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

October 27, 2003

Abstract

We study the determination of specialized codes under bounded rationality, and its implications for organizations. Agents may decrease communication costs by designing codes that fit their own environment, using for example more precise words for more frequent events. Bounded rationality imposes sharply decreasing returns to scope, since when similarly skilled agents in different services must communicate with one another they must share common codes, which in turn degrades communication within each service. Thus the decision of whether to segregate services or integrate them trades off the synergies that result from better coordination between services against the loss due to the need for a common, more vague, code than the one that would optimize communication within services. Alternatively, more skilled 'translators' may be used to allow separate services to appropriate the synergies while keeping their own codes. A decrease in diagnosis costs leads to increasing integration among services and to the substitution of hierarchies for common codes, as common codes allow for the direct interaction among agents in different services. When adoption decisions are decentralized and non contractible, the common code will be inefficiently biased towards the needs of early adopters and there will be too little commonality of codes.

1

1 Introduction

The treatment of communication costs in economics is, at best, sketchy. Communication is generally deemed to consist in the incentive compatible revelation of some signal at no cost. In reality, agents appear to spend a large portion of their waking hours engaging in communication. Why is it costly to communicate with others? What are the implications of changes in communication costs for the organization of economic activity? How can we study these changes? In his classic book *The Limits of Organization* Arrow (1994) observed that organizations deal with the complexity of the environment by creating specialized codes. In this paper, we build on this observation to provide a theory of how agents establish specialized codes that respond to their own environment to communicate with each other, and how these codes constrain them and the way they organize their activities.

A natural language, such as English, is a general code: it is flexible, as it can be adapted to a wide range of situations. Subsets of agents dealing repeatedly with particular subsets of problems can design specialized codes that fit their special needs. These specialized codes reduce the cost of communicating information, by allowing the use of words that summarize complex information. Of course, agents are boundedly rational and cannot learn as many terms as problems they confront. Thus the problem of the code design is to allocate concepts or tasks to words under the constraints on rationality of agents.

Examples of such specialized codes in organizations are widespread. In some instances, organizations explicitly decide to create a code to facilitate communication. For example, when several firms come together in a common project, they usually create a Project Management Dictionary (Blankevoort 1986).[1] More generally, different accounting systems,[2] human resource databases and other organizational data bases are different codes, in the sense that they map differently the primitive objects in the environment (money

---

[1] One such case is the SEMATECH consortium, where 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 cultures the consortimum decided to "compile a dictionary of common techical terms and accronyms. Before this attempt at standardization, many firms prided themselves on having unique names for things." (Browning, Meyer and Shetler, 1995: 125).

[2] For example, the New York State Office of the Controler's manual mandates the coding of certain income flows into New York State as Taxes, of others others as Fees, of others as licenses, of others as Commissions. See the Accounting System User Procedures Manual of the Office of the Comptroller of NY State at: http://www.osc.state.ny.us/agencies/accmanual/actcodes/31180.htm.

flows in the accounting example) into words.[3]

Our analysis begins by proposing a simple model of a code. A code is a partition of the space of signals. Such partition is designed to achieve the maximum possible precision in communication, subject to the constraints imposed by the agents' bounded rationality. In particular, agents aim to communicate their information so as to minimize the extra effort involved in exactly figuring out what the agent communicating the message meant. In creating the code agents take into account the fact that some signals are generated more frequently than others.

We then characterize the solution of this problem. In particular, we show that the optimal code allocates precise words to frequent events and more vague words to unusual events, and that the more imprecise words are used less often, even though they make allusion to a wider array of events. We also show that a given code is more valuable the more unequal the distribution of events it makes reference to, since in that case the precision of the words can be more tightly linked to the characteristics of the environment.

Up to this point, our analysis focuses on the use of a common code to facilitate communication among agents facing the same distribution of events. In reality, agents who deal with different sets of tasks or problems must sometimes communicate with one another. A code that would be ideal for a salesman dealing with high end customers, i.e. one having words describing different sociological types of wealthy customers, is inadequate for a salesman dealing with low-end customers. If both agents have to communicate with a common superior, or if they must deal with each other or each other's customers, then a less specialized code may be preferred. In fact, as we show, bounded rationality implies that whenever two agents facing different sets of events must communicate with the same third agent (such as a superior, for example), and such agent is not more skilled than them, a common code will be chosen, rather than separate or partially common codes ('dialects').

The need for a common code thus introduces an important source of diminishing returns to scope in organizations. Having agents who deal with different activities communicate with each other reduces the precision of the specialized codes that agents are allowed to use in their activity. The organization thus faces a choice: it can group agents together improving coordination at the cost of a decrease in the precision of the code, or it can keep them separate to enjoy the benefits of separate codes.

In adopting a common code among different services which deal with

---

[3]Accounting scholars have long recognized that accounting is an information system, but the emphasis up to now has been (like in economics) on agency information costs, rather than on the coding costs (see Watts and Zimmerman, 1990) for a review

different events, the organization must trade-off the improved coordination between services that results from the common code against the degradation of the within-service communication that results. Common codes will be preferred when the between-services improvements offset the within-service losses. In particular, we show that if the two services face similar tasks, a common code is not too costly. Even when the two original codes are relatively different, the commonality will be justified if the synergies are sufficiently big. Also, the lower the diagnosis cost the more likely that the loss caused by receiving an imprecise signal is less important than the synergy gain that results from the improved ability to communicate.

Rather than incurring the loss of a common code, an organization may choose to hire a more skilled agent, that is, an agent able to acquire richer codes, and utilize her as a specialized 'translator.' While this decreases the communication loss required to acquire the integration synergies among the different units, it incurs the cost of an extra agent and some extra information costs.

Under which circumnstances is each of these three simple organizational arrangements (separate, non-communicating units that cannot capture across-unit synergies; units with their own codes that communicate through a specialized translator; and units sharing a common code and able to communicate horizontally without translator) preferred? As information costs decrease, we expect to observe two types of changes: first, an increase in integration among previously non-communicating units; second, within units that were communicating hierarchically, a move towards a more centralized system of information sharing, in the form of a centralized code, together with more decentralized communication. The reason for this second prediction is that a reduction in diagnosis costs, i.e. a reduction in the cost of identifying a problem and matching it with its solution, increases the value of a common code. The introduction of a common code reduces the 'translation' role of hierarchy, by facilitating 'horizontal' communication. Thus we expect to observe, following reductions in information costs, the substitution of hierarchies for common codes and horizontal communcation.

After studying the efficiency aspects of the choice of common codes, we discuss the constrained adoption decisions that result when agents independently choose their codes. We show that if the adoption decisions are sequential, common codes are chosen when they are optimal, but they are inefficiently biased towards the needs of those agents who adopt first. Codes can also be considered as specific investments, as a consequence, when services can separately choose their codes and adoption costs are non contractible, there exists too little commonality of codes and too much fragmentation. Communication between different organizations is (inefficiently) worse than

within each given organization, as organizations do not internalize the benefits that adopting a common code brings to other organizations. This result provides a rationale for the finding[4] that coordination tends to be better within a particular organization than between organizations, even when the objective content of the jobs of those coordinating is similar.

We conclude the paper by studying the empirical evidence on the theory in two ways. First, we look at some systematic evidence on the impact of the drop in information costs on the internal organization of firms. Second, we present two detailed case studies, which aim to provide some detailed insights on the link between information costs and common codes, the link between common code and decentralization and finally the conflicts generated by decentralized code adoption.

