DISCUSSION PAPER SERIES

No. 4205

CODES IN ORGANIZATIONS

Jacques Crémer, Luis Garicano
and Andrea Prat

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP4205.asp**

ISSN 0265-8003

CODES IN ORGANIZATIONS

Jacques Crémer, Universite des Sciences Sociales de Toulouse and CEPR
Luis Garicano, University of Chicago and CEPR
Andrea Prat, London School of Economics (LSE) and CEPR

Discussion Paper No. 4205
January 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4205

January 2004

ABSTRACT

Codes in Organizations*

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 simple communication model and characterizes optimal codes. There exists a fundamental trade-off between choosing a specialized code that simplifies internal communication and a common code that facilitates external communication. We identify the drivers of this trade-off and we study the strategic aspects of code adoption. The results are used to interpret some existing organizational structures.

JEL Classification: D23
Keywords: coding, communication cost and organization theory

Jacques Crémer
Institut d'Economie Industrielle
Université des Sciences Sociales
Place Anatole France
31042 Toulouse Cédex
FRANCE
Tel: (33 5) 6112 8634
Fax: (33 5) 6112 8637
Email: jacques@cremeronline.com

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=116204

Luis Garicano
University of Chicago
Graduate School of Business
Department of Economics
1101 East 58th Street
Chicago, IL 60637
USA
Tel: (1 773) 834 2862
Fax: (1 773) 702 0458
Email: luis.garicano@gsb.uchicago.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=145086

Andrea Prat
STICERD
LSE, London School of Economics
Houghton Street
London
WC2 2AE
Tel: (44 20) 7955 6992
Fax: (44 20) 7955 6951
Email: a.prat@lse.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=138721

*We thank Phillippe Aghion, Matthias Dewatripont, Bob Gibbons, Jerry Green, Augustin Landier, Jean Tirole, workshop participants at the CEPR/Toulouse Organizations workshop and at the Harvard/MIT Seminar in Organizations for useful comments and Pedro Vicente for outstanding research assistance. This Paper is produced as part of a CEPR Research Network on 'Product Markets, Financial Markets and the Pace of Innovation in Europe', funded by the European Commission under the Research Training Network Programme (Contract No: HPRN-CT-2000-00061).

Submitted 19 November 2003

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

January 16, 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 simple communication model and characterizes optimal codes. There exists a fundamental tradeoff between choosing a specialized code which simplifies internal communication and a common code which facilitates external communication. We identify the drivers of this tradeoff and we study the strategic aspects of code adoption. The results are used to interpret some existing organizational structures.

1 Introduction

Because codes form an important part of the communication infrastructure of firms and organizations, they have been discussed quite extensively in

*We thank Phillippe Aghion, 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.

the economic literature, since, at least, Arrow (1974). The general insights stemming from these discussions are well understood. Whereas English and other natural languages are general codes, flexible and adapted to a wide range of situations, subsets of agents dealing repeatedly with each other find it useful to design specialized codes, which are more efficient for their specific circumstances. For instance, firms often initiate a common project by creating a Project Management Dictionary (Blankevoort 1986).[1] However, despite this ubiquity of codes, there has been little formal analysis of their properties, and their consequences for the organizations of firms have been nearly completely neglected. In this paper, we show how a theory of codes can enrich a theory of the optimal organization of firms.

The integration of a theory of codes into the theory of the firm yields new insights on the relationship between centralization of information and decentralization of decision making, and, as a consequence, enriches the study of organizational forms. In particular, the theory sheds some light on some aspects of the relationship between decentralization and information technology which have been discussed both in the economics literature[2] and in the business press. Accounting systems, human resource and other organizational data bases are codes, in the sense in which economics understands them. In recent years, the management of these codes within firms has become more centralized,[3] 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

[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 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] See Bresnahan, Brynjolfsson, and Hit (2002); Brynjolfsson and Hitt (2000), Caroli and van Reenen (2001), Rajan and Wulf (2001).

