+p_2+p_0p_2} b_B.$$

The language which minimizes the diagnosis cost of engineer A is the same language that would be used in the case of one service working in isolation facing the distribution $F(x; b'_A)$, and the corresponding expected diagnosis cost per customer is $D^*(b'_A)$. Because he serves a customer with probability $p_1+p_2+p_0p_2$, the profits generated by engineer A are therefore

$$\left(p_1+p_2+p_0p_2\right)\left(1-\lambda D^*(b'_A)\right.$$

A similar reasoning holds for engineer B and the profits from the hierarchical firm are therefore

$$\Pi^T = 2(p_1+p_2+p_0p_2)(1-\lambda\frac{D^*(b'_A)+D^*(b'_B)}{2}) - \mu,$$

which must be compared to

$$\Pi^I = 2(p_1+p_2+p_0p_2)\left(1-\lambda D^*\left(\frac{b_A+b_B}{2}\right)\right), \tag{10}$$

and

$$\Pi^S = 2(p_1+p_2)\left(1-\lambda\frac{D^*(b_A)+D^*(b_B)}{2}\right). \tag{11}$$

Because $b'_A+b'_B = (b_A+b_B)/2$, the concavity of D^* implies

$$D^*\left(\frac{b_A+b_B}{2}\right) > \frac{D^*(b'_A)+D^*(b'_B)}{2}, \tag{A.1}$$

and furthermore, because b'_A and b'_B both belong to (b_A, b_B)

$$\frac{D^*(b'_A)+D^*(b'_B)}{2} > \frac{D^*(b_A)+D^*(b_B)}{2}. \tag{A.2}$$

Notice that $\Pi^I = \Pi^S$ for

$$\lambda^* = \frac{p_0 p_2}{(p_1 + p_2 + p_0 p_2) D^* \left(\frac{b_A + b_B}{2}\right) - (p_1 + p_2) \frac{D^*(b_A) + D^*(b_B)}{2}}.$$

We first show that under the hypotheses of the proposition, λ^* belongs to $\left[1, 2/D^* \left(\frac{b_A + b_B}{2}\right)\right]$. Notice first that, by concavity of D^*, the denominator of λ^* is smaller than $p_0 p_2 D^* ((b_A + b_B)/2)$, which establishes the upper bound.

To show that $\lambda^* > 1$, let $D^{\min} = D^*(1) = \left(44 - 7\sqrt{7}\right)/54$ be the minimum value of D^*. Because $D^*(b) \leq 1/2$ for all b, if

$$p_0 p_2 \geq 1 - 2D^{\min} > 0.06$$

(the $p_0 p_2$ large enough of the proposition), the denominator of λ^* is smaller than or equal to

$$\begin{aligned} \frac{p_1 + p_2 + p_0 p_2}{2} - (p_1 + p_2) D^{\min} &= (p_1 + p_2) \left(\frac{1}{2} - D^{\min}\right) + \frac{p_0 p_2}{2} \\ &< p_0 p_2 \end{aligned}$$

and λ^* is greater than 1.

We solve for μ the equation $\Pi^T = \Pi^I$ to obtain

$$\mu^* = 2\lambda(1 - p + p^2)(1/2 - D^*(b')).$$

Fixing $\mu = \mu^*$, the three functions of λ, Π^T, Π^I and Π^S are represented on figure A.1; they are equal for $\lambda = \lambda^*$. The relative slopes of the functions are consequences of the fact that

$$\begin{aligned} &(p_1 + p_2 + p_0 p_2) D^* \left(\frac{b_A + b_B}{2}\right) \\ &\qquad > (p_1 + p_2 + p_0 p_2) \frac{D^*(b'_A) + D^*(b'_B)}{2} \qquad \text{(by (A.1))} \\ &\qquad\qquad > (p_1 + p_2) \frac{D^*(b_A) + D^*(b_B)}{2}. \qquad \text{(by (A.2))}. \end{aligned}$$

A decrease in μ shifts the graph of Π^T upswards, and for μ smaller than μ^* but large enough that there exists values of λ such the three organizational forms can be optimal (this determines the μ^{**} of the proposition), the comparative statics of the proposition hold.