The paper is structured as follows. Section 2 introduces the model of coding that will be used in the rest of the paper. Section 3 focuses on coding when the organization is considered in isolation. We begin by providing a general characterization of optimal codes when there are only two agents. We then discuss how the value of a code depends on the type of problem the organization faces and we analyze the marginal benefit of increasing the complexity of the code. We then extend the analysis to situations in which there are more than two agents. Section 4 considers multiple organizations and asks whether the two organizations should have common or separate codes. We also consider the benefit of integrating the two organizations. Section 5 introduces strategic consideration in code adoption. We show that in a non-cooperative equilibrium sequential code adoption leads to a first-mover advantage and to inefficient distortions. We also discuss the possibility that the codes that are arise in equilibrium are inefficiently heterogeneous. Section 6 discusses the empirical evidence.Section 7 reviews the existing literature and concludes.

2   Codes and bounded rationality

Agents can improve the processing of information among them by designing specialized codes. Of course, as Arrow (1974) points out, coding does not avoid diminishing returns: since agents are boundedly rational they can only learn a limited code. To capture this idea, we consider agents who can only deal with a maximum number of words.

---

[4]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.

Consider an agent who receives a signal $x \in X$ and must communicate it to another agent. For simplicity, assume that the set of signals X is finite and that every signal x has a strictly positive probability f_x of occurring.

Agents are grouped in services. A *service* is a group of agents dealing with the same distribution of task f_x. To fix ideas, suppose the tasks are client types, the individual who drew the task is a salesman, and the individual to whom the type must be communicated an engineer. A code[5] $\mathcal{C}$ is a partition of the type space X into K disjoint subsets: $W_1, \ldots, W_K$.[6] A particular $k \in \{0, \ldots, K\}$ can be thought of as a *word* and the corresponding subset W_k as its meaning.

Definition 1 *A code is a partition of the set of signals X.*

Words that are vague, in the sense that could be transmitted whenever one of a wide set of events took place, communicate little information. To fix ideas, think of the salesman communicating the geographical position of the customers in the city for the engineer to place a visit. If his code is very coarse, so that the words are vague, once the word is communicated ('the customer is in the North') the engineer who receives it must spend a lot of time searching for the client. The search time depends on the size of area of the city that is covered by this word.

More generally, beyond the geographical interpretation, if the engineer receives a coarse message, he must spend a lot of time refining his understanding of the problem, i.e. diagnosing the problem or processing the information on the problem. This *diagnosis cost* is higher the more imprecise is the word.[7] In particular, we assume throughout that, as the geographic example suggest, the diagnosis cost is proportional to the number of underlining events that are referred by the word.

Diagnosis costs incurred depend on what event is drawn and which code is used. Suppose x is realized and $\mathcal{C}$ is such that $x \in W_k$. The salesman then transmits to the engineer the word W_k. After that, the engineer must further diagnose the client need. Since the diagnosis cost is linear in the number of events in that set, the diagnosis cost of x in a particular code $\mathcal{C}$ is

$$d(x, \mathcal{C}) = \gamma \times \sharp W_k,$$

---

[5] We use the word 'code' rather than language since grammar plays no role in this problem.

[6] As will be clear, the fact that the subsets are disjoints could trivially be derived from first principles.

[7] 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.

where $\sharp$ denotes the number of elements of a set. The expected cost of diagnosis in code $\mathcal{C}$ is therefore

$$D(\mathcal{C}) = \sum_{x \in X} f_x d(x, \mathcal{C}).$$

We will use a very simple definition of the bounded rationality of the agents: they can learn at most $K \geq 2$ words; on the other hand, there will be no cost in increasing the number of words that the agents know as long as this constraint is satisfied. Therefore, there will be no possible trade-off between diagnosis costs and the richness of the language, an interesting topic that we leave for future research.

To summarize, a code is optimal if it minimizes the expected diagnosis cost subject to the constraint that each agent knows no more than K words. In the next section we derive some properties of optimal codes.

3 Communication within a Service

3.1 The structure of the optimal code

We begin by studying the optimal code when one agent needs to communicate with a single other agent. The problem can be rewritten in a useful form by introducing three additional pieces of notation. For every k, let $n_k = \frac{\sharp W_k}{N}$, where N is the cardinality of X, and let $p_k = \sum_{x \in W_k} f_x$. Therefore, n_k can be interpreted as the *breadth* of word k, that is the number of events that are described by k, whereas p_k can be seen as the *familiarity* of word k, which is the probability that the event belongs to W_k. For example, if X is the set of meteorological events that occur in the Netherlands, the word "drizzle" is narrow (because it defines a very specific phenomenon) and familiar (because it occurs all the time), "bad weather" is broad and familiar, "good weather" is broad and (relatively) unfamiliar, and "hurricane" is both narrow and unfamiliar.

With the new notation, the diagnosis cost of event x becomes $d(x, \mathcal{C}) = \gamma n_k N$ and the expected diagnosis cost of code $\mathcal{C}$ is

$$D(\mathcal{C}) = \sum_k p_k \gamma n_k N = N\gamma \sum_k p_k n_k.$$

We use n and p to denote the respective vectors. Let $A(n)$ be the set of all p that are possible given n. Formally,

$$A(n) = \{p \in [0,1]^K \mid \exists \{W_1, \dots W_K\}, \frac{\sharp W_k}{N} = n_k, \sum_{x \in W_k} f_x = p_k\}.$$

As the number of codes is finite $A(n)$ is a finite set for every n.

We can redefine the objective function as $D(\mathcal{C}) = \sum_k p_k n_k$. Hence, the optimal code problem becomes:

$$\min_{p,n} \sum_k p_k n_k \qquad (1)$$

$$\text{s. t.} \begin{cases} p \in A(n); \\ n_k \in \{0, \frac{1}{K}, \ldots, 1\} \quad \text{for all } k; \\ \sum_k n_k = 1. \end{cases}$$

This is an integer problem, and in general it is difficult to characterize the solution completely. However, we can show that the solution has two important properties. In order to describe them, we introduce another piece of notation: for any event x and any code $\mathcal{C}$ let

$$p(x, \mathcal{C}) = p_{k:x \in W_k}.$$

Given an event x, $p(x, \mathcal{C})$ is the familiarity of the word of code $\mathcal{C}$ that includes event x. When there is not ambiguity, we will drop the code from the notation and simply write $p(x)$.

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

Proof. In the solution (p, n), choose n_k and $n_{k'}$ such that $n_k > n_{k'}$ and choose $x \in W_k$ and $x' \in W_{k'}$. Exchanging x and x' must (weakly) increase the diagnosis cost. Therefore,

$$n_k p(x') + n_{k'} p(x) - (n_k p(x) + n_{k'} p(x')) = (n_k - n_{k'})(p(x') - p(x)) \geq 0.$$

Hence, $p(x') \geq p(x)$. ∎

Proposition 1 is a consequence of the fact that the objective function in (1) is linear in p given n. Hence, if we hold the breadth of each single word fixed, the best thing we can do to reduce expected diagnosis time is to put the frequent events into narrow words and the rare ones into broad words. See figure 1 for a graphical illustration of the argument.

Proposition 1 also suggests that the problem of finding the optimal code can be separated into two steps.