[3] The centralization of information took the form 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. 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). 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. EDI systems allow for the exchange of electronic data between suppliers and customers by standardizing the format of the data exchanged.

and centralizing certain systems also provided, perhaps surprisingly, benefits usually associated with decentralization" - we discuss this relationship below.

The first step of our analysis is to build a simple model of codes, as partitions of the set of information that an agent might want to send to another. Good codes provide information which is as relevant and as precise as possible while respecting the bounded rationality of the agents. We then derive some of their properties. An optimal code uses precise words for frequent events and vaguer words for more unusual ones. We show that the less precise words are used less often, even though they make allusion to a wider array 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 variety in 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 in Section 3 that agents will either have entirely separate codes or common codes. 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 want to use a common code for all of its activities? In Section 4, 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 show that two services which face similar tasks will not find a common code too costly. We model the synergies between services and show that a common code will be justified if they are sufficiently large.

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. We show that 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.

Different organizations may sometimes choose to use the same code, in order to facilitate cooperation. Of course, the choice of this code will be affected by strategic considerations that we study next. We show in Section 5 that there exists a first mover advantage: a firm will adopt a code sufficiently attractive to potential partners, but more adapted to its need. Moreover, when adoption costs are not contractible, too little code commonality will exist, as investing in a common code generates an externality in improving other's communications. This explains why coordination with organizations is better than coordination between organizations.[4] Some evidence for these inefficiencies is provided by the cooperation 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 Codes with two agents

2.1 A model of codes

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. Because of bounded rationality, the salesman uses a "code" $\mathcal{C}$, that is a partition $\{W_1, W_2, \ldots, W_K\}$ of the set X to transmit this information to the engineer: he only communicates the word k such that $x \in W_k$ — W_k is called the "meaning" of word k. We will use the metaphor of the communication between a salesman and an engineer throughout the paper, but our theory applies more generally to the case of an agent who needs to communicate to another agent some information about event in a set X, which leads to the following definition.

Definition 1 *A code C is a partition of the set X of events*

Vague words contain many events and lead to high search or *diagnosis costs* for the receiver of the information. For example, if he receives information that the characteristics of the customer belong to a large set W_k, the engineer must spend a lot of time searching for the best solution to the customer need. [5] More specifically, throughout the paper, the diagnosis cost will be assumed to be proportional to the cardinality of the word. If x is realized and $x \in W_k$, the diagnosis cost is $\lambda \times \sharp W_k$,where $\sharp$ denotes the number

[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.

[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.

of elements of a set and λ a proportionality factor. Therefore the expected cost of diagnosis when using code $\mathcal{C}$ is

$$\lambda \sum_{k=1}^{K} \sum_{x \in W_k} (f_x \times \sharp W_k) .$$

Of course, linear search costs are special cases, and we use them because they provide a tractable technique to study the organizational consequences of codes. Moreover, some models of search do yield linear costs. For instance, suppose that a client is described by a vector of valuations for the different products sold by the company. The company then tries to find the product that will best fit the needs of the client. The salesman observes the vector of valuations, and communicates to the engineer the word containing the most valuable product. Since the search for the maximum requires the engineer to examine all of the events in the word, the search costs will be proportional to the size of the word.[6]

Finally, we need to describe the bounded rationality of the agents: they can learn at most $K \geq 2$ words; on the other hand, there is no cost in increasing the number of words up to K. We leave for future research the study of possible trade-off between diagnosis costs and the richness of the language.

A code is optimal if it minimizes the expected diagnosis cost subject to the constraint that each agent knows no more than K words. The next section presents some properties of optimal codes.

