Codes in Organizations*

Jacques Crémer
Université de Toulouse, IDEI-GREMAQ and CEPR

Luis Garicano
University of Chicago, MIT, and CEPR

Andrea Prat
London School of Economics and CEPR

April 14, 2004

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. We also we study how conflicts of interests among organizations can lead to the adoption of biased codes or to excessive code variety. Our analysis illuminates some aspects of the impact of IT innovation on organizational structure.

*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.

1 Introduction

A code is a collection of technical expressions that members of one or more organizations share with each other but not with the general population. A code can be a technical language, like the one shared by professional economists ("subgame perfect equilibrium" does not mean much to the rest of the world). Or it can be a convention for categorizing and sharing data, like an accounting system or an organizational database. Sometimes these codes emerge in an uncoordinated way over the course of decades, like in the case of the economic lingo. Other times they are created ex novo with a precise purpose in mind: For instance, firms often initiate a common project by creating a Project Management Dictionary (Blankevoort 1986).[1]

Despite the ubiquity of codes, there has been little formal analysis of their consequences for the organization of firms. The importance of organizational codes was recognized by Arrow (1974), who also provided an informal discussion of the properties of optimal codes. However, we are not aware of any formal theory of organizational codes in economics.[2] The goal of the present paper is to introduce a parsimonious model of codes. After characterizing the properties of optimal codes, we aim to study the interaction between the determination of the code and the adoption of an organizational form.

We will use our results to interpret recent changes in organizational structure. In particular, we will study how the management of 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." Starting from IT innovation, our theory of codes will offer a coherent account of these phenomena as equilibrium outcomes.

The first step of our analysis is to build a simple model of codes. Our agents are subject to two forms of bounded rationality. First, they have a limited ability to learn codes. Second, they have a limited ability to solve

[1] One such case is the SEMATECH consortium, in which all domestic US manufacturers of semiconductors and the US government came together in an effort to engineer the recovery of the US semiconductor industry. In order to bridge the differences between all the different company codes the consortimum decided to "compile a dictionary of common technical terms and acronyms. Before this attempt at standardization, many firms prided themselves on having unique names for things." (Browning, Meyer and Shetler, 1995: 125).

[2] The relation of the present paper with the existing literature is discussed in detail in the Conclusions.

problems that involve incomplete information. Suppose an agent is faced with a problem (think of the "problem" as a potential business opportunity), the solution of which depends on some event, which the agent does not observe. Another agent in the organization is familiar with the problem at hand and is willing to communicate what he knows to the first agent (suppose for now that there are no conflicts of interests among agents). Communication occurs through a previously agreed code. Each word of the code restricts the possible set of events to a specific subset. This simplifies, but does not eliminate, the problem-solving stage. Given a distribution of possible problems, each code determines a certain expected surplus, defined as the difference between benefit of solved problems and cost of problem solving.

In Section 2 we derive some of the properties of efficient codes. A code should use precise words for frequent events and vague words for more unusual ones. This observation leads to a simple algorithm for identifying the optimal code for any particular distribution of events. A more unequal distribution of events increases the value of the creation of a specialized code, since the precision of the words can be more tightly linked to the characteristics of the environment.

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 is costly as it limits the set of agents among whom the words are useful. As a consequence, we show that agents will either have 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 it shapes both their scope and their use of integrating mechanisms, which we study next.

If having agents communicate with each other requires that they share a common code, would an organization composed of multiple services want to use a common code for all of its parts? In Section 3, we argue that it would want to do so in order to improve coordination between services, a benefit that it must trade off with the resulting degradation of within-service communications due to the use of a less well-adapted code. We identify the variables that determine the terms of these tradeoff: a common code is more likely to be adopted if the degree of synergy among services is high, the difficulty of the underlying task is great, the distribution of events in different services is similar.

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. Hierarchies are more efficient when communication costs are high, whereas low communications costs favor their replacement by common codes and horizontal communications. We discuss the reorganization of Microsoft under Robert J. Herbold in the 1990s as an example of the trade-offs involved.

Finally, in Section 4 we study how the presence of conflicts shapes code adoption and organizational structure. Different organizations may sometimes choose to use the same code in order to facilitate cooperation but this choice is affected by strategic considerations. We show that there exists a first mover advantage: a shared code is suboptimally skewed towards the needs of early adopters. Moreover, when adoption is not contractible, there will exist too little code commonality, as investing in a common code generates externalities by improving other's communications. This implies that coordination within organizations will be better than coordination between organizations, which appears consistent with the evidence.[3] Some 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 the paper.

At the end of the paper, we discuss links with previous literature, and discuss some directions for future research.

2 A Theory of Codes

We begin by introducing a simple theory of endogenous codes. After laying out the model, we derive a series of properties of optimal codes. We discuss how the cost of miscommunication varies with the characteristics of the environment. Initially, we study the case of just two agents, but we then discuss what happens when the model is extended to more agents.

2.1 Model

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

The salesman and the engineer share the same code $\mathcal{C}$, which 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. Agents have a limited ability to learn languages. The code can contain at most K words.

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

A certain word has two characteristics. First, how many events it contains We call $n_k \equiv \sharp W_k$ the *breadth* of word k. Second, how likely it is to occur. We call $p_k \equiv \sum_{x \in W_k} f_x$the *frequency* of word k.

Once the engineer receives the word from the salesman, she still has to identify exactly what the client's characteristics x are. This *diagnosis* stage takes time and/or energy, which translates into a monetary cost d. The broader the word, the harder the search: we assume that the diagnosis cost is an increasing function of the number of events in the word: the function $d(n_k)$ is strictly increasing in the breath n_k. The expected diagnosis cost given a certain code $\mathcal{C}$ is

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

For now, we assume that 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. If all clients are served, profit maximization is equivalent to finding the code that minimizes the expected diagnosis cost.

To give a concrete interpretation of the diagnosis cost, consider the following simple example. The salesman is an interior decorator who is remodeling a client's house. The engineer is a craftsman who collaborates with the decorator (abstract from conflict of interests between the decorator and the craftsman). Let x be the client's preference about a particular fixture of the house. 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. The key assumption here is that the decorator cannot just relate the client's informal conversation because it would take too long or the craftsman would not understand what the client means.[4] Instead, the decorator and the craftsman have a common code to describe fixtures. Their bounded rationality makes such a code somewhat coarse. After receiving the approximate description of the desired characteristics of the fixture, the craftsman collects

[4] The craftsman could spend resources trying to improve his understanding of the original client's language, but this would come at a cost (regular reading of fancy architectural magazines? trips to Paris?). We assume here for simplicity that this cost is infinitely high (K is fixed). The results would change only qualitatitatively if we assumed that K can be increased at a cost.

Equally, the decorator could familiarize herself with the commercial supply of fixtures, so she could just ask the craftsman to bring a specific model. But that too would come at a cost.

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. If the decorator and the craftsman can agree on a code to describe fixtures, what would the optimal code look like?[5]

Another concrete example is supplied by knowledge management systems

2.2 Optimal codes

Given that a code is an allocation of a finite number of events in a finite number of words, finding an optimal code is an integer problem. We cannot offer a full characterization of the solution but we can derive two useful properties of optimal codes.

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

Proof. Suppose $\mathcal{C}$ is the optimal code. Take any two words, k and k', such that $n_k > n_{k'}$ (if all the words have the same length, the proposition is automatically true). Suppose for contradiction that there are events $x \in W_k$ and $x' \in W_{k'}$ such that $f_x > f_{x'}$.

Construct a new code $\tilde{\mathcal{C}}$ that is identical to the given code except that event x now belongs to word k' and event x' now belongs to word k.

The difference between the diagnosis costs of code $\mathcal{C}$ and code $\tilde{\mathcal{C}}$ is

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

But then $D(\mathcal{C}) > D\left(\tilde{\mathcal{C}}\right)$, which is a contradiction because we assumed that code $\mathcal{C}$ minimizes the expected diagnosis cost. ■

Proposition 1 is a direct consequence of the assumption that the diagnosis cost $d(W_k)$ is increasing in the breadth of word k. If we hold the breadth of each single word fixed, the best thing we can do to reduce expected diagnosis

[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.

time is to put the frequent events into narrow words and the rare ones into broad words.