1. Attribution of Meaning: For each possible n, re-order k in such a way that $n_1 \leq n_2 \leq \cdots \leq n_K$. Put the n_1 most frequent events in W_1; put the next n_2 frequent events in W_2; and so on up to the least frequent N_K event that go into W_K. This yields an expected diagnosis cost $\tilde{d}(n)$, which, by Proposition 1, is the lowest we can attain given n.
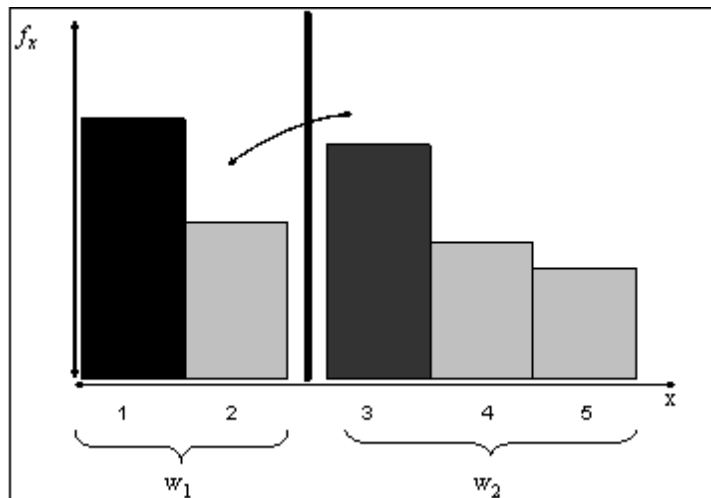
Figure 1: *Suppose code has one word for 1 and 2 and one word for 3, 4, and 5. Swap events 2 and 3. Diagnosing 2 becomes more expensive, but diagnosing 3 becomes cheaper. Because 3 is more frequent, the code is more efficient.*

2. Choice of Word Breadth: Compare $\tilde{d}(n)$ for all possible n and find the highest.

This algorithm for finding optimal codes turns out to be extremely useful, and it is applied throughout the rest of the paper.

Step 1 of the algorithm is fully characterized by Proposition 1. Step 2 is less simple because it involves integer programming. We have a partial characterization:[8]

Proposition 2 *Unless integer constraints make it impossible, in an optimal code broader words are less familiar. Formally, if $n_k \geq n_{k'}$, then $p_{k'} + f_{\tilde{x}} \geq p_k - f_{\tilde{x}}$ where $f_{\tilde{x}}$ is the lowest probability event in W_k. Furthermore, if $n_k N \geq n_{k'} N + 2$, then $p_{k'} \geq p_k$.*

Proof. As the code is optimal by assumption, transferring word $\tilde{x}$ from W_k to $W_{k'}$ cannot lower costs. Hence, we must have

$$\left(n_k - \frac{1}{N} \right) (p_k - f_{\tilde{x}}) + \left(n_{k'} + \frac{1}{N} \right) (p_{k'} + f_{\tilde{x}}) \geq n_k p_k + n_{k'} p_{k'}$$

---

[8] Although Proposition 1 and Proposition 2 are closely related, neither of the two implies the other directly.
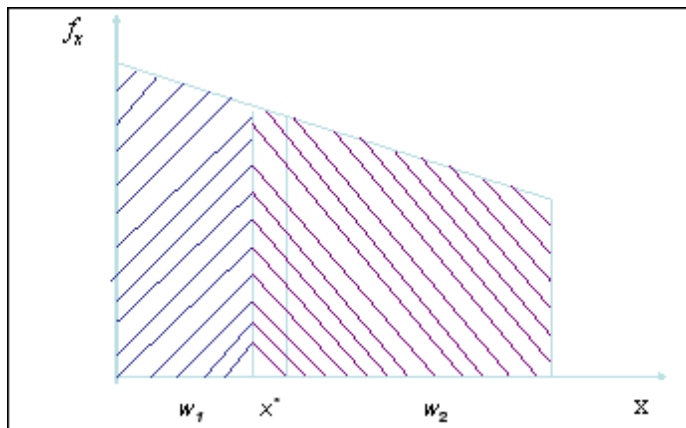
Figure 2: *Transferring x^* from w_2 to w_1 decreases diagnoisis costs.*

This inequality rewrites as

$$\frac{1}{N}\left[(p_{k'} + f_{\tilde{x}}) - (p_k - f_{\tilde{x}})\right] + f_{\tilde{x}}(n_{k'} - n_k) \geq 0, \qquad (2)$$

which proves the first statement in the proposition.

To prove the second statement, rewrite inequality (2) as

$$(p_{k'} - p_k) + f_{\tilde{x}}\left(n_{k'} - n_k + \frac{2}{N}\right) \geq 0.$$

∎

To understand Proposition 2, suppose there are a large number of events with infinitesimal probability (See Figure 2). The first part of the proposition then says that broader words are less familiar: if $n_k \geq n_{k'}$, then $p_{k'} \geq p_k$. To see this, consider the costs and benefits of transferring an infinitesimal event x^* from a broad word to a narrower word. Now, the event x^* is captured by a narrower word, and this is a certain benefit. However, the broad word is now less broad and the narrow word is less narrow. This is a benefit if the broad word is more familiar. But that would create a contradiction because the initial code would be suboptimal. Hence, a broad word must be less familiar.

The intuition above is based on a marginal argument. To complete the argument, we must account for the presence of integer constraints. There may exist words that are both broader and (slightly) more familiar than others because they only contain non-infinitesimal words which – if moved – would make another word both broader and more familiar. The last part of the proposition puts an upper bound to the importance of integer constraints.

10

If word k contains at least two events more than word k', then k must be less familiar than k'.

3.2 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 is "unequal".

To give a precise meaning to distributions inequality, take a distribution p and assume without loss of generality that p is ordered in a nondecreasing way: if $x' < x''$, $p(x') < p(x'')$. We say that distribution $\tilde{p}$ is more *unequal* than p if for every x $\tilde{P}(x) \geq P(x)$.[9] Intuitively, a more unequal distribution is one that puts even more probability on events that were already likely to happen. With this definition, we can show that communications cost is decreasing in inequality:

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

Proof. Let $\mathcal{C}$ be the optimal code for distribution p. Use the same code for distribution $\tilde{p}$. The cost for distribution p is $\sum_k p_k n_k$ while the cost for $\tilde{p}$ is $\sum_k \tilde{p}_k n_k$. By Proposition 1, word size is nondecreasing in k. Then,

$$
\begin{aligned}
\sum_k p_k n_k &= P_1 n_1 + (P_2 - P_1) n_2 + ... + (P_{k-1} - P_{k-2}) n_{k-1} + (1 - P_{k-1}) n_k \\
&= P_1 (n_1 - n_2) + P_2 (n_2 - n_3) + ... + P_{k-1} (n_{k-1} - n_k) + n_k \\
&\geq \tilde{P}_1 (n_1 - n_2) + \tilde{P}_2 (n_2 - n_3) + ... + \tilde{P}_{k-1} (n_{k-1} - n_k) + n_k \\
&= \sum_k \tilde{p}_k n_k.
\end{aligned}
$$

As $\sum_k \tilde{p}_k n_k$ is not lower than the minimal diagnosis cost for $\tilde{p}$, the statement is proven. ∎

To understand the proposition, see figure 3. 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 cost, 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.