2.2 Optimal codes with two agents

We begin by studying the optimal code when one agent needs to communicate with a single other agent. The breath n_k of word k is the number of events that are described by k, that is $n_k = \frac{\sharp W_k}{N}$, where N is the cardinality of X. The *frequency* p_k of word k is the probability that an event belongs to W_k, that is $p_k = \sum_{x \in W_k} f_x$. 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 is used frequently (because drizzle occurs all the time), “bad weather” is broad and used frequently, “good weather” is

[6] In reality, finding a solution to the customer's problem that gives a payoff close to the maximum could in many cases be sufficient, and would lead to different search strategies. On the other hand, if search costs are low compared to the benefits of a better fit, the search for a maximum would be (close to) optimal and the analysis of the text carries through.

broad and used relatively infrequently, and "hurricane" is both narrow and infrequently used.

The problem of finding an optimal code is an integer problem, which makes it difficult to characterize the solution completely. However, we can show that the solution has two important properties.

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

Proof. Exchanging x and x' must (weakly) increase the diagnosis cost. Therefore,

$$n_k(p_k + f_{x'} - f_x) + n_{k'}(p_{k'} - f_x + f_{x'}) \geq n_k p_k + n_{k'} p_{k'},$$

which implies $(n_k - n_{k'})(f_{x'} - f_x) \geq 0$, which implies the result. ■

Proposition 1 is a consequence of the fact that the objective function in (**??**) 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. Notice however that the fact that the diagnosis cost is linear in the number of events in the words play no role in the argument. If the cost was $\Gamma(n)$, then we would obtain $(\Gamma(n_k) - \Gamma(n_{k'}))(f_{x'} - f_x) \geq 0$, which yields the result as long as Γ is increasing.

Proposition 1 describes the familiarity of events in different words. The following proposition describes the familiarity off the words themselves.

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

Proof. Transferring word $\tilde{x}$ from W_k to $W_{k'}$ cannot lower costs. Hence, we must have

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

which implies

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

and therefore the first statement in the proposition.

To prove the second statement, rewrite inequality (1) as

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

■

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. 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 last part of 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 $k^{'}$.

Up to now, 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". We treat this problem in the framework of section 2.2, without a natural order for the events.

To give a precise meaning to the inequality of distributions, assume without loss of generality that the events are indexed by real numbers and that

$f_x \leq f_{x'}$ if $x < x'$ (the probabilities of the events are increasing in their indices). We say that another distribution of events $\tilde{f}$ (where the probabilities are not necessarily increasing 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.

With this definition, we can show that communications 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 $\sum_k p_k n_k$ while the cost associated with $\tilde{f}$ is $\sum_k \tilde{p}_k n_k$, where $\tilde{p}_k = \sum_{x \in W_k} \tilde{f}_x$ ($\tilde{p}_k$ is not the probability of the k^{th} word in the optimal code associated with $\tilde{f}$).

By Proposition 1, word size is nondecreasing in k. Then, writing $P_k = \sum_{k' \leq k} p_{k'}$ and $\tilde{P}_k = \sum_{k' \leq k} \tilde{p}_{k'}$, we have

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

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

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.

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

3 Code commonality

3.1 Using the same code to communicate with different agents

Up to now we have considered only the optimal choice of codes with two agents. Suppose now that two salesmen with different types of clients, such as the salesman from region A and the salesman from region B, must communicate with the same engineer. 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 will a common code be chosen?

The trade-off between a common code and different codes or dialects runs 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.

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

Given the strict constraints on bounded rationality that we have assumed, it is optimal for the same code will be used by both services. Intuitively, if the engineer incurs the cost of learning an extra word to communicate with one service, he might as well also use this word to communicate with the other service.

Proposition 4 *Only a common code can be efficient*

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 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[8], and suppose 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 $W \in W'/(W' \cap W_k)$ 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. ■

Three examples will illustrate the proof.