Proposition 1 relates word breadth to event frequency. One may also wonder about word frequency. To prove a result in this area we need to make the additional assumption that the diagnosis cost is linear or convex:

Proposition 2 *Suppose that $d''(n) \geq 0$ for all n. Unless integer constraints make it impossible, in an optimal code broader words are used less frequently. Formally, if $n_k - n_{k'} \geq 2$, then $p_{k'} \geq p_k$.*

Proof. Suppose $\mathcal{C}$ is the cost-minimizing code. Take any two words, k and k', such that $n_k - n_{k'} \geq 2$ and suppose for contradiction that $p_{k'} < p_k$.

Construct a new code $\tilde{\mathcal{C}}$ that is identical to the given code except that an event $\tilde{x}$ that used to belong to word W_k is now moved to work $W_{k'}$. The difference between the diagnosis costs in $\mathcal{C}$ and $\tilde{\mathcal{C}}$ is

$$\begin{aligned} & D(\mathcal{C}) - D\left(\tilde{\mathcal{C}}\right) \\ = \; & d(n_k)\, p_k + d(n_{k'})\, p_{k'} - d(n_k - 1)(p_k - f_{\tilde{x}}) - d(n_{k'} + 1)(p_{k'} + f_{\tilde{x}}). \end{aligned}$$

As $n_k - n_{k'} \geq 2$ implies $n_k - 1 \geq n_{k'} + 1$,

$$\begin{aligned} & d(n_k - 1)(p_k - f_{\tilde{x}}) + d(n_{k'} + 1)(p_{k'} + f_{\tilde{x}}) \\ \leq \; & d(n_k)(p_k - f_{\tilde{x}}) + d(n_{k'})(p_{k'} + f_{\tilde{x}}) \\ < \; & d(n_k)\, p_k + d(n_{k'})\, p_{k'}, \end{aligned}$$

where the first inequality is because d is weakly convex and the second inequality is because d is increasing. But then $D(\mathcal{C}) > D\left(\tilde{\mathcal{C}}\right)$, which is a contradiction. Hence, it must be that $p_{k'} \geq p_k$. ■

To see that broader words are used less frequently: if $n_k \geq n_{k'}$, then $p_{k'} \geq p_k$, start from an optimal code and suppose we transfer an infinitesimal event x^* from a broad word to a narrower word. After transferring it, the event x^* is captured by a narrower word, and this is a certain benefit. Moreover, the broad word is now less broad and the narrow word is less narrow. If the broad word were used more often, this would also yield a benefit; but the original code was optimal, thus it cannot be a benefit, and the broad word must be used less frequently.

The intuition above is based on a marginal argument (which explains why convexity is needed). To complete the argument, we must account for the presence of integer constraints. There may exist words that are both broader and used (slightly) more than others because they only contain non-infinitesimal words which – if moved – would make another word both broader and more frequently used. The proposition puts an upper bound to the importance of integer constraints. If word k contains at least two events more than word $k^{'}$, then k must be used less often than $k^{'}$.

We have assumed that events could be allocated between words arbitrarily. In some instances, however, events have a natural order which imposes constraints on the ways in which words can be constructed. For example, if we are partitioning the color spectrum into discrete color words, we cannot create words that group non contiguous points of the spectrum. We show in the Appendix that, when events have a natural ordering and cannot be reorganized, propositions equivalent to 1 and 2 can be proven. In particular, it is the case that for two contiguous words, the broader word is used less often, and that words describes events which have a lower average frequency.

2.3 The value of a code

How does communication cost depend on the features of the underlying environment? This section shows that the cost goes down when the distribution of events becomes more "unequal".

To give a precise meaning to the inequality of distributions, assume without loss of generality that the events are indexed by natural numbers and that they are ordered in a non-decreasing fashion according to their frequency: i.e. $f_x \leq f_{x'}$ if $x < x'$. We say that another distribution of events $\tilde{f}$ (where the probabilities are not necessarily non-decreasing in the indices) is more unequal than f if we have

$$\sum_{x' \leq x'} f_x \geq \sum_{x' \leq x} \tilde{f}_{x'} \text{ for all } x.$$

Intuitively, a more unequal distribution puts even less probability on events that were already less likely to happen.[6]

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

Proposition 3 *If the distribution of events $\tilde{f}$ is more unequal than distribution f, the minimal diagnosis cost associated with $\tilde{f}$ is not greater than the minimal diagnosis cost associated with f.*

Proof. Let $\mathcal{C}$ be the optimal code for distribution f. Use the same code for distribution $\tilde{f}$. The cost associated with distribution f is $\sum_k p_k d\left(n_k\right)$ while the cost associated with $\tilde{f}$ is $\sum_k \tilde{p}_k d\left(n_k\right)$, where $\tilde{p}_k = \sum_{x \in W_k} \tilde{f}_x$.

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

Writing $P_k = \sum_{k' \leq k} p_{k'}$ and $\tilde{P}_k = \sum_{k' \leq k} \tilde{p}_{k'}$, we have:

$$\begin{aligned} & \sum_k p_k d(n_k) \\ = \; & P_1 d(n_1) + (P_2 - P_1) d(n_2) + ... + (P_{K-1} - P_{K-2}) d(n_{K-1}) + (1 - P_{K-1}) d(n_K) \\ = \; & P_1 (d(n_1) - d(n_2)) + ... + P_{K-1} (d(n_{K-1}) - d(n_K)) + d(n_K) \\ \geq \; & \tilde{P}_1 (d(n_1) - d(n_2)) + ... + \tilde{P}_{K-1} (d(n_{K-1}) - d(n_K)) + d(n_K) \\ = \; & \sum_k \tilde{p}_k d(n_k), \end{aligned}$$

where the inequality is due to two facts: (1) By Proposition 1, n_k in the optimal code $\mathcal{C}$ is nonincreasing in k and therefore $d(n_{k-1}) - d(n_k) \geq 0$ for all k; (2) By the definition of inequality, $P_k > \tilde{P}_k$ for all k. As $\sum_k \tilde{p}_k d(n_k)$ is not lower than the minimal diagnosis cost for $\tilde{p}$, the statement is proven. ■

An unequal distribution means that there are few extremely likely events and a large number of rare events. The optimal code involves narrow words for the likely events and broad words for the others. This is a good situation from the viewpoint of communication costs, because the organization is likely to end up with an event that is represented by a narrow word. The worst-case scenario occurs when all events are equiprobable. Then, words will divide the event space into equiprobable sets, and this will impose a high communication cost.

An immediate consequence of this argument is that increasing the number of words from 1 to a strictly positive number lowers communication costs more for a more unequal distribution that for a more equal one. 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. The reason is that communication costs are independent of the distribution of events when the number of words is either 1 or very large.[7] But by proposition 3, the communication costs are lower for the more unequal distribution for all $k > 1$.

Two contrary effects determine the marginal value of enriching the code by a word. First, each word is more precise when the distribution is more concentrated, so that each word is more valuable in this case. On the other hand, if the distribution is concentrated a few words added are sufficient to transmit the bulk of the necessary information. Which effect dominates? The value of an additional word is not monotonic in the equality of the underlying density, by the argument in the previous paragraph. We expect

[7] More precisely, as the number of words become very large, communication costs converge to 0 whatever the distribution of events.

that, when the language is poor (it has few words) adding additional words is more valuable the more unequal the environment, but when the language is already quite rich, adding more words eventually is more valuable in a more equal environment.

2.4 Shared codes and dialects

Up to now we have considered only the optimal choice of codes with two agents: a salesman and an engineer. What happens if there are multiple agents on one side? For instance, suppose that there are two salesmen: A and B. They both face the same set of events X but they have different distribution functions: f_x^A and f_x^B.

All three agents – the two salesmen and the engineer – 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, uses code $\mathcal{C} = \mathcal{C}_A \cup \mathcal{C}_B$.