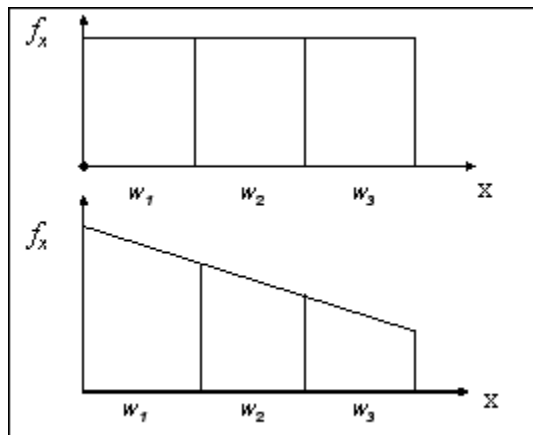[9]The new distribution $\tilde{p}$ need not be nondecreasing in x

Figure 3: *The suboptimal equal sized word code in the panel below achieves the same cost (1/3) as the code in the top panel. Since the words are not optimal below, the lmore 'unequal' distribution generates less costly communication.*

The following is an immediate consequence of the proposition above:

Corollary 1 *If distribution $\tilde{p}$ is more unequal than distribution p, then moving from 0 to k words is more valuable in $\tilde{p}$. On the other hand moving from k words to ∞ words (perfect communication) is more valuable in p.*

Proof. The first part of the argument follows from the previous proposition. The search costs with 1 word are 1 for both distributions. The search cost is lower for the more unequal one for any n=2,. by the previous proposition, thus the value of any given language of the first word is higher in the more unequal one. On the other hand, in the limit diagnosis costs are equal in both languages (0 in both), thus adding a sufficiently large number of words is more valuable in the more equal language. ∎

Two elements affect 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 information necessary. Which effect dominates? While there is no general answer, the corollary shows that at least in the beginning adding a new word is better for an unequal environment, while if the language is already extremely rich, the marginal benefit is higher for a more equal environment.

12

3.3 Code commonality among similarly skilled agents

Consider now a case in which between service communication is needed. In particular, suppose that the salesman from region A and the salesman from region B must communicate with the engineer e. In this case, the two salesmen may use the same code, completely different codes, or they may use 'dialects', that is codes with some common words and some different words that refer specifically to the events that each one confronts. When is e going to use the same code to get information from the two agent types?

The trade-off between a common code and different codes or dialects is as follows: when the same code is used, the precision of each salesman information diagnosis goes down. Thus tailoring a code for each type of agent may make communication more precise, as the codes are specialized to the specific density of events confronted. However, the precision of the words they can transmit is sharply limited by the fact that the engineer must learn both codes.

Given the strict constraints on bounded rationality that we have assumed, it can be shown, as the next proposition states formally, that the same code will be used by both services. Intuitively, if the engineer must in any case incur the costs of learning the extra word to communicate with the alternative service, a more precise word that comes from the intersection of the one used with one service and the one used with the other service will always improve communication.

Proposition 4 *Only a common code can be efficient*

Proof. We will show that a code that is not entirely common cannot be efficient because it is strictly dominated by another code.

Suppose that the code $\mathcal{C}_A$ that e 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[10]. Suppose without loss of generality that $W_k \in \mathcal{C}_A$. Transform $\mathcal{C}_B$ into $\tilde{\mathcal{C}}_B$ as follows. Make W_k a word of $\tilde{\mathcal{C}}_B$. Reduce all the words that contained what are now elements of W_k. That is, $W \in \tilde{\mathcal{C}}_B$ if and only if $W \in W'/(W' \cap W_k)$ for some $W' \in \mathcal{C}_B$ or $W = W_k$.

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

---

[10]That is k is an element of $\text{argmin}_k \#W_{\tilde{k}}$ subject to $W_{\tilde{k}} \in \mathcal{C}_1 \cup \mathcal{C}_2$ and $W_{\tilde{k}} \notin \mathcal{C}_1 \cap \mathcal{C}_2$.

narrower word in $\tilde{\mathcal{C}}_B$ than it was in $\mathcal{C}_B$. The new code is strictly more efficient than the older. ∎

These three examples 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: $\hat{\mathcal{C}} = \{\{1,4\}, \{2,3\}, \{5,6\}, \{2,5\}, \{3,6\}\}$.

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

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

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

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

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.

Corollary 2 *If both salesmen send the same number of messages to the engineer, then propositions 1 and 2 apply as stated if one lets $\tilde{f}_x = \frac{1}{2}(f_x + g_x)$.*

Note that an essential aspect of the proof is that all agents are similarly bounded. An alternative that we will consider later on is for an organization to avoid the need for common codes by hiring agents who may learn more words and who use this skill to 'translate' among different sets of agents.

3.4   Codes with two words

In the rest of the paper, we use our characterization of the optimal codes to discuss organizational choices when communication is important. To simplify our analysis, we restrict our attention to two-word codes. Doing this does still allow us to study issues such as the commonality of codes, since we know from the previous section that agents choose either fully common codes (when those in A and B communicate with one another) or separate codes (when they do not). For simplicity, we assume that there is a continuous set of events.

Suppose that a salesman deals with consumers $x \in [0,1]$ drawn from a distributions with density $f(x) = (1-b) + 2bx$ with $b \in [-1,1]$ and must transmit his information about the characteristics/identity of the customer to an engineer using a two word code. The distribution is given of x is given by

$$F(x) = (1-b)x + bx^2 \qquad \text{with } b \in [-1,1]$$

The optimal two-word code is

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

with first order conditions

$$f(x)\, x - f(x)(1-x) + F(x) - (1 - F(x)) = 0$$

The solution is[11]

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

and the optimal diagnostic costs are[12]

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

In what follows, we start to 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 next. We have presented the model in this section as a model of communications between two agents. It is obvious that all the results still hold true if there are multiple agents that need to communicate with each other, as long as all the agents who need to transmit information to others face the same probability of events.

---

[11]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 $(\hat{x}, 1]$.

[12]The function $D^*(b)$ is a concave symmetric function, which reaches its maximum, equal to 0.5 at $b = 0$.

4 Codes and Organization

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 optimal. For expositional reasons, it is best to focus first on the comparison between the two pure forms (1) and (2), and then introduce the third form

4.1 Integration or separation?

In order 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,

$$
\begin{aligned}
F_A(x) &= (1 - b)\, x + bx^2, \\
F_B(x) &= (1 + b)x - bx^2.
\end{aligned}
$$

with $b \in [-1, 1]$ measuring the similarity between the two distributions. The respective densities are

$$
\begin{aligned}
f_A(x) &= (1 - b) + 2bx, \\
f_B(x) &= (1 + b) - 2bx.
\end{aligned}
$$

Each engineer has the ability to attend to the needs of at most one client. Salesmen bring sales leads randomly to each engineer. The arrival process is as follows (see Figure 4):

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

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.
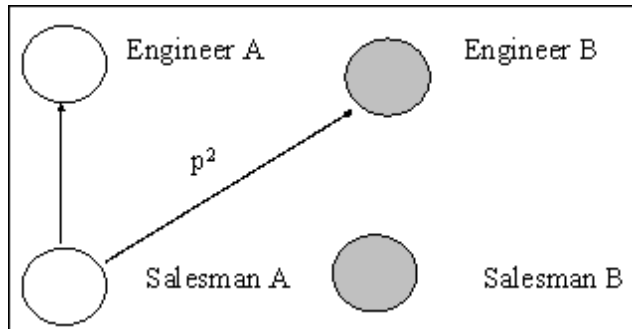

Figure 4: *Synergies exist when there is excess demand on one service and excess capacity in the other.*

We study the two possible organizations of these services: one where the salesmen from service A only communicate sales leads to engineers in A; a second, where a salesman from A may communicate sales leads to either engineer. 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. 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. As we know from section 2, this requires in turn the use of a common code.