Let $\mathcal{C}_A = \{\{1,4\},\{2,5\},\{3,6\}\}$ and $\mathcal{C}_{\mathcal{B}} = \{\{1,2,3\},\{4,5,6\}\}$. The narrowest noncommon words are $\{1,4\}$, $\{2,5\}$, and $\{3,6\}$. Take $W_k = \{1,4\}$. Then, $\tilde{\mathcal{C}}_{\mathcal{B}} = \{\{1,4\},\{2,3\},\{5,6\}\}$. 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 $\tilde{\mathcal{C}}_{\mathcal{B}}$ 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

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

[8] 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 $W_{\tilde{k}} \notin \mathcal{C}_1 \cap \mathcal{C}_2$.

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

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

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.

Corollary 1 *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. We will introduce later the possibility for the firm to introduce agents who can learn more words and use this skill to 'translate' among different sets of agents.

While the strict commonality of the codes is a feature of the tight bounds on the rationality of the agents we have imposed, the general point that the proof points out is true more broadly: bounded rationality together with the need to communicate among wide groups of agents imposes sharply decreasing returns to the variety in codes. Tailoring words to particular agents is costly as it limits the set of agents among whom the words are useful. This important aspect of code commonality is the key determinant of the decreasing returns to scope in organizations that shape their scope and their use of integrating mechanisms such as hierarchy, which we study in section 4. First, we discuss the assignment of different agents to alternative codes.

3.2 Two Word Codes

The result in the previous subsection, implying that agents will either share a fully common code or have separate codes, allows us to substantially simplify our framework by restricting our attention to two–word codes. This still allows us to study issues such as the commonality of codes, since agents choose either common codes (when those in A and B communicate with one another) or separate codes (when they do not). For simplicity, we also assume in what follows that the number of events is sufficiently high that X can be approximated as continuous.

Suppose that a salesman deals with consumers $x \in [0,1]$ drawn from a distributions with cdf $F(x) = (1-b)\,x + bx^2$ and 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. Here b is a measure of how unequal the distribution of events is. In this case the optimal two-word code is the solution to:

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

This optimization has a closed form solution[9] 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.[10]

3.3 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.[11] 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 $b^* \in (-1, 1)$.

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

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

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

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

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

.

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

agents, there is a limit to engineer's span: the organizational costs of getting diverse salesmen, as it leads to less adapted codes.

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 next.

4 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 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.

4.1 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,

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

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.

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. 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.

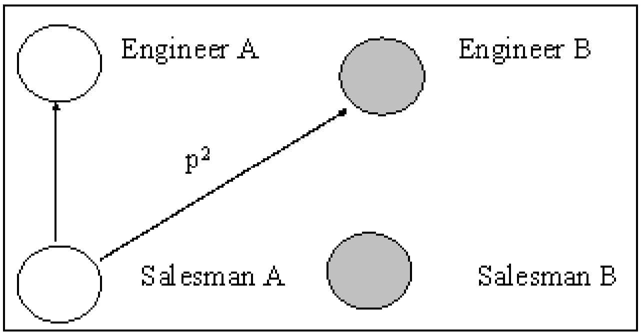

Figure 1: *Synergies exist when there is excess demand on one service and excess capacity in the other.*

4.2 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:

$$\Pi(p, b|\mathcal{C}_j) = 2(1 - p(1 - p))(1 - \frac{\lambda}{2}). \tag{3}$$

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

$$\Pi(p, b|\mathcal{C}_s) = 2(1 - p)(1 - \lambda D^*(b^2)). \tag{4}$$

The organization should be integrated rather than separated[12] 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:

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

Define p^*, λ^* and b^* as any set of parameters such that (5) holds with strict equality, that is the firm is indifferent between integration and separation. Then 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.

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

Proposition 6 *For any* (p, λ, b) *such that* $p \geq p^*$*,* $\lambda \leq \lambda^*$ *and* $|b| \leq |b^*|$ *with at least one of these inequalities strict, the integrated organization is strictly more profitable that the separated organization. Generally, a decrease in* p*, and increase in* λ *or an increase in* $|b|$ *favors a separated organization.*

Proof. To prove the first part, just rearrange the first part rearrange (5). For the comparative statics

Define

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

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 $\partial D^*(b)/\partial b < 0$ for $b \in [0, 1]$. ■

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 is the use of the most appropriate code, the lower the cost for each service of adopting the common code rather than their optimal code, and the greater the benefits of an integrated organization. Finally, using a common code is least costly when the distributions of problems is more similar in the two services (a small $|b|$).