For instance, let $X = \{\{1, 2, 3, 4, 5, 6\}\}$. We could have codes:

$$\begin{aligned} \mathcal{C}_A &= \{\{1,4\}, \{2,5\}, \{3,6\}\} \\ \mathcal{C}_B &= \{\{1,2,3\}, \{4,5,6\}\} \\ \mathcal{C} &= \{\{1,4\}, \{2,5\}, \{3,6\}, \{1,2,3\}, \{4,5,6\}\} \end{aligned}$$

It easy to see that in the optimal solution the upper bound on the engineer's language ability must be binding: code $\mathcal{C}$ will have exactly K words. One may wonder whether it can be the case that codes $\mathcal{C}_A$ and $\mathcal{C}_B$ differ, and therefore the two agents use codes with less than K words. They codes could have entirely different words, as in the example above, or they could share some words but have other words that differ (*dialects*). We can prove that the optimal configuration involves the agents using identical codes:[8]

Proposition 4 *Only a common code can be efficient: in the optimal solution, $\mathcal{C} = \mathcal{C}_A = \mathcal{C}_B$.*

Proof. We will show that a code that is not entirely common is strictly dominated by another code.

Suppose that the code $\mathcal{C}_A$ that the engineer uses to communicate with A and the code $\mathcal{C}_B$ that he uses to communicate with B are not entirely common. Let W_k be the narrowest noncommon word in the codes,[9] and suppose

[8] Note that there is a trivial sense in which we can always have common codes. Starting from $\mathcal{C}_A$ and $\mathcal{C}_B$ just assign $\mathcal{C} = \mathcal{C}_A \cup \mathcal{C}_B$ to both salesmen. This does not mean that salesman A uses $\mathcal{C}$. He may continue to use only $\mathcal{C}_A$.

A more formal way to state this is that $\mathcal{C}$ contains only disjoint sets.

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

without loss of generality that $W_k \in \mathcal{C}_A$. Transform $\mathcal{C}_B$ into $\tilde{\mathcal{C}}_B$ as follows by adding W_k. That is, $W \in \tilde{\mathcal{C}}_B$ if and only if $W = W'/(W' \cap W_k)$ for some $W' \in \mathcal{C}_B$ or $W = W_k$. Notice that this is feasible, as the bounded rationality constraint of B cannot be saturated by $\mathcal{C}_B$, as agent B knows at least one word less than e.

By construction, $\tilde{\mathcal{C}}_B$ has one more word than $\mathcal{C}_B$ but this word is common to $\mathcal{C}_A$. Thus, the total number of words is unchanged and the new code is feasible. Yet, for every event 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 code is strictly more efficient than the older. ■

Two examples will illustrate the proof.

Let $\mathcal{C}_A = \{\{1,4\},\{2,5\},\{3,6\}\}$ and $\mathcal{C}_B = \{\{1,2,3\},\{4,5,6\}\}$. The narrowest noncommon words are $\{1,4\}$, $\{2,5\}$, and $\{3,6\}$. Take $W_k = \{1,4\}$. Then, $\tilde{\mathcal{C}}_B = \{\{1,4\},\{2,3\},\{5,6\}\}$. Each event 1 through 6 is now represented by a shorter word. Diagnosis cost must go down. The total number of words is still five: $\tilde{\mathcal{C}} = \{\{1,4\},\{2,3\},\{5,6\},\{2,5\},\{3,6\}\}$.

As a second, example, we show that $\mathcal{C}_A$ and $\tilde{\mathcal{C}}_B$ are still not efficient. Take $\{2,5\}$ as the narrowest noncommon word. The new code is $\tilde{\mathcal{C}} = \{\{1,4\},\{2,5\},\{3\},\{6\},\{3,6\}\}$, still five words but obviously more efficient.

A more complicated example is

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

Take $\{1,2,3\}$ as the narrowest noncommon word. The new code for B is

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

Events $\{1,2,3,4,7,11,5,8,12,6,9,13\}$ are now represented by shorter words and $\{10,14,15,16\}$ is unchanged.

Proposition 4 has a useful implication. The problem with two salesmen is mathematically identical to the problem with one salesman who faces a distribution of events corresponding to the expected distribution of events for the two salesmen.[10]

Corollary 1 *Suppose that salesman A has m_A clients and salesman B has m_B clients. Then, Propositions 1 and 2 apply as stated if one defines* $f_x = \frac{m_A f_x^A + m_B f_x^B}{m_A + m_B}$.

[10]If there are more than two salesmen, Proposition 4 and Corollary 1 can be applied iteratively.

These results have implications for the design of optimal organizations. If an agent belongs to a certain communication network, it is optimal that he shares the code used by the rest of the network. Agents cannot speak 'pidgin versions' of the team code. The organizational designer then faces a stark choice: either the agent is fully integrated in the team or he is entirely out.

The two results of this section have been obtained under the assumption that all agents suffer from bounded rationality in the same way (K is the same for all). If agents are heterogeneous, things may change. In particular, if the engineer has a higher K than the salesmen it may be optimal to have different codes. This is obviously the case is the engineer can learn twice as many words as the salemen. In Section 3 we entertain this possibility formally by assuming that the firm can hire, at an additional cost, an agent with a higher K.[11]

In what follows, we explore the organizational implications of the need for a common code to support communication. If agents must employ a common code when communicating to the same third party, the organization must determine whether the benefit from having them communicate with each other outweights the loss in precision that is required by the need for a common code. This is the question that we deal with in the next section. But first, since the choice is simply between common or not common code, we simplify our model to two word codes.

2.5 Two Word Codes

In the remaining of the paper, we will focus on a simple specification.[12] Suppose that a salesman deals with consumers $x \in [0,1]$ drawn from a distributions with cumulative distribution function

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

and density

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

with $b \in [-1,1]$. Here b is a measure of how unequal the distribution of events is.

[11] In Appendix A.3, we extend the present model to a situation with multiple engineers and a continuum of salesmen. We discuss the interaction between optimal code and optimal *span*, i.e. the proportion of salesmen who report to each engineer.

[12] Strictly speaking, this is not a special case of the general problem because X was assumed to be finite.

Also assume that codes can have at most two words, $K = 2$, and that the diagnosis cost is linear: $d(n_k) = n_k$.

By Proposition 1, the optimal code will be such that one word contains the more frequent events and the other word contains less frequent events. The optimal code problem reduces to

$$S(b) = \min_x F(x)\, x + (1 - F(x))(1 - x)$$

This optimization has a closed form solution[13] yielding the optimal cut-off point x and the minimum search costs $S(b)$, which allow us to simplify the analysis of the organizational implications of the common code.[14]

For future reference, we note that the optimal cut-off point is:

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

and the resulting expected diagnosis cost is

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

3 Integration, separation and hierarchy

The previous section studied communication in exogenously given organizations. This section endogenizes the organizational structure and looks at how the need to achieve optimal communication shapes the organization. We will ask who should communicate with whom and what code they should use.

We develop a simple model with two services, A and B. Each of them is composed of one salesman and one engineer. We shall study communication and coordination among the two services. We focus on three possible organizational forms: (1) Separation (the two services use different codes); (2) Integration (the two services share the same code); and (3) Translation (there exists a hierarchical structure supplying an interface between the services). This section determines the circumstances under which each form is

[13] The diagnostic cost S is not convex in x. However, its derivative is a second degree polynomial, of which it is possible to show that it is negative on $[0, \hat{x})$ and positive on $(\hat{x}, 1]$.

[14] The optimal cut-off point is: $\hat{x} = \frac{1}{6b}\left(3b - 2 + \sqrt{(3b^2 + 4)}\right)$ with

$$D^*(b) = \frac{8 + 36\, b^2 - \left(4 + 3\, b^2\right)^{\frac{3}{2}}}{54\, b^2}.$$

.

optimal. For expositional reasons, it is best to focus first on the comparison between the two pure forms, separation and integration, and then introduce the third form. But before that, we explicitly model the source of synergies.

3.1 Preliminaries: Synergies

To generate a need for coordination, there must be a potential synergy among the two services, which we model as follows. Customers arrive randomly, and there may be excessive load in one service and excessive capacity in the other. If that happens, the two services benefit from diverting some business from the overburdened service to the other. Formally, suppose that salesmen from services A and B deal with consumers from two different distributions F_A and F_B,

$$F_i(x) = (1 - b_i)\, x + b_i x^2, \qquad i = A, B \tag{4}$$

with $b_A = b$ and $b_B = -b$ and $b \in [-1, 1]$ measuring the similarity between the two distributions. Let x_i^* the cutoff between words of each service, with (by symmetry) $x_B^* = 1 - x_A^*$, and $D_i^*(x)$ the expected diagnosis cost in either service.

In a given period, salesman i receives a random number y_i of clients, with the following distribution (see Figure 4):

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

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.

Each engineer has the ability to attend to the needs of at most one client. If p is low, then each salesman is likely to find one client per period. 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.