What are the diagnosis costs in this case? To obtain the common language, we use Corollary 2, which says that the common language is the one that would be chosen when the density of tasks is the average of the two. In this case, the average problem density

$$f_x = 1/2((1 - b) + 2bx) + 1/2((1 + b) - 2bx) = 1,$$

i.e. a uniform density. The optimal code in this case has two equally imprecise words, with each word identifying the sales lead as coming from one half of the distribution, $x^* = 1/2$. Let the per-client diagnosis costs be $\lambda \in (1, 2)$ to ensure positive profits, and the output that can be obtained when a client's

17

problem is solved be 1. The total benefits attained by both agents in this case is:

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

Next examine a *separated organization*. The two services use different codes. Service A selects a code with cut-off

$$x_A^* = \frac{3b - 2 + (4 + 3b^2)^{1/2}}{6b};$$

and by symmetry service B adopts $x_B^* = 1 - x_A^*$. The expected diagnosis cost in either service is

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

Note that because we have assumed $\lambda > 1$, it will not be profitable for the salesman of service A to send a problem to the engineer of service B in the absence of a common code, as the diagnosis cost will be higher than the revenue that can be generated.

In either service, the expected profit is

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

where $D^*(b)$ is as defined above.

Should the organization be separated or integrated? The choice depends on the inequality:

$$2(1 - p(1 - p))(1 - \frac{\lambda}{2}) \geq 2(1 - p)(1 - \lambda D^*(b)) \tag{4}$$

The following proposition[13] characterizes this choice:

Proposition 5 *An integrated organization is superior to a separated organization if the coordination gain is larger than the communication loss, that is:*

$$\frac{1 - p(1 - p)}{1 - p} > \frac{1 - \lambda D^*(b)}{1 - \lambda/2} \tag{5}$$

Therefore, an integrated organization is more advantageous when:

- *the synergy parameter p increases;*

---

[13]It is easy to check that there exist parameter values that lead to each one of these choices.

- *the diagnosis cost λ decreases;*

- *the underlying distribution of tasks becomes less unequal (b becomes closer to 0).*

Proof. To prove the first past, just rearrange the first part rearrange (4). For the comparative statics define[14]

$$V(b, \lambda, p) = 2(1 - p(1 - p))(1 - \frac{\lambda}{2}) - 2(1 - p)(1 - \lambda D^*(b)) \qquad (6)$$

Taking derivatives of this expression, we obtain

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

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 $\lambda < 2$ is required for positive profits.

Finally,

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

since $\partial D^*(b)/\partial b < 0$. ∎

The role of synergy is clear. The higher the probability that the two services benefit from communicating, the greater the advantage of being able to communicate. Instead, the diagnosis cost operates through a different channel. The lower λ, the less important it is to use the most appropriate code. This reduces the loss in terms of communication cost that occurs when a common code is adopted and it leads to an integrated organization. Finally, also the shape of the distribution affects the reduction in communication cost due to separation. This reduction is large when the distribution is unequal. Conversely, shifting to a common code is least costly when the distribution is flat $(b = 0)$.

---

[14]One should still check that one of the two organizations is not better than the other for all values of the parameters. This can be seen for instance by setting $\lambda = 1.5$ and $p = 0.25$. Then, V considered as a function of b is a concave function, which is positive on $(-0.684, 0.684)$ and negative outside this interval.

4.2   When is a hierarchy useful?

We know consider the third organization form, which is more complex than the other two because it includes 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 will communicate to the translator the type of the 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.

There is fixed cost μ of hiring the translator. However, we assume that the translator is faster than regular engineers when it comes to diagnosis, since when he gets a 'small word' he can translate it without search. For simplicity we make the extreme assumption that the translator's λ is zero. The qualitative results here go through even if the cost is positive as long as it is lower than those of the engineers.

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

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

Proof. The expected payoff with translation is

$$
\pi_{\text{translate}} = 2(1-p)(1 - \lambda D^*_{\text{translate}}(b)) + 2p^2 \left(1 - \lambda \tilde{D}^*_{\text{translate}}(b)\right) - \mu,
$$

where

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

and

$$
x^*_{\text{translate}} = \arg\min_x 2(1-p)\left(F\left(x\right) x + \left(1 - F\left(x\right)\right)\left(1 - x\right)\right) + 2p^2 \left(F\left(x\right)\left(1 - x\right) + \left(1 - F\left(x\right)\right) x\right).
$$

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

$$
\begin{aligned}
\pi_{\text{integrated}} &= 2(1 - p(1-p))\left(1 - \frac{\lambda}{2}\right) \\
\pi_{\text{separate}} &= 2(1-p)(1 - \lambda D^*(b))
\end{aligned}
$$

For a given b and p, let

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

This is the solution of

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

and it is the value of λ for which $\pi_{\text{integrated}} = \pi_{\text{separate}}$.

A sufficient condition for $\lambda^* \in [1, 2]$ is that $p \geq 0.213$. To see this, note that

$$D^*(b) \in \left[\frac{22}{27} - \frac{7}{54}\sqrt{7}, \frac{1}{2}\right].$$

This means that

$$\lambda^* \in \left[2\frac{p^2}{p^2 + \left(1 - 2(\frac{22}{27} - \frac{7}{54}\sqrt{7})\right)(1 - p)}, 2\frac{p^2}{p^2 + \left(1 - 2\frac{1}{2}\right)(1 - p)}\right]$$

But

$$2\frac{p^2}{p^2 + \left(1 - 2\frac{1}{2}\right)(1 - p)} = 2$$

and

$$2\frac{p^2}{p^2 + \left(1 - 2(\frac{22}{27} - \frac{7}{54}\sqrt{7})\right)(1 - p)} \geq 1$$

if and only

$$p \geq -\frac{7}{54}\sqrt{7} + \frac{17}{54} + \frac{1}{54}\sqrt{\left(-1204 + 518\sqrt{7}\right)} \simeq 0.21$$

Now define

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

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

$$\pi_{\text{translate}} = \pi_{\text{integrated}} = \pi_{\text{separate}}.$$

If $\lambda = \lambda^*$ and $\mu < \mu^*$, 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 $\pi_{\text{translate}}$, $\pi_{\text{integrated}}$, and π_{separate} 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. ∎

To understand the intuition for this proposition, refer to Figure 5 and begin with the comparison between separation and translation. The latter has a fixed cost μ but makes inter-service communication possible. The net benefit is given by the probability of getting extra business minus the cost of extra communication minus the fixed cost of hiring a translator. If the diagnosis cost λ is high, the cost of extra communication is high and the net benefit is likely to be low. So, translation is more likely to beat separation when λ is low.
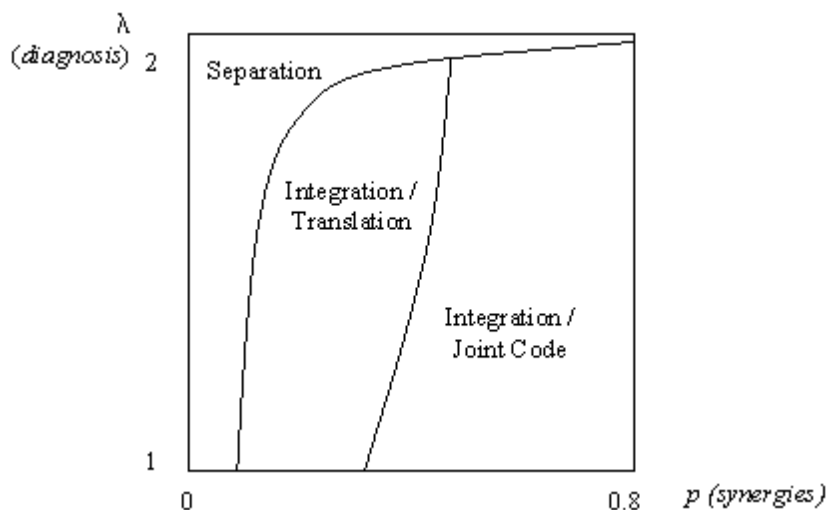

Figure 5: *Numerical simulation of optimal organizational structure and code design.* $b=0.4$ *and* $\mu = 0.006$.