4.3 Hierarchy

Suppose instead that 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.

There is fixed cost μ of hiring the translator, but because 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 presented below go through even 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 7 *Consider any b and $p \geq p^*$, such that an integrated organization is optimal for some λ. If μ is low enough, there exist $1 \leq \lambda_{\min} < \lambda_{\max} \leq 2$ such that the unique optimal organization is*

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

To understand the intuition for this proposition, refer to Figure 2 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.

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.

4.4 Evidence from the case of 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,.[13] 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

[13] 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.

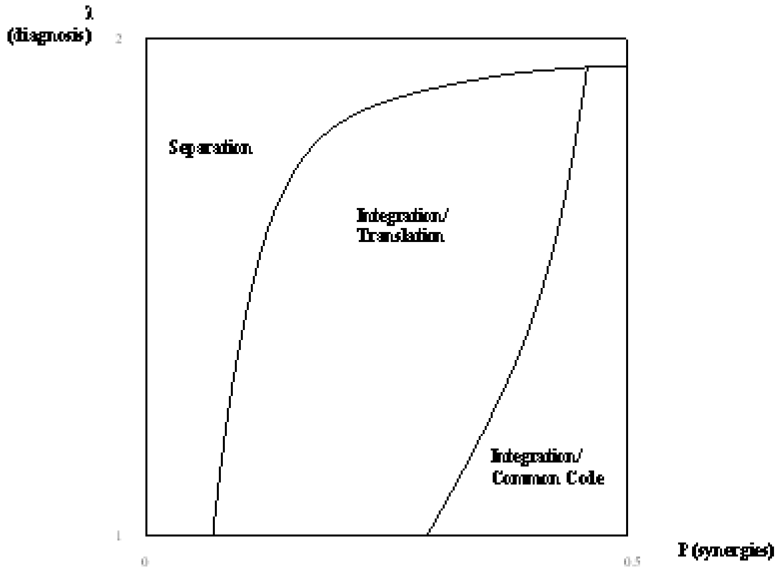

Figure 2: Simulation of organizational structure and code design. $b = 0.4$ and $\mu = 0.006$.

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 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."

Even though the adoption of a common code appears (particularly ex-post) 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. Germany certainly shouldn't be characterizes as just another European country." [14] 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. The next section studies how the adoption decision changes when it depends on the decisions of independent agents.

5 Common codes in different firms

The previous analysis abstracted from strategic considerations: codes were chosen to maximize total surplus. In this 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.[15]

Obviously, with complete contracts, agents would agree to select the surplus-maximizing code and, eventually, to make appropriate side payments. However, in many cases 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, at prohibitive cost, full access to the firm.

We first examine sequential code adoption. We ask what are the incen-

[14] 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.

[15] This is not to say such concerns are non-existent within firms. Microsoft ex-COO Herbold himself (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.

tives 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 can adopt a common codes, at some (fixed) cost. The presence of externalities may inhibit the adoption of an efficient common code.

5.1 Biased codes

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 2). 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 (this is equivalent to assuming that the salesman makes a take-it-or-leave-it offer with probability σ).

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. 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 $(1-p)(1-\lambda D^*(b^2))$ as in the previous section.

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

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

And thus the expected profit is:

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

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

$$\begin{gathered} \max \Pi_A(p,b,\sigma|\mathcal{C}_j) \\ \text{subject to } \Pi_B(p,b,\sigma|\mathcal{C}_j) \geq \pi^S. \end{gathered} \tag{6}$$

Then we can show the following result:

Proposition 8 *A joint code is adopted if and only if the optimal joint code yields aggregate profits larger than two separate codes. The joint code that is actually chosen will be better adapted to the needs of the second mover, firm B, than the code that firm A would have chosen in isolation.*

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 Excessive code variety

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 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 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

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

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 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 9 *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.