Finally, we assume that the profit that can be obtained when a client's problem is solved is 1. The per-client diagnosis costs is $\lambda \in (1, 2)$, so that if the engineer knows that the client's characteristics fall in an interval of size s, his diagnosis cost is $s\lambda$. This ensures positive profits. It also ensures that information must transit through a salesman before being sent to an engineer; indeed an engineer without information on the client's problem would have diagnosis costs greater than the profits obtained from solving it.

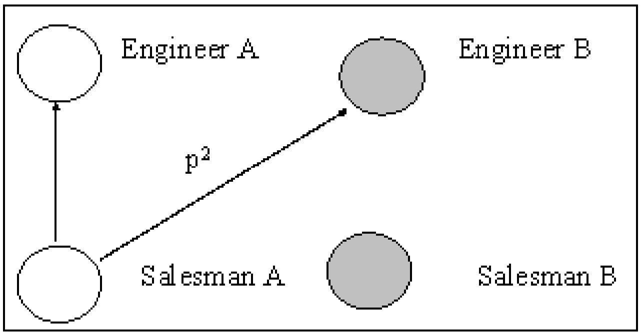


Figure 1: *Synergies exist when there is excess demand on one service and excess capacity in the other.*

3.2 Integration or separation?

The organizational choice here is between segregating the services, so that salesmen from service A only communicate sales leads to engineers in A; and integrating them so that a salesman from A may communicate sales leads to either engineer. Should services communicate with each other, even at the expense of a common code?

Consider an *integrated organization* first. This requires that a salesman from service A explain to an engineer in B the needs of his customer: indeed, because $\lambda > 1$, sending the problem to the engineer without explanation is not profitable. However, in this case the codes must be common in both services.

What are the diagnosis costs in this case? Because the two services have the same language, it is easy to adapt Corollary 1 to prove that the common language is the one that would be chosen when the density of tasks is the average of the two densities of the two services. In this case, since both services have opposing distributions, the average problem density is uniform. The optimal code has two equally imprecise words, with each word identifying the sales lead as coming from one half of the distribution. The total profits then are:

$$\Pi(p, b, \lambda | \mathcal{C}_j) = 2(1 - p(1 - p)) \left(1 - \frac{\lambda}{2}\right). \quad (5)$$

In a *separated organization*, where the two services use different codes, the expected profit is:

$$\Pi(p, b, \lambda | \mathcal{C}_s) = 2(1 - p)(1 - \lambda D^*(b^2)), \quad (6)$$

where $D^*(b^2)$ is determined as in (3).

The organization should be integrated rather than separated if the between service improvement in communication (measured by the synergy gain) is larger than the within service loss in precision due to the worsening of the code used:[15]

$$\frac{1-p(1-p)}{1-p} \geq \frac{1-\lambda D^*(b)}{1-\frac{\lambda}{2}}. \tag{7}$$

The following proposition shows that an increase in the synergy parameter p, a decrease in the diagnosis cost λ or a decrease in $|b|$, the divergence in the distribution of tasks, makes the integrated organization more profitable.

Proposition 5 *An integrated form becomes relatively more profitable when: the synergy parameter p increases, the diagnosis cost λ decreases, or the inequality in task distributions $|b|$ decreases.*

Proof. Using (5) and (6), define

$$V(b, \lambda, p) = 2(1-p(1-p))\left(1-\frac{\lambda}{2}\right) - 2(1-p)(1-\lambda D^*(b^2)),$$

where $D^*(b^2)$ is determined as in (3). We have

$$\partial V/\partial \lambda = -(1-p)(1-2D^*(b)) - p^2 < 0,$$

since the diagnosis cost $D^*(b)$ is bounded above by $1/2$, i.e. the cost incurred when words are equal length. We also have

$$\partial V/\partial p = \lambda(1-2D^*(b)) + 4p(1-\lambda/2) > 0$$

since $\lambda < 2$. Finally, the comparative statics with respect to b are a direct consequence of the fact that $D^*(b)$ is decreasing in b^2 for $b \in [0,1]$. ■

Proposition 6 characterizes the determinants of the trade-off between separate, well-adapted codes optimized for within-service communication and broader, common codes that allow for between-service communication. Separate codes are preferable when synergies are relatively low, when the underlying probability distributions confronting the different units are sufficiently different, or when diagnosis costs are high so that there is a high premium on communicating precisely. As a result, increases in synergies, in the equality of the distributions or decreases in diagnosis costs should result in more code commonality.

[15] 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 (7) is a concave function of b, which is positive on $(-0.684, 0.684)$ and negative outside this interval.

3.3 Hierarchy

Suppose instead that the two services may exploit the synergy by employing a fifth agent who provides translation among the two services. Each service adopts a separate code. 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. For simplicity we make the extreme assumption that the translator's λ is zero. The qualitative results of the analysis go through even if his λ is strictly positive, as long as it is lower than the engineers' λ.

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.

Proposition 6 *For any b, if p is high enough and μ is low enough, there exist $1 \leq \lambda_{\min} < \lambda_{\max} \leq 2$ such that the unique optimal organization is*

integrated	*if* $\lambda < \lambda_{\min}$
hierarchical	*if* $\lambda \in (\lambda_{\min}, \lambda_{\max})$
separated	*if* $\lambda > \lambda_{\max}$

Proof. See Appendix. ■

To understand this proposition, refer to Figure 2 which presents a comparison between the three organizational forms as a function of synergies p and diagnosis costs λ. Consider first the comparison between translation 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 communication cost incurred by translation is low and the net benefit is likely to be high. Thus, translation is more likely to beat separation when λ is low. Consider now the choice between translation and integration. Translation saves on communication cost by allowing services to keep efficient service-specific codes – thus translation is likely to beat integration when λ is high, since communication savings are more important when λ is high. Thus as the proposition shows, if the fixed cost μ of hiring a translator is low enough, there exists an interval of λ for which the hierarchical structure is optimal.

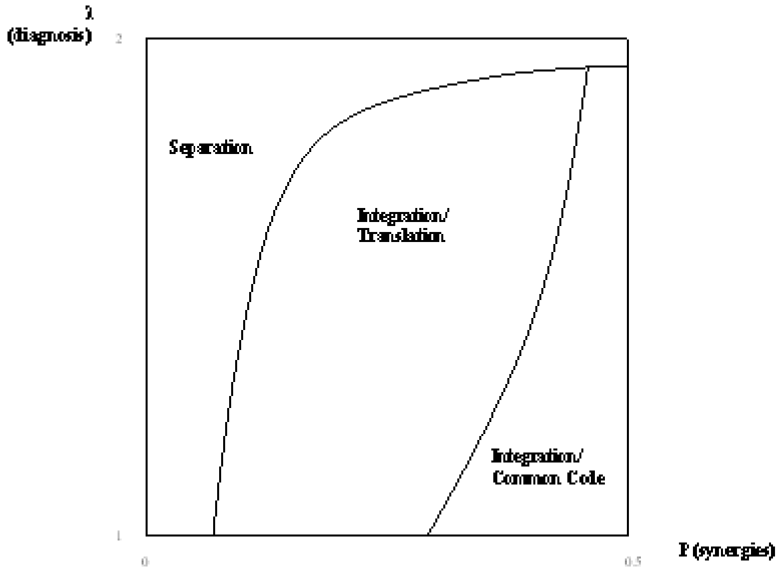


Figure 2: Simulation of organizational structure and code design. $b = 0.4$ and $\mu = 0.006$.

3.4 Information Costs and Code Commonality: Evidence

In the past two decades, there has been a dramatic drop in information processing costs caused by the large advances in information and communication technology.[16] In particular, searching for certain kind of information has become extremely cheap. To give an example close to home, suppose a colleague tells me that a certain paper "has the word X in the abstract and was published on journal Y in the last five years" Today, I can retrieve the paper in a matter of seconds on my personal computer. Not too many years ago, I would have had to spend quite some time sifting through past journal volumes.[17]

In the context of our model such a technological change can be interpreted as an exogenous reduction of the diagnosis cost parameter λ. Applying the results above, we should expect to see an increase in 'integration' in the form

[16] For example, according to Intel Corp. the average price of a transistor was 10,000 times higher in 1980 than in 2000.