Instead, translation is better than integration when the diagnosis cost λ is high. This is because translation saves on communication cost by allowing services to keep efficient service-specific codes. These savings are more important when λ is high.

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.

5 Strategic Code Adoption

The previous analysis abstracted from strategic considerations. We implicitly assumed that codes are adopted in order to maximize total surplus. This section's goal is to consider the effect of conflict of interest in the choice of organizational codes.

Obviously, if there are complete contracts, the presence of multiple players would not prevent them from agreeing to select the surplus-maximizing code and then make appropriate side payments. Instead we assume that code adoption is non-contractible. Firms cannot sign contracts that commit them to adopting a particular code because the outcome is difficult to verify. Outsiders cannot check that a firm is indeed using a certain code for internal communication unless they are given full access to the firm (which no firm would agree to do).

We first examine sequential code adoption. We ask what are the incentives for the firm that moves first. Knowing that its decision affect other firms's code policies, what kind of code should the first mover choose? We then analyze how free-riding affects code adoption. We start from a situation in which firms have different codes but could adopt common codes if they incur a fixed cost. Although adopting a common code is efficient, we show that the presence of externalities may prevent the move to a common code.

5.1   First-mover bias

As in the previous section, we assume that there are two services A and B. However, now we identify the two services with two separate profit maximizing firms: salesman A and engineer A belong to firm A while the other two agents make up firm B. Timing is sequential. First, firm A adopts a code. Then, firm B observes the code adopted by A and selects its own code.

The advantage of joint codes is that firms can "trade". We then need to model contracting between firms, and in particular bargaining power. We assume that, when a salesman of firm i offers a problem to an engineer of firm j, the surplus that is created by the relationship goes in proportion a to the salesman and $1 - a$ to the engineer (this is equivalent to assuming that the salesman makes a take-it-or-leave-it offer with probability a). When salesman i has one customer, he communicates only with his engineer. When he has two customers, he will offer one of the two customers to the engineer of the other firm. The other engineer accepts if she has not received a customer from her salesman.

For simplicity, and as in the previous section, we assume that the two firms have symmetric distribution functions $F_B(x) = 1 - F_A(1 - x)$. The associated diagnosis cost for a customer drawn from salesman i is denoted with $s_i(x)$:

$$s_A(x) = F_A(x) x + (1 - F_A(x))(1 - x)$$
$$s_B(x) = F_B(x) x + (1 - F_B(x))(1 - x)$$
$$= (1 - F_A(1 - x)) x + F_A(1 - x)(1 - x)$$

The expected payoff in a separate code is the same for each of the two firms:

$$(1 - p)(1 - \lambda s_A(x^s))$$

where

$$s_A(x^s) = \arg\min_x s_A(x).$$

With a joint code, the payoffs for A are then

$$\begin{cases} 1 - \lambda s_A(x) & \text{with probability } 1 - 2p + p(1-p) \\ (1+a)(1 - \lambda s_A(x)) & \text{with probability } p^2 \\ (1-a)(1 - \lambda s_B(x)) & \text{with probability } p^2 \\ 0 & \text{otherwise} \end{cases}$$

while the payoffs for 2 are

$$\begin{cases} 1 - \lambda s_B(x) & \text{with probability } 1 - 2p + p(1-p) \\ (1+a)(1 - \lambda s_B(x)) & \text{with probability } p^2 \\ (1-a)(1 - \lambda s_A(x)) & \text{with probability } p^2 \\ 0 & \text{otherwise} \end{cases}$$

The expected payoff in a joint code with x for the two firms is respectively

$$\pi_A^J(x) = 1 - p + p^2 - \lambda\left(\left(1 - p + p^2 a\right) s_A(x) + p^2(1-a) s_B(x)\right)$$
$$\pi_B^J(x) = 1 - p + p^2 - \lambda\left(\left(1 - p + p^2 a\right) s_B(x) + p^2(1-a) s_A(x)\right)$$

The first mover, firm A, solves the following problem

$$\max\left(\pi^S, \pi_A^J(x)\right)$$
$$\text{subject to } \pi_B^J(x) \geq \pi^S.$$

In the solution, the participation constraint may or may not be binding. Either $\hat{x}$ is the solution to the unconstrained maximization problem or it is the value such that the participation constraint is binding: $\pi_B^J(x) = \pi^S$. In the former case we already know that $\hat{x} > \frac{1}{2}$. In the latter case, note that if $\pi_B^J\left(\frac{1}{2}\right) < \pi^S$ a joint code cannot be optimal. Then assume that $\pi_B^J\left(\frac{1}{2}\right) \geq \pi^S$. But then there is an $x > \frac{1}{2}$ such that $\pi_B^J(x) = \pi^S$. Thus, whether the participation constraint is binding or not, $\hat{x} > \frac{1}{2}$. In other words, a joint code will be adopted, but it will be biased:

Proposition 7 *If a joint code is strictly superior to a separate code, then a joint code is adopted, and the first mover will adapt its code so that it fits better the needs of the second mover than the code it would have chosen in isolation.*

Proof. To see the first part of the proposition, assume that a joint code is strictly superior to a separate code form an efficiency point of view. This means that

$$\pi_A^J \left(\frac{1}{2}\right) = \pi_B^J \left(\frac{1}{2}\right) > \pi^S.$$

But then, if firm 1 sets $x = \frac{1}{2}$, the participation constraint of firm 2 is satisfied and firm 1's expected payoff increases. If a joint code is strictly superior, a joint code is adopted.

For the second part, assume that a joint code is strictly superior. Note that

$$\frac{d}{dx}\pi_A^J(x) = -\lambda \left(\left(1 - p + p^2 a\right) s_A'(x) + p^2 (1 - a) s_B'(x)\right)$$
$$= -\lambda \left(\left(1 - p + p^2 (2a - 1)\right) s_A'(x) + p^2 (1 - a) \left(s_A'(x) + s_B'(x)\right)\right)$$

By symmetry,

$$s_A'(x) + s_B'(x) \le 0$$

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

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

Hence, ,

$$\text{if } x \le \frac{1}{2} \qquad \frac{d}{dx}\pi_A^J(x) > 0. \tag{7}$$

If the participation constraint is not binding, firm 1 faces an unconstrained maximization problem over $\pi_A^J(x)$. By (7), 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 $\pi_B^J(x) \ge \pi^S$. But, by (7), this implies that the optimal x is to the right of $\frac{1}{2}$. ∎

Because code choice is non-contractible, the first mover only takes 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.

5.2   Free-riding and excessive code differentiation

Let us begin from a situation in which firms are endowed with separate codes. A firm that wants to switch to a different code must sustain a fixed cost c. Suppose that the environment changes and it is now efficient to have a common code (even considering the switching cost). However, switching costs are 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 code; one firm adopts the code of the other firm; or both firms adopt a joint code. Suppose we are in a situation in which the efficient solution is for one firm to adopt the other firm's code (which occurs for intermediate values of c).