5.3 Evidence from 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 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 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). [16] 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

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

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.[17] 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 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 code adoption decision was largely decentralized and so individual strategic considerations of the type discussed previously played a more important role. Indeed, 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 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). 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, all parties adopted Northrop's (Argyres, 1999:167).

6 Related literature and conclusions

This paper has presented 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

[17] Argyres, personal communication to the authors.

discusses 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) has also dealt with optimal codes, but with a different focus: the main constraint is *channel capacity*, and codes are chosen to minimize the cost of transmitting information. We do not take into account the cost of transmission, and the main constraint is the ability of agents to learn codes.

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, but the paper contains no models of codes and no analysis of organizational forms.

More recently, Battigalli and Maggi (2002) construct a more 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, we are interested in organizational structure rather than contract incompleteness.

Like us, 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.

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.

In Garicano (2000), the bounded rationality of agents prevent 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. Focussing 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 tht the causal mechanism we propose is consistent with the one present in some detailed case studies of decentralization and organization.

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] 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.

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

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

[19] 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.

[20] 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.

[21] Wermerfelt, Birger

A Appendix

A.1 Proof of Propositions

Proposition 5 (Span of a code)

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

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

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

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

Note that

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

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

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

$$\begin{aligned}
\Psi(b^*) &= g(b^*)\left((2x_A(b^*) - 1)\,F(x_A(b^*), b^*) - (2x_B(b^*) - 1)\,F(x_B(b^*), b^*) + (x_B(b^*) - x_A(b^*\right. \\
&= g(b^*)\,(c_A(b^*) - c_B(b^*))
\end{aligned}$$

Thus, the optimum is when

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

It is easy to see that

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

but I think this is the opposite, the A's have higher coststhe reason is that the average because

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

Note that

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

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

Proposition 7 (Hierarchy)
Proof. 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 $p \geq p^* = 0.213$ 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. ■

Proposition 8(First mover bias) **Proof.** Given the symmetry of the two firms, a joint code is strictly superior to a separate code form an efficiency point of view if and only if

$$\Pi_A(p,b,\frac{1}{2}|\mathcal{C}_j) = \Pi_b(p,b,\frac{1}{2}|\mathcal{C}_j) > \pi(p,b|\mathcal{C}_s). \tag{7}$$

If this inequality does not hold the problem (6) has no solution, and firm A will choose its optimal code as if B did not exist. Similarly, B will choose its optimal code. On the other hand, if (7) is satisfied, there is clearly a common code accepted by B that A will propose.

To prove the second part of the proposition, assume a joint code is strictly superior to separate codes. Firm A will choose the code that minimizes

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

Note that

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

By symmetry,

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

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,

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

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

Hence,

$$\text{if } x\leq\frac{1}{2} \qquad \frac{d}{dx}\pi_A^J\left(x\right)>0. \tag{8}$$

If the participation constraint is not binding, firm 1 faces an unconstrained maximization problem over $\pi_A^J\left(x\right)$. By (8), 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 $\pi_B^J\left(x\right)\geq\pi^S$. But, by (8), this implies that the optimal x is to the right of $\frac{1}{2}$. ∎

A.2 Extension of proposition 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[18] $[t_k, t_{k+1}]$, is the probability that the word is used, now $F[t_{k+1}] - F[t_k]$, and the breadth of 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 10 (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 foten. .*
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

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

[18]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 first-order conditions[19] are

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

which implies

$$f(t_k) = \frac{[F(t_{k+1}) - F(t_k)] - [F(t_k) - F(t_{k-1})]}{(t_k - t_{k-1}) - (t_{k+1} - t_k)} \tag{9}$$

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 (9):

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

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. ■

[19] 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.