[17] In the example of the interior decorator and the craftsman discussed earlier, the diagnosis cost has two components: the time/cost spent in retrieving the fixtures and the time/cost spent in transporting them to the client's site. The IT revolution has probably affected only the first part.

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. We discuss next the evidence of the impact of information technology on common codes and on decentralization, and the links between them.

First, the reduction in information costs is correlated with increasing code commonality. Historically, the information generated by each business unit of a firm and within each business unit by each function has been coded and processed separately, according to the needs of that business unit or function. This meant that the different pieces of information were defined in different ways and could not be easily aggregated.[18] As information costs dropped, companies seek ways to integrate this disperse information. This integration was obtained, within firms, through the use of company-wide Enterprise Resource Planning (ERP) systems (such as those produced by German company SAP or Dutch company Baan); and between firms, through the use of Electronic Data Exchanges (EDI),[19] which allow for the exchange of electronic data between suppliers and customers by standardizing the format of the data exchanged. Through these systems, firms have substituted flexible ways to code their data for more rigid but unified central databases.[20]

Second, the reduction in information costs is correlated with increasing 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 Reenen (2001) also find, on entirely different data, evidence that the degree of decentral-

[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] 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).

[20] 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).

ization 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 the CEO is increasing, in particular, through the disappearance in the role of the COO.

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. However, it is not clear that these two changes are causally linked in the way our model suggest, that is, that the introduction of a common code does indeed allow for the substitution of the 'translation' role of hierarchy and allows for direct horizontal communication between units that otherwise would be 'speaking a different language.' Some case study evidence, however, does suggest that this type of change may indeed have taken place.

Consider the organizational changes undergone by the Microsoft Corporation starting in the mid 90s to which we referred in the introduction of this paper. According to Robert J. Herbold,[21] COO of Microsoft at the time, Microsoft had in 1994 a completely decentralized set of information systems. These were separate codes, in the sense of this paper, in that each system produced a different mapping of data to category. In the finance area, the managers of the different units had all set up their own techniques of financial reporting, stressing what they believed were the important components. 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." The situation in the human resources field was the same: there was no consistent, between units, way to keep track of human resources, with eighteen HR-related databases. "When asked about head counts, managers answers usually were, to put it charitably, poetic." The price of having this tailored information was that between unit communication was compromised, as different measures often needed to be reconciled.

Taking advantage of the drop in information costs discussed above, Microsft decided to move towards 'common codes' in those two areas. Among the main advantages of these moves, according to Herbold, were that all managers could easily make sense of that information. Paradoxically, and as our model predicts, this centralizing move provided "benefits usually associated with decentralization" as managers had instant access to relevant information and could operate on it directly.

[21] We rely heavily on his personal account, in Harvard Business Review, January 2000. All the quotes below proceed directly from his account.

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." [22] That the adoption is in the interest of the company does not mean that all it is in the interests of each agent involved. This matters when 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. These are particularly important when separate firms, necessarily involving separate decision makers, are involved.[23]

4 Strategic Code Adoption

So far, we have assumed that all agents share the goal of maximizing social surplus. We now allow for the presence of a conflict of interest and we study its effect on organizational codes.

Even in the presence of diverging objectives, with complete contracts, agents would agree to select the surplus-maximizing code and, if necessary, to make appropriate side payments. It is thus interesting to leave such a Coasian world and assume that code adoption is non-contractible. It is conceivable that firms cannot sign contracts that commit them to adopting a particular code. Perhaps, this is because outsiders cannot verify the actual code that is being used in internal communications unless they are given, at prohibitive cost, full access to the firm. The fact that a particular organization formally adopts a certain code does not mean that its members actually use it.

We first examine sequential code adoption. We ask what are the incentives for the firm that moves first. Knowing that its decision affect other firm's decisions on codes, what kind of code should the first mover choose? We then analyze how free-riding affects code adoption. We start from a sit-

[22] 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 viewpoint corresponds more closely to reality.

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

uation 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

Consider the same set-up as in the previous section, with two services, A and B with cdf's given by $F_i(x)$ as in the previous section (see (4)). However, these two services are now two separate firms: salesman A and engineer A belong to firm A while the other two agents make up 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 she has not received a customer from her salesman. The surplus created by the relationship goes in proportion σ to the salesman and $1-\sigma$ to the engineer.[24]

Timing is sequential. First, firm A adopts a code. Then, firm B observes the code adopted by A and selects its own code, either a specific code, or the code that A has chosen. Finally, customers arrive and the services behave as in the previous section given the organizational codes available.

As in the previous section, the advantage of choosing the same code is to open the possibility of "trade". The payoffs with separate codes are given by

$$\pi(p, b|\mathcal{C}_s) = (1-p)(1-\lambda D^*(b^2))$$

as in the previous section.

With a joint code, the profit of firm i is:

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

where $s_i\,(x)$ represents the expected diagnosis cost for an engineer who receives a task from her own salesman and $s_j\,(x)$ is the expected diagnosis cost for an engineer who receives a task from the other salesman. These costs are:

$$\begin{aligned} s_i\,(x) &= F\,(x)\,x + (1-F\,(x))\,(1-x)\,; \\ s_j\,(x) &= F\,(x)\,(1-x) + (1-F\,(x))\,x. \end{aligned}$$

The expected profit is:

$$\Pi_i(p, b, \sigma|\mathcal{C}_j) = 1-p+p^2-\lambda\left(\left(1-p+p^2\sigma\right)s_i\,(x)+p^2\,(1-\sigma)\,s_j\,(x)\right).$$

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

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

$$\begin{aligned} &\max \quad \Pi_A(p, b, \sigma|\mathcal{C}_j) \\ &\text{subject to } \Pi_B(p, b, \sigma|\mathcal{C}_j) \geq \pi^S. \end{aligned} \tag{8}$$

Then we can show the following result:

Proposition 7 *If adopting a joint code is efficient, in equilibrium a joint code is adopted. However, the adopted code is not the optimal (symmetric) code with* $x = \frac{1}{2}$,*but it is inefficiently biased towards the environment of the first mover* ($x > \frac{1}{2}$).

Proof. See Appendix. ■

In this setting, contractual incompleteness does not result in situations in which firms keep separate codes even when it would be efficient to have a common code. However, the code that is adopted is suboptimally skewed towards the needs of the first mover.

The firm that moves first takes only into account its expected profit. This includes the cost of internal communication and a portion of the cost of inter-firm communication; but it does not take into account the cost of internal communication for the follower. The first mover minimizes his communication cost by selecting a code that fits its environment. The equilibrium code differs from the efficient code which fits the "average" environment that the two firms face. The 'selfishness' of the first-mover is limited only by the 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

In the previous result, there is inefficiency with respect to *what* common code is adopted but not with respect to *whether* a common code is adopted at all. We now show that the presence of fixed adoption costs can lead to situations in which a common code is not adopted at all.

Suppose firms are endowed with separate codes, so that a firm that wants to switch to a different code must sustain a fixed cost c. Suppose that the environment changes and that it is now efficient, taking into account switching costs, to have a common code. However, switching codes is non-contractible: a firm cannot make side payments to the other for adopting a new code.

There are potentially three cases: both firms keep separate codes; one firm adopts the code of the other firm; or both firms adopt a joint code. Suppose the efficient solution is for one firm to adopt the other firm's code (which occurs for intermediate values of c), and denote the common code when B adopts A's code with the superscript SJ (as in "semi-joint"). The expected payoffs are respectively

$$\pi_A(p,b,\sigma|\mathcal{C}_{SJ}) = 1-p+p^2-\lambda\left(\left(1-p+p^2\sigma\right)s_A\left(x_A^S\right)+p^2\left(1-\sigma\right)s_B\left(x_A^S\right)\right)$$
$$\pi_B(p,b,\sigma|\mathcal{C}_{SJ}) = 1-p+p^2-\lambda(\left(1-p+p^2\sigma\right)s_B\left(x_A^S\right)+p^2\left(1-\sigma\right)s_A\left(x_A^S\right))$$

From an efficiency point of view, SJ is optimal when:

$$\pi_A^{SJ}+\pi_B^{SJ}-c \geq \max\left(\pi_A^S+\pi_B^S, \pi_A^J+\pi_B^J-2c\right)$$

where $\pi_A^{SJ}=\pi_A(p,b,\sigma|\mathcal{C}_{SJ})$ etc. for simplicity. Note that if λ is small enough and c is high enough, the above must be satisfied. Suppose thus that we are in the region in which SJ is efficient. Firm B switches to firm A's code if:

$$\pi_B^{SJ}-c \geq \pi_B^S$$

Note however that $\pi_A^{SJ} > \pi_B^{SJ}$ (because code x^S is geared toward firm 1's needs). Thus,

$$\frac{1}{2}\left(\pi_A^{SJ}+\pi_B^{SJ}-c\right) > \frac{1}{2}\left(\pi_A^{SJ}+\pi_B^{SJ}\right)-c > 2\pi_B^{SJ}-c$$

So, the fact that SJ is efficient is no guarantee that B is willing to adopt it.[25]

Proposition 8 *In the presence of switching costs, firms may end up keeping separate codes when it would be more efficient for one firm to switch to the other firm's code.*

By adopting a common code, a firm generates a positive externality for the other firm, since a common code between A and B facilitates the communication of messages from A to B and from B to A. As a result, under certain conditions, firms that should have a common code inefficiently remain with separate codes and lose the potential synergy benefits that could be obtained with better communication.