Let us denote the situation in which firm B adopts the code of firm A with the superscript SJ (as in "semi-joint"). The expected payoffs are respectively

$$\pi_A^{SJ} = 1 - p + p^2 - \lambda \left(\left(1 - p + p^2 a\right) s_A \left(x_A^S\right) + p^2 \left(1 - a\right) s_B \left(x_A^S\right) \right)$$
$$\pi_B^{SJ} = 1 - p + p^2 - \lambda \left(\left(1 - p + p^2 a\right) s_B \left(x_A^S\right) + p^2 \left(1 - a\right) s_A \left(x_A^S\right) \right)$$

From an efficiency point of view, SJ is optimal when

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

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. We shall now examine adoption in a non-cooperative environment. Firm 2 switches to firm 1'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 2 is willing to adopt it.

Proposition 8 *If there is a cost c of adoption, there are circumstances in which firms keep separate codes when it would be more efficient for one firm to switch to the other firm's code.*

It is interesting to note that this result is still true if 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. The firm that is supposed to switch code would generate a non-contractible positive externality to the other firm. In certain circumstances, firms keep separate codes when it is optimal for one firm to adopt the other firm's code.

6 Centralized Code, Decentralized Communication?: Evidence

Our model makes several predictions about the organizational implications of codes. We will focus our analysis on the following three predictions, mostly related to the effects of a decrease in the cost of identifying a problem and matching it with its solution, the diagnosis costs:

- A reduction in the diagnosis costs leads to increasing the links across firms and within previously separate units in a firm by facilitating the adoption of integrating mechanisms such as hierarchies and common codes.

- Decreases in diagnosis costs in an organization already integrated reduce the 'translation' role of hierarchy, by facilitating 'horizontal' communication – the substitution of codes for hierarchies.

- When agents adoption decisions are decentralized and adoption costs non-contractible, we expect to find too little code commonality, as each individual only enjoys part of the benefits that can be derived from a move towards unifying codes; moreover, codes will be in this context inefficiently biased towards the needs of early adopters.

The next subsections explore the evidence on these issues in two ways. First, we explore the evidence on the correlation between the drop in information costs and decentralization. Although the patterns in this evidence seem to go in a similar direction as the ones in the theory, the evidence is too vague about the causal link to either reject or confirm the theory. For this reason, we also present two detailed case studies, which aim to provide some detailed insights on the link between information costs and common codes, the link between common code and decentralization and finally the conflicts generated by decentralized code adoption.

6.1 Information Technology and Decentralization

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 where defined in different ways and could not be aggregated in a simple way. For example the database company Oracle had 70 incompatible databases for its human-resources department.

This incompatibility of the codes made it impossible to answer simple queries, such as how many employees were working at any time at the company.[15]

This state of affairs started to change during the mid 90s, as information costs dropped and business seek to integrate their information. Within firms, this integration took the form particularly of company-wide Enterprise Resource Planning (ERP) systems (such as those produced by German company SAP or Dutch company Baan) whose purpose was to integrate the information in all the separate databases so as to treat it in a unified way. Through these systems, firms have substituted flexible ways to code their data for more rigid but unified central databases.[16] Between firms, the increase in the commonality of the information has taken place with the integration of supplier and buyer networks using both Electronic Data Exchanges (EDI), and other similar systems to link suppliers and buyers.[17] These EDI systems allow for the exchange of electronic data between suppliers and customers by standardizing the format of the data exchanged. Again, in order to benefit from these networks, it is necessary for the individual firms to treat data in ways that are compatible with their suppliers – "improvements in coordination through EDI are dependent on the willingness of an EDI partners to invest in computer systems to improve its internal flow of information" (Hart and Saunders, 1997).

Since these information related changes have been widespread, our model would lead us to expect that the reduction in information costs should be accompanied by an increase in decentralization throughout the economy. Is there any evidence that this is indeed the case?

The economics literature has only recently turned its attention to the internal organizational structure of firms. But the robust finding by a number of authors working on micro-level data has been that the reduction in information costs is correlated with increasing decentralization. In particular, Brynjolfsson and Hitt, (2000) were the first to find the increasing complementarity between IT and decentralization. Bresnahan, Brynjolfsson and

---

[15]: "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.)

[16]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).

[17]Among these other related approaches are 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).

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 decentralization of authority is complementary with the use of IT. Finally, 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. This could be a consequence of the increasing commonality of codes, which makes the translation function of firms obsolete.

This evidence suggests that indeed the drop in information costs led to increasing integration in the codes of organizations. It also points to the existence of an empirical paradox between centralized information and decentralized decision making. However, it is not clear that the causal mechanism at play is the one identified by the model, namely that the introduction of the common code allows for the substitution of a key role of hierarchy, that of 'translation' and allows for direct horizontal communication between units that otherwise would be 'speaking a different language.' The following case studies provide detailed evidence that this causal mechanism may indeed have played an important role in the existence of a ink between reduction in information costs and decentralization.[18]

6.2   Common Codes at Microsoft

The organizational changes undergone by the Microsoft Corporation starting in the mid 90s offer an interesting case study of the trade-offs involved in the adoption of a common code. According to Robert J. Herbold,[19] COO of Microsoft at the time, Microsoft had in 1994 a completely decentralized set of codes. In the finance area, "the general managers of Microsoft's business and geographical units would sometimes decide to redefine or change, for their own purposes, a key measure used in financial reporting ... because these systems were incompatible, each quarter, the company shipped countless sheets of paper presenting the company's and individual units' financial results." The managers of the different units had all set up their own techniques of

---

[18]Garicano and Rossi-Hansberg (2003) study the impact of two other types of information costs, communication costs and the cost of accessing knowledge, on organization and inequality. They do not consider diagnosis costs. They find that decentralization can result from drops in the cost of accessing knowledge, and that such changes can also increase within worker class wage inequality.

[19]We rely heavily on the personal account of the COO of MS at the time, Robert J. Herbold in Harvard Business Review, January 2000. All the quotes below proceed directly from his account.

financial reporting, stressing what they believed were the important components. 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."

An advantage of such a situation was that managers could measure exactly what they needed to measure. 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." On the other hand, between unit communication was compromised, since lots of different measures had to be understood by top managers, and different measures often needed to be reconciled.

The company decided to move towards 'common codes' in those two areas. Among the main advantages of these moves, according to Herbold, were, first that business unit performances could be easily compared to one another, and second, that all managers could easily make sense of that information.

Paradoxically, this centralizing move provided "benefits usually associated with decentralization" as managers had instant access to information and could operate on it directly. "Giving managers instant access to company information accelerates decision making."

Inside the organization the cost and benefits of the adoption of common codes within these two areas differed, in ways similar to the ones discussed in section 5. For example, the German Country Manager refused 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. Germany certainly shouldn't be characterizes as just another European country." [20] The fact that the adoption decision was centralized, however, reduced the scope of these individual concerns to either bias the code or delay its introduction.[21]

---

[20]Obviously, it is hard to interpret this compaints as arguing that the move is efficient but not in the interest of managers – we only know that the center thought the codes were inefficiently different while the country managers think that the codes are just appropriately adapted to the different circumnstances of the country environments.

[21]This is not to say it eliminates such concerns. Herbold himself 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.

6.3   A common code for the B-2 Bomber

Advances in information technology allowed the design of the 'stealth' B -2 bomber by Northrop, Boeing, Vaught (a division of LTV) and General Electric to be the 'first major aeropsace program to rely on a single engineering database to coordinate the activities of the major subcontractos on a large-scale design and development project' (Argyres, 1999:163). [22] A key element in this program was 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. 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).

The use of the grammar had two consequences. First, it allowed for designers proceeding from different companies to participate jointly in the design. In previous projects, the difficulty of cross-company communication had meant all designers, with the exception of those fo the the motors (which are a relatively stand-alone component requiring little coordination) had belonged to the same firm.[23] Thus the existence of a common code allowed integration of several teams where before there was none possible.