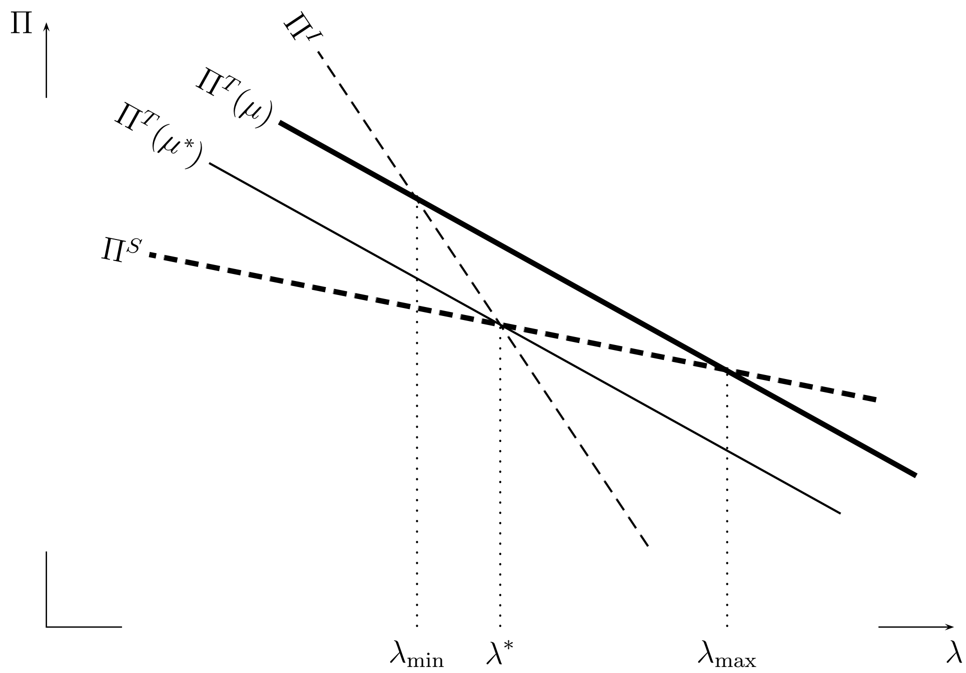

Figure A.1: This figure illustrates the proof in Appendix 6. The line marked $\Pi^T(\mu^*)$ is the graph of Π^T as a function of λ when $\mu = \mu^*$; the line marked $\Pi^T(\mu)$ shows the graph of Π^T with a smaller μ.

Appendix B

Extension of propositions 1 and 2 to the 'Natural Ordering' case

Suppose that there is a continuum of events with $X = [0,1]$. The frequency of events is described by a continuous and differentiable, but possibly non-monotonic, probability density f on $[0,1]$. Words are constrained to be intervals. Writing $t_0 = 0$ and $t_K = 1$ a code is therefore a partition[36] $\{[t_{k-1}, t_k]\}_{k=1,K}$.

The best K-words code is solution of

$$\min_t \sum_{k=1}^{K} (F(t_k) - F(t_{k-1}))(t_k - t_{k-1})$$

subject to

$$t_{k-1} \leq t_k \text{ for } k = 1, \ldots, K.$$

As in the text, the familiarity of a word, $[t_k, t_{k+1}]$, is the probability $F[t_{k+1}] - F[t_k]$ that the word is used; its breadth, $t_{k+1} - t_k$, is the 'number of events' in the word. Finally, the average frequency' of the events in the word is the average density of these events,

$$\phi_k = \frac{F(t_{k+1}) - F(t_k)}{t_{k+1} - t_k}.$$

Then the following proposition contains the results equivalent to propositions 1 and 2 for the case where events are naturally ordered.

Proposition B.1 (Natural order)**.** *When words must contain contiguous events, the following two properties hold in an optimal code:*

1. *For two contiguous words, the broader word is used less often .*
2. *For two contiguous words, the broader word describes events which have a lower average frequency.*

We begin by proving the following lemma.

Lemma B.1. *In the optimal code $t_{k-1} < t_k$ for all $k = 1, \ldots, K$.*

[36] As the text is written, t_k belongs to two words. To avoid this, words should be described by semi-open intervals, at the cost of heavier notation. It should be obvious to the reader that the results are not affected.

Proof. Assume for instance that we had $t_0 < t_1 = t_2 = t < t_3$. Increase t_2 by a small x. The diagnosis cost increases by λ multiplied by

$$(F(t+x) - F(t))x + (F(t_3) - F(t+x))(t_3 - t - x).$$

The derivative of this expression with respect to x for $x = 0$ is equal to

$$-f(t)(t_3 - t) - (F(t_3) - F(t)) < 0,$$

which proves the result. □

We can know prove the proposition.

Proof of proposition B.1. The first-order conditions are

$$F(t_k) - F(t_{k-1}) + f(t_k)(t_k - t_{k-1}) = F(t_{k+1}) - F(t_k) + f(t_k)(t_{k+1} - t_k),$$

which imply

$$f(t_k) = \frac{[F(t_{k+1}) - F(t_k)] - [F(t_k) - F(t_{k-1})]}{(t_k - t_{k-1}) - (t_{k+1} - t_k)} \tag{B.1}$$

The numerator is the difference between the familiarities of contiguous words, while the denominator is the opposite of the difference between their breadths. Thus, optimality requires that the differences between breadth and familiarity of contiguous words have opposite signs, as part 1 of the proposition states.

To prove the second statement, rewrite (B.1) as

$$f(t_k) = \frac{\phi_{k+1}(t_{k+1} - t_k) - \phi_k(t_k - t_{k-1})}{(t_k - t_{k-1}) - (t_{k+1} - t_k)}$$

Thus $\phi_{k+1} - \phi_k$ and $(t_{k+1} - t_k) - (t_k - t_{k-1})$ must be of opposite sign: that is, events in the broader word have a lower average frequency. □