[25] Note that this result is still true even when firms share the cost of adoption in equal parts. This is because firm 2 still incurs the cost of adopting a code that is suboptimal for internal communication.

4.3 Strategic 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.

Advances in information technology made it possible for the companies involved in the design of the 'stealth' B-2 bomber by Northrop, Boeing, Vaught (a division of LTV) and General Electric to create a centralized database to facilitate the design of the bomber. A key element of the program would be the 'B-2 Product Definition System'. This was essentially a common code, a "technical 'grammar' by which engineers and others conveyed information to each other."[26] However, the strategic considerations analyzed above played an important part in the complicated process towards the adoption of a common code. First, Boeing and Vaught were unenthusiastic about the adoption of a common code during the negotiations leading to the creation of a centralized database. A Boeing engineer explained that '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). That the common approach was probably, in spite of the resistance, optimal is seen by the fact that the Air Force – arguably concerned with achieving the efficient outcome in this context- was willing to pick up the training costs incurred by Boeing and Vaught (Argyres, 1999: 166). It appears thus that the presence of such a central player avoided the tendency towards 'excessive code variety' identified in Proposition 9. Second, not only the adoption was hard to attain, but it was also, as Proposition 8 suggests, biased towards the needs of the early adopter, Northrop. Rather than generating a common code, which would presumably fit the needs of all players, all parties adopted Northrop's (Argyres, 1999:167) system.

When the code was finally agreed, this 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).[27] The use of the grammar had two consequences. First, it allowed for designers proceeding from different companies

[26] "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).

[27] The account that follows draws heavily on a detailed cased study by Argyres (1999).

to 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.[28] Thus, the existence of a common code allowed integration of several teams were before there was none possible, an effect illuminated by Proposition 6. Moreover, this integration happened with little need for hierarchical coordination, since among the main consequences of the creation of a relatively rigid, unifying codes was an increase in decentralized decision making and the reduction in the need for a hierarchy vis-a-vis previous projects: "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 the predictions of Proposition 7 concerning the substitution of hierarchies by codes.

5 Related literature and conclusions

This paper constitutes an initial step towards an integrated theory of codes and organization. In this section, we review briefly the existing literature on codes.

The idea that there is a trade-off between generality and specialization of codes, explored for instance in Sections 3 and 4 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 a particular organization.

Information theory (Shannon, 1948) also deals with the design of optimal codes. The constraint is *channel capacity*, and codes are chosen to minimize the cost of transmitting information. The theory studies questions such as which are the binary sequences resulting in minimum mean length message attainable, what is the trade-off between decoding error and mean length, what is the impact of noisy channels on mean length and decoding etc. Clearly, there is a key difference with our approach. In information theory, the sender must transmit all of the information, and the question is: what code minimizes transmission cost? In our setting, the transmission cost is given, but the sender is prevented from transmitting all information. The question is: what code maximizes the value of information transmitted? As a consequence, we we pose are entirely different set of questions which have

[28] Argyres, personal communication to the authors.

to do with organizational structure rather than channel utilization.[29]

Crémer (1993) 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. 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 consider the possibility of communication within the team.

Like us, Wernerfelt (2003) considers codes that are enacted to minimize communication costs within an organization. But the focus of his work is different. First, his paper does not characterize the optimal codes, but focuses instead on whether, when agents have common interests but decisions are decentralized, equilibria with one or multiple codes exist, and whether these equilibria are symmetric or asymmetric. His analysis does not consider either the organizational implications of the use of codes nor the strategic aspects of such choices. Instead, our approach here studies how the environment in which the organization operates determines the optimal code, studies explicitly the organizational implications (for scope, hierarchy, the relation between information and centralization etc.) and considers the strategic aspects of such choices.

Building on Marschak and Radner's (1972) *team theory* a number of authors have studied the limits that bounded rationality places on communications 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 an organization theory built on the trade-off between communication costs and returns to specialization.

[29] 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.

In Garicano (2000), the bounded rationality of agents prevents them from learning how to solve all the problems that the organization faces. On the other hand, agents can request help from other agents when they do not know how to solve one of these problems. He shows that the firm will organize itself in a 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 (see also Garicano and Ross-Hansberg (2003) for the application of this model to the determinants of wages in hierarchies).

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 and we have shown the causal mechanism we propose is consistent with the one present in some detailed case studies of decentralization and organization.

There are several avenues for future research uncovered by the model. Perhaps the most interesting of these are empircal. The availability of large databases of business texts and their ease of access allows for a study of the commonality or lack thereoff of the language used across different serives 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 'de-layering' (less hierarchy) and more horizontal communication where the commonality of codes increased more.

On a theoretical front, 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 teh 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).[30]

Another promising research avenue concerns the interaction between organizational codes and labor market dynamics. A worker who learns an orga-

[30] See Appendix A.4 for a preliminary effort in that direction

nizational 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 direct 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.

Economists have a comparative advantage in the study of incentive problems, but we feel that the problems of bounded rationality are important elements of a theory of organizations: even a firm composed of honest agents, who do not lie and work to the maximum of their abilities, would face organizational problems. It is therefore important that these elements be integrated in our theories. In particular, we have shown that the bounded rationality of the employees makes it necessary to limit the flexibility of individual divisions to choose their own codes. We believe that the study of the homogeneity of the decision making processes within the firm is an important topic both on theoretical and applied grounds, that much more work is needed in this area, and that advances in this direction will require both a richer theory of codes and attention to the other dimensions of bounded rationality.

References

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

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

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

[4] Blankevoort, P. J. "Contradictory effects of a project management dictionary." *International Journal of Project Management*, Vol. 4 (4), pp. 236-238, 1986.

[5] Patrick Bolton and Mathias Dewatripont. "The firm as a communication network.", 104(4): 809–839, 1994.

[6] 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, Vol. 117, February 2002.

[7] Brynjolfsson, Eric, and Loren Hitt. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *Journal of Economic Perspectives*, Vol. 14 (4), pp. 23-48.

[8] Browning, Larry D., Janice M. Beyer, Judy C. Shetler. "Building Cooperation in a Competitive Industry: SEMATECH and the Semiconductor Industry." *The Academy of Management Journal,* Vol. 38, No. 1. (Feb, 1995), pp. 113-151.

[9] Caroli, Eve and John Van Reenen, "Skill Biased Organizational Change? Evidence from a Panel of British and French Establishments." *Quarterly Journal of Economics*, Vol. 116, pp. 1449-1492, November 2001.

[10] Crémer, Jacques. "Corporate Culture: Cognitive Aspects," *Industrial and Corporate Change*, 3, 2, pp. 351-386, 1993,

[11] Crémer, Jacques. "A Partial Theory of the Optimal Organization of a Bureaucracy", *Bell Journal of Economics*, Autumn 1980, 11, 2, 683-693.

[12] Garicano, Luis. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy*, Vol. 108, No. 5, pp. 874-904, October 2000.

[13] Garicano, Luis and Esteban Rossi-Hansberg. "Organization and Inequality in a Knowledge Economy." Mimeo, 2003.

[14] Hart, Paul and Carol Saunders, "Power and Trust: Critical Factors in the Adoption and Use of Electronic Data Interchange." *Organization Science*, Vol. 8 (1), pp. 23-42, 1997.