Moreover, this integration happened with little need for the 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). By reducing the need for the coordinating role of the managers, this code provided for a larger scope for horizontal communication.

Of course, unlike in the Microsoft case just discussed, the decision on which code to adopt was largely decentralized and so individual strategic considerations of the type discussed in Section 5 played a more important role here. First, the B-2 project provides some evidence that there may be excessive code variety. Boeing and Vaught were unenthusiastic during the negotiations leading to the creation of a centralized database. A Boeing engineer argued explained that 'we were developing our own system CATIA [...] We knew we wouldn't be using CATIA if we had to be compatible

---

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

[23] Argyres, personal communication to the authors.

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). Not only the adoption was hard to attain, but it was also, as the theory 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, they all adopted the one of Northrop (Argyres, 1999:167).

7   Conclusions and Literature

This paper has presented an intitial step towards a joint theory of organizaiton and codes. We review here briefly existing literature on codes and conclude with what we view as our main contributions to the study of this problem

In his celebrated book on the limits of organization, Arrow (1974) discusses the endogenous development of codes within organizations. He also identifies a trade-off between having generally understood codes that allow for wide communication and designing specialized codes that fit the needs of a particular organization. In this paper, we have aimed to develop some of Arrow's ideas in a formal setup and to explore the properties of optimal codes and the terms of the trade-off between generality and specialization.

Information theory (Shannon, 1948) deals with optimal codes as well but with a different focus. The main constraint there is *channel capacity*. A code is chosen to minimize the cost of transmitting information. Instead, here the cost of transmission is not taken into account. The main constraint is the ability of agents to learn codes.

Marschak and Radner's (1972) *team theory* studies decision making when there are several agents with a common goal but limited communication. While our paper shares the team-theoretical emphasis on communication problems rather than incentive issues, it differs from Marschak and Radner in that it allows for endogenous communication protocols (at the cost of downplaying the decision-theoretic aspect).

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 knowl-

edge of rules of behavior. He then goes on to study, within a team theoretic framework, the benefits of shared knowledge. The model does not study the issue of codes or the communication costs of achieving the shared knowledge.

Radner (1993) and other authors (see Van Zandt, 1999 for a survey) study organization from the viewpoint of information processing. The key bounded rationality assumption is that agents have limited computation capacity. The analysis focus on the properties of optimal information processing hierarchies. Within the information processing literature, the work that is closest to ours is Bolton and Dewatripont (1994). They allow agents in an information processing hierarchy to achieve returns to specialization and they 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. Our paper is complementary to the information processing literature in that it explores similar issues but from the point of view of codes.

Garicano (2000) studies the problem of matching problems with solutions when codes do not exist, so that problems cannot be labeled at all. In that case, agents who draw a problem and look for a solution must ask other agents until they find someone who knows the solution. He shows that the optimal organization of knowledge is a knowledge-hierarchy, in which agents closer to the production floor deal with the most common and easier problems and shield the experts from these problems, so that these experts can focus on the exceptions.

Wernerfelt (2003) is similar to ours in that it considers codes that are enacted to minimize communication costs within an organization. But the focus is different: in his paper, the codes are designed in a decentralized way and the paper focuses on the existence of multiple equilibrium codes due to independent decision making. Instead our approach focuses on how the environment in which the organization operates determines the optimal code and the level of commonality with the codes of other organizations.

Battigalli and Maggi (2002) construct a model of language and its associated costs, and they use it 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. This paper shares Battigalli and Maggi's aim to take into explicit account the cost of using language to partition the set of events. However, the area of application of our model is different. We are interested in organizational structure rather than contract incompleteness. Also, the building blocks of our language differ substantially from Battigalli and Maggi.

Thus none of the previous literature studies the relations between the organizational code and its environment, nor the consequences that the constraints on individual's ability to learn words imposed on organizational design. Our focus on the organizational consequencs of codes has alowed us, for example, to discuss the decision of whether to segregate services or integrate services, or the use of specialized translators that allow, at a cost, the codes of units facing different environment to be preserved to some extent. This analysis has proved useful in providing insights on the effects of decreases in information costs on organizational design, as we argued in Section 6. In particular, we have shown that when diagnosis is less costly, organizations tend to more common codes, and to more horizontal communication as a result. Finally, we have explored independent individual adoption decisions, and have shown that two biases may be present when they are costly and non-contratible: too little code commonality, and codes that are inefficiently biased towards the needs of early adopters.

Beyond the substantive results we have obtained, we see the contribution of this paper as pointing out ways to think about and formulate a central problem to economic organization. We trust that future work will advance on the study of codes in organizations.

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

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

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

[12] 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.

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

[14] 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.

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

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

[17] 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.

[18] 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.

A   Appendix

A.1   Assigning agents to codes: Span and range of a code

We have shown in Section 3.3 that all different agents dealing with one particular engeneer must share a code. The question that naturally follows is which agents should then communicate to share a code. Suppose that there are a number of boundedly rational engineers, and that salesman should be assumed to them to minimize communication loss. When should many agents ahare the same codes? When should only a few do it? What determines this division?

Suppose in particular there are two engineers, e_A and e_B, and a continuum of salesmen.[24] 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 $S_A = [0, b^*]$ and $S_B = (b^*, 1]$, where $b^* \in (-1, 1)$.

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 *range* of engineer i be the proportion of salesman types that the engineer communicates to: $1 + b^*$ for e_A and $1 - b^*$ for e_B.

Proposition 9 *Suppose that $g(b)$ is increasing and linear. In the optimal organization, $c_A(b^*) = c_B(b^*)$. Furthermore, the span of engineer A is smaller than the span of engineer B and the range of engineer A is greater than the range of engineer B.*

Proof. Given b^* the problem for language A is

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

---

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

and for language B its is

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

where

$$F(x_A, b < b^*) = \frac{1}{G(b^*)} \int_0^{b^*} F(x_A, b) g(b) db$$

$$F(x_B, b > b^*) = \frac{1}{1 - G(b^*)} \int_{b^*}^1 F(x_B, b) g(b) db$$

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

$$c_A(b^*) = \frac{2x_A(b^*) - 1}{G(b^*)} \int_0^{b^*} F(x_A(b^*), b) g(b) db + (1 - x_A(b^*))$$

$$c_b(b^*) = \frac{2x_B(b^*) - 1}{1 - G(b^*)} \int_{b^*}^1 F(x_B(b^*), b) g(b) db + (1 - x_B(b^*))$$

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

$$= (2x_A(b^*) - 1) F(x_A(b^*), b^*) g(b^*) + g(b^*)(1 - x_A(b^*))$$

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

$$= -(2x_B(b^*) - 1) F(x_B(b^*), b^*) g(b^*) - g(b^*)(1 - x_B(b^*))$$

$$\Psi(b^*) = g(b^*)((2x_A(b^*) - 1) F(x_A(b^*), b^*) - (2x_B(b^*) - 1) F(x_B(b^*), b^*) + (x_B(b^*) - x_A(b^*$$

$$= g(b^*)(c_A(b^*) - c_B(b^*))$$

Thus, the optimum is when

$$c_A(b^*) = c_B(b^*)$$

It is easy to see that [TO PROVE]

$$c_A(0) < c_B(0)$$

because

$$E[b|b > 0] > -E[b|b < 0]$$

Note that

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

is nondecreasing in b^* [TO PROVE].

Then the unique value of b^* for which $\Phi(b^*) = 0$ is to the right of 0, which proves the statement. ∎