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

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

[17] 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.

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

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

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

[21] 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.

[22] Wermerfelt, Birger

A Appendix

A.1 Proof of Propositions [RE-ARRANGE LAYOUT. CHECK PROOFS]

A.1.1 Proof of Proposition 6 (Hierarchy)

The expected payoff with translation is

$$\Pi(p,b,\mu|\mathcal{C}_t) = 2(1-p)(1-\lambda D^t(b)) + 2p^2\left(1-\lambda\tilde{D}^t(b)\right) - \mu,$$

where

$$\begin{aligned} D^t(b) &= F\left(x^t\right)x^t + \left(1-F\left(x^t\right)\right)\left(1-x^t\right), \\ \tilde{D}^t(b) &= F\left(x^t\right)\left(1-x^t\right) + \left(1-F\left(x^t\right)\right)x^t, \end{aligned}$$

and

$$x^t = \arg\min_x 2(1-p)\left(F\left(x\right)x + \left(1-F\left(x\right)\right)\left(1-x\right)\right) + 2p^2\left(F\left(x\right)\left(1-x\right) + \left(1-F\left(x\right)\right)x\right).$$

We compare it to the expected payoffs in the other two forms:

$$\begin{aligned}\Pi(p,b|\mathcal{C}_j) &= 2(1-p(1-p))\left(1-\frac{\lambda}{2}\right)\\ \Pi(p,b|\mathcal{C}_s) &= 2(1-p)(1-\lambda D^*(b))\end{aligned}$$

For given b and p,

$$\lambda^* = 2\frac{p^2}{p^2+(1-2D^*(b^2))\,(1-p)},$$

is the value of λ for which $\Pi(p,b|\mathcal{C}_j) = \Pi(p,b|\mathcal{C}_s)$. Because

$$\frac{44-7\sqrt{7}}{54} \leq D^*(b^2) \leq \frac{1}{2},$$

it is straightforward to check that $p \geq p^* = 0.213$ (the "p high enough" of the proposition) implies $\lambda^* \in [1,2]$. Let

$$\mu^* = 2p^2\left(1-\lambda^*\tilde{D}^*(b^2)\right).$$

If $\lambda = \lambda^*$ and $\mu = \mu^*$,

$$\Pi(p,b,\mu|\mathcal{C}_t) = \Pi(p,b|\mathcal{C}_j) = \Pi(p,b|\mathcal{C}_s).$$

Consider $\mu < \mu^*$ (the "μ low enough" of the proposition). Il $\lambda = \lambda^*$, translation dominates the other two forms. If $\lambda > \lambda^*$, the optimal form cannot be separation. If $\lambda < \lambda^*$, the optimal form cannot be integration. These last three statements, combined with the observation that $\Pi(p,b,\mu|\mathcal{C}_t)$, $\Pi(p,b|\mathcal{C}_j)$, and $\Pi(p,b|\mathcal{C}_s)$ are all linear in λ proves that the set of λ's for which translation is optimal is an interval that contains λ^*. To the left of the interval, separation is optimal. To the right, integration is optimal.

A.1.2 Proposition 7 (First mover bias)

Given the symmetry of the two firms, a the most efficient joint code is the symmetric one. A joint code is strictly superior to a separate code form an efficiency point of view if and only if

$$\Pi_A(p,b,\frac{1}{2}|\mathcal{C}_j) = \Pi_B(p,b,\frac{1}{2}|\mathcal{C}_j) > \pi(p,b|\mathcal{C}_s). \tag{9}$$

If this inequality does not hold, the value of (8) is lower than the expected profit under a separate code: firm A will choose its optimal code as if B did not exist, and B will choose a separate code.

Suppose instead that (9) is satisfied. Firm A will choose the code that minimizes

$$\left(1-p+p^2\sigma\right) s_A(x)+p^2(1-\sigma) s_B(x) = (1-p+p^2(2\sigma-1))s_A(x)+p^2(1-\sigma)(s_A(x)+s_B(x))$$

subject to the participation constraint for B. Note that

$$\begin{aligned}\frac{d}{dx}\pi_A^J(x) &= -\lambda\left(\left(1-p+p^2\sigma\right) s'_A(x)+p^2(1-\sigma) s'_B(x)\right)\\ &= -\lambda\left(\left(1-p+p^2(2\sigma-1)\right) s'_A(x)+p^2(1-\sigma)\left(s'_A(x)+s'_B(x)\right)\right)\end{aligned}$$

By symmetry,

$$s'_A(x)+s'_B(x) \leq 0$$

if and only if $x \leq \frac{1}{2}$. Also, it is easy to see that if $x \leq \frac{1}{2}$ and, as we have assumed, f is strictly increasing,

$$s'_A(x) = 2f(x)(2x-1)+2F(x)-1 < 0.$$

$\pi_i^J(x) = 1-p+p^2-\lambda\left(\left(1-p+p^2\sigma\right) s_i(x)+p^2(1-\sigma) s_j(x)\right)$

Hence,

$$\text{if } x \leq \frac{1}{2} \qquad \frac{d}{dx}\pi_A^J(x) > 0. \tag{10}$$

If the participation constraint is not binding, firm 1 faces an unconstrained maximization problem over $\pi_A^J(x)$. By (10), the optimal x is to the right of $\frac{1}{2}$.

Suppose instead that the participation constraint is binding. Because a joint code is strictly superior, there is an interval $\left[\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon\right]$ with $\varepsilon > 0$ such that for all the x in the interval $\pi_B^J(x) \geq \pi^S$. But, by (10), this implies that the optimal x is to the right of $\frac{1}{2}$.

A.2 Extension of propositions 1 and 2 to 'Natural Ordering' case

Suppose that events are aligned along the real line, and that the frequency of events is described by a continuous and differentiable, but possibly non-monotonic, probability distribution f on $[0,1]$. Words are contiguous intervals in the real line. As before, the familiarity of a word[31] $[t_k, t_{k+1}]$, is the probability that the word is used, now $F[t_{k+1}] - F[t_k]$, and the breadth of

[31] 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.

the word is the 'number of events' in the word, that is here the size of the interval, that is $t_{k+1} - t_k$.We define finally the 'average frequency' of the events in the word as the average height of the density over those 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 9 (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.*

Proof. The best K-words code is solution of

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

subject to

$$\begin{aligned} &t_k \text{ increasing in } k, \\ &t_0 = 0, \\ &t_K = 1. \end{aligned}$$

The first-order conditions[32] are

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

[32] It is easy to check that at the optimum $t_k < t_{k+1}$ for all k. Assume for instance that we have $t_0 < t_1 = t_2 = t < t_3$. Increase t_2 by a small x. The increase in cost is equal to

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

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 in this special case. It is clear that the reasoning generalizes.

Economically, the marginal cost of a very small word is zero, but adding it has a marginally strictly negative impact on the cost of adjoining words.

which implies

$$f(t_k) = \frac{[F(t_{k+1}) - F(t_k)] - [F(t_k) - F(t_{k-1})]}{(t_k - t_{k-1}) - (t_{k+1} - t_k)} \tag{11}$$

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, rearrange (11):

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

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

A.3 Assigning agents to codes

Agents within a service, i.e. those dealing with one particular engineer, must share a code. The question that naturally follows is: which agents should share a code to minimize communication costs?

Suppose in particular that there are two services, each with one engineer e_A and e_B, and a continuum of salesmen.[33] Each salesman has a linear density function characterized by the parameter $b \in [-1, 1]$. The distribution of salesmen is $g(b)$. Each salesman and each engineer can learn a maximum of two words. The solution of the problem is to divide the salesmen into two subsets, S_A and S_B. Salesmen in S_i know the same language as engineer e_i. Moreover, in the optimal solution $S_A = [-1, b^*]$ and $S_B = (b^*, 1]$, where $b^* \in (-1, 1)$.

[RE-DO: define diversity; proposition is unclear, check proof, intuition after proposition is incomplete]

Let the *span* of engineer i be the proportion of salesmen that she serves: $F(b^*)$ for e_A and $1 - F(b^*)$ for e_B. Let the *diversity* of engineer i be the range of salesman types that the engineer communicates to: $1 + b^*$ for e_A and $1 - b^*$ for e_B.

Proposition 10 *If $g(b)$ is increasing and linear, then in the optimal organization, $c_A(b^*) = c_B(b^*)$, with $b^* > 0$. The span of engineer A is smaller*

[33] Dealing with N engineers can be done identically and the solution has the same characteristics as the one discussed below.

than the span of engineer B and the diversity of engineer A is greater than the diversity of engineer B.

Proof. Recall that $f(x,b)$ is the density of tasks x and $g(b)$ is the density of salesmen. We must show that if g(b) is increasing and linear, then in the optimal organization, $c_A(b^*) = c_B(b^*)$, with $b^* > 0$. To obtain this result we proceed in two steps: first, we obtain the optimal language for a given assignment of agents to language, which amounts to obtaining the cutoffs between words x_A and x_B in each language. This generates an optimized cost function that depends on the assignment of agents to language $b.^*$The second step then is choosing b^* so as to minimize this cost.

Given b^*, the problem for language A is

$$c_A(b^*) = \min_{x_A} F(x_A, b < b^*) x_A + (1 - F(x_A, b < b^*))(1 - x_A)$$

and for language B it is

$$c_B(b^*) = \min_{x_B} F(x_B, b > b^*) x_B + (1 - F(x_B, b > b^*))(1 - x_B)$$

where

$$\begin{aligned} F(x_A, b < b^*) &= \frac{1}{G(b^*)} \int_0^{b^*} F(x_A, b) g(b) db \\ F(x_B, b > b^*) &= \frac{1}{1 - G(b^*)} \int_{b^*}^1 F(x_B, b) g(b) db \end{aligned}$$

The problem for b^* is then

$$\min G(b^*) c_A(b^*) + (1 - G(b^*)) c_B(b^*)$$

Consider

$$\Psi(b^*) = \frac{d}{db^*}(G(b^*) c_A(b^*) + (1 - G(b^*)) c_B(b^*))$$

Note that

$$\begin{aligned} c_A(b^*) &= \frac{2x_A(b^*) - 1}{G(b^*)} \int_0^{b^*} F(x_A(b^*), b) g(b) db + (1 - x_A(b^*)) \\ c_b(b^*) &= \frac{2x_B(b^*) - 1}{1 - G(b^*)} \int_{b^*}^1 F(x_B(b^*), b) g(b) db + (1 - x_B(b^*)) \end{aligned}$$

$$\begin{aligned} \frac{d}{db^*} G(b^*) c_A(b^*) &= \frac{d}{db^*}\left((2x_A(b^*) - 1) \int_0^{b^*} F(x_A(b^*), b) g(b) db + G(b^*)(1 - x_A(b^*))\right) \\ &= (2x_A(b^*) - 1) F(x_A(b^*), b^*) g(b^*) + g(b^*)(1 - x_A(b^*)) \end{aligned}$$

$$\frac{d}{db^*}\left(1-G\left(b^*\right)\right)c_B\left(b^*\right) = \frac{d}{db^*}\left(\left(2x_B\left(b^*\right)-1\right)\int_{b^*}^{1}F\left(x_B\left(b^*\right),b\right)g\left(b\right)db+\left(1-G\left(b^*\right)\right)\left(1-x_B\left(b\right.\right.\right.$$
$$= -\left(2x_B\left(b^*\right)-1\right)F\left(x_B\left(b^*\right),b^*\right)g\left(b^*\right)-g\left(b^*\right)\left(1-x_B\left(b^*\right)\right)$$

$$\Psi\left(b^*\right) = g\left(b^*\right)\left(\left(2x_A\left(b^*\right)-1\right)F\left(x_A\left(b^*\right),b^*\right)-\left(2x_B\left(b^*\right)-1\right)F\left(x_B\left(b^*\right),b^*\right)+\left(x_B\left(b^*\right)-x_A\left(b^*\right.\right.\right.$$
$$= g\left(b^*\right)\left(c_A\left(b^*\right)-c_B\left(b^*\right)\right)$$

Thus, the optimum is when

$$c_A\left(b^*\right)=c_B\left(b^*\right)$$

It is easy to see that

$$c_A\left(0\right)<c_B\left(0\right).$$

To see this, note that for the salesman who faces distribution with $b=0$, the best possible code is $x=0$. Because g is increasing, there are more salesmen similar to 0 in service A than in service B, and therefore the optimal code for 0 is closer to the optimal code for service A than to the optimal code for service B, and hence the result in the paper. Moreover, note that

$$\Phi\left(b^*\right)=c_A\left(b^*\right)-c_B\left(b^*\right)$$

is nondecreasing in b^*, and that this function is negative for $b^*=-1$ and positive for $b^*=1$. Then the unique value of b^* for which $\Phi\left(b^*\right)=0$ is to the right of 0, which proves the statement. ■

Since $g(b)$ is increasing, there is a bigger mass of salesmen with $b>0$ than with $b<0$, and the boundary between codes has $b^*>0$. This means that the diversity of the types of salesmen an engineer deals with is inversely related to their number. Thus even though there is no cost in managing more agents, there is a limit to engineer's span: the organizational costs of getting diverse salesmen, as it leads to less adapted codes.

A.4 Dynamics: Persistence and Variability of Codes

In the main text we have studied the static problem of choosing the optimal code. Suppose now that the environment can change over time, and that agents may decide, at a cost, to adapt the environment to these changes. How persistent is the code? how does this persistence depend on the variability in the environment?

Let the probability distribution of clients that salesmen confront each period be given by $b_t\in\{-b,b\}$. With a certain probability, the environment changes and the probability distribution of clients changes. In particular, let the period-by-period transition probability between states be π. Suppose, finally, that the cost of changing codes (which may be a time cost of adapting, a monetary cost, etc.) is γ and assume no discounting.

Since there are only two values for the state variable, and the value function is symmetric,there are only two possible policy functions: either agents change their code whenever the new environment does not match their code, or they keep the same code regardless. This considerably simplifies the problem, as it means we can actually compute the values for the two policies and find out the cut-offs that determine which policy agents pursue:

1. Persistent Code: In this case, and since there is no discounting, the optimal language is as if $b = 0$.The expected period by period cost is then given by $V(x_{t-1}, b_t) = D(0) = (1 - \frac{\lambda}{2})$,independent of b.

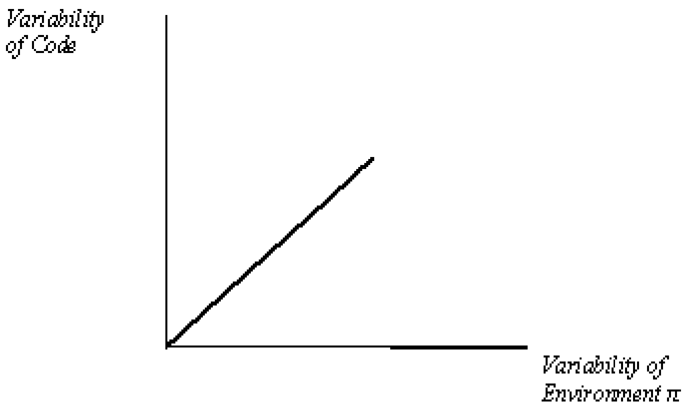
2. Variable Code: Agents change the code whenever the state changes, and that leads to a value function that is, period by period: $V(x_{t-1}, b_t) = D(b_t) + \pi\gamma$

Given this two policies, the following proposition is easy to prove:

Proposition 11 *The persistence of the optimal code is non-monotonic in the variability of the environment. In particular, there exists a value $\pi^* \in R$ such that for all $\pi < \pi^*$, the optimal code adapts to the environment, and is more variable the more variable the environment, and for $\pi > \pi^*$, the optimal code is persistent even in the presence of environmental changes.*

Proof. Define $\pi^* = (D(0) - D(b^*))/\gamma$.Compare the persistent and variable code period by period value functions. Then we have that whenever $D(b^*) + \pi\gamma < D(0)$,that is whenever $\pi < \pi^*$,the language changes every time the state changes, and the probability of change is π.Moreover, if $\pi > \pi^*$,the persistent policy function changes and ■

Figure 3 illustrates this proposition. Intuitively, there are two kinds of environments where we will observe little code variability. In a stable environment (with π low) organizations will have codes that are well adapted to the environment, but that change little because the environment changes little. On other hand, when the environment changes a lot (π high), they will have 'generic' codes, which being unadapted to the environment can be used regardless of the state of the world.


Variability
of Code
Variability of
Environment π