Organizational Restructuring in Response to Changes in Information-Processing Technology

Jacek A. Cukrowski
CERGE-EI, Prague, Czech Republic

May 1996

Abstract

This paper analyzes the effects of changes in information-processing technology on the efficient organizational forms of data-processing in decision-making systems. Data-processing in decision-making is modelled in the framework of the dynamic parallel-processing model of associative computation with an endogenous duration of operations. In such a model, the conditions for the efficient organization of information-processing are defined and the architecture of the efficient structures is considered. It is shown that decreasing returns to scale of the information-processing function and the information overload of the system are necessary and sufficient conditions for the decentralized (hierarchical) information-processing, respectively. Moreover, the analysis shows that the size of the efficient structures is determined exclusively by their information workload and the existing information-processing technology. In particular, the results indicate that, for a given information workload, the size of information-processing structures in decision-making systems is inversely related to the degree of homogeneity of the information-processing function. Consequently, the organizational restructuring of decision-making systems towards flatter hierarchical structures can be explained exclusively by those technological changes which affect economies of scale in information-processing, i.e., which increase the degree of homogeneity of the information-processing function.

Keywords: Technological changes, organizational restructuring, hierarchy, decision-making, information-processing.

JEL classification: O33

Abstrakt

Tato práce analyzuje vlivy změn v technologii zpracování informací na efektivní organizační formy zpracování dat v systémech rozhodování. Zpracování dat v systémech rozhodování je modelováno v rámci dynamického paralelního modelu zpracování asociativních výpočtů s endogenním trváním operací. Pro tento model jsou zde zadefinovány podmínky pro efektivní organizaci zpracování informací a je zde pojednáno o architektuře efektivních struktur. Je zde také ukázáno, že klesající výnosy z rozsahu zpracování funkce zpracování informací a informační přetížení systému jsou nutné a postačující podmínky pro decentralizované, resp. hierarchické zpracování informací. Analýza navíc ukázala, že velikost efektivních struktur je určována výlučně informačním zatížením a stávající technologií zpracování informací. Výsledky naznačují, že velikost struktur zpracování informací při daném informačním zatížení je v inverzním vztahu se stupněm homogenity funkce zpracování informací. Organizační restrukturalizace systému rozhodování směrem k méně členitým hierarchickým strukturám může být následně vysvětlena výhradně technologickými změnami ovlivňujícími úspory ze zvýšeného rozsahu zpracování informací, tzn. zvýšením stupně homogenity funkce zpracování informací.

1. Introduction

Several leading economists have recognized the importance of informational structures for the internal organization of business firms and large corporations (see, e.g., Aoki, 1986, or Milgrom and Roberts, 1990). In particular, decision-making systems of such enterprises are widely perceived to have complex informational structures (see, for example, Radner, 1992 and 1993, Radner and Van Zandt, 1992, or Bolton and Dewatripont, 1994). The architecture of these structures is often considered as a factor which determines the profitability of the firm as a whole (see, e.g., Milgrom and Roberts, 1990, or Radner, 1992). An overview of the contributions made by recent research to understanding the economic significance of data-processing in the management of modern enterprises is presented by Radner (1992). In the same paper, however, he emphasizes that it is still not clear what formal requirements have to be satisfied in order to observe decentralized (hierarchical) organization of informational processes.

Taking into account the very limited ability of individuals to handle and process information (for given information-processing technology), some economists argue that hierarchical forms have been developed to respond to the growing problem of handling an increasing flow of information (Chandler, 1966, or Bolton and Dewatripont, 1994). Not surprisingly, therefore, organizational changes (the flattening of hierarchical structures) observed in the last decade (see, e.g., *Business Week*, 1989) are usually explained by an improvement in information-processing technology (Schein, 1989, Kennedy, 1994, or Bolton and Dewatripont, 1994). However, research to date has not provided an adequate economic explanation of the conditions under which improvements in information-processing technology might reduce the depth of hierarchical structures. Thus, an examination of the criteria which have to be satisfied for the restructuring of the information-processing structures of decision-making systems in response to changes in information-processing technology is the subject of the analysis below.

The paper is concerned with the analysis of data-processing in the decision-making sector of the firm. The consideration focuses on the computation of associative operations[1] because a number of decision-making paradigms involve primarily operations of this kind (see Radner, 1992 and 1993).

[1] A binary operation (*) is associative if the following is true: $A*(B*C)=(A*B)*C$, where A,B,C denote items of data.

As an example of associative computation in the control and management of the firm, a decision-making process where actions are chosen using the pattern-recognition (nearest-neighbor classification) procedure can be considered. In this decision-making scheme, decision-makers compare sets of data about the environment with elements of the finite set of reference patterns, and to each element there corresponds a particular action (decision) to be taken. Thus, the problem of decision-making consists in the calculation of the closest reference pattern, i.e., calculating the distances between the reference patterns and data analyzed and choosing the minimum. Note that calculating the distances and choosing a minimum are both associative operations (see Radner, 1992, for details).

Another popular example of associative computations in decision-making is the linear decision rule, where the value of the linear function $c_1x_1+c_2x_2+...+c_Nx_N$ is computed (c_i is a coefficient of a conversion to a common unit and x_i is a numerical data item, i=1,2,...,N), and the decisions are made based on the result of this computational process. In practice, the items aggregated may not be just numbers but vectors or matrices. Computations of such a kind are commonly used in the methods of statistical prediction or statistical control (see Marschak and Radner, 1972, Aoki, 1986, or Radner and Van Zandt, 1992 and 1993).

Since both addition and choosing a minimum (or maximum) are associative operations, a dynamic parallel-processing model of associative computation (Radner, 1992 and 1993, and Radner and Van Zandt, 1992) is frequently used in the economic literature to describe computational processes in decision-making.

In the present paper, the model of associative computation is extended to include the assumption that the speed of information-processing in each individual processor depends upon the capital and labor allocated to it, i.e., the duration of individual operations is endogenized. Moreover, it is assumed that in order to produce a flow of decisions, the same computational procedure is repeated again and again. Consequently, the analysis is restricted to the organization of the single cohort of data-processing (one-shot mode). In such a model, the conditions for the efficient organization of data-processing are defined, and the architecture of efficient structures is determined. We show that the size of the efficient information-processing structure is determined solely by the existing information-processing technology and the information workload of the system. Finally, assuming a fixed information workload, we find that the size of the efficient structures might vary only in response to those technological changes which affect returns to scale of the information-processing function and, in particular, that the flattening hierarchical structures observed recently can be explained only by those technological improvements[2] which increase the degree of homogeneity of the information-processing function. Technological changes which leave returns to scale of the information-processing function unaffected decrease the delay in information-processing but do not influence efficient organizational forms.

The remainder of the paper is organized as follows: In Section 2, the decision-making sector of the firm is characterized and the model of information-processing in decision-making is presented; Section 3 analyzes the architecture of efficient processing networks; Section 4 examines conditions for decentralized information-processing in decision-making systems. Section 5 discusses the relationship between the characteristics of the information-processing function and the size of the efficient structures. Section 6 concludes.

2. Data-Processing in Decision-Making

Empirical studies of the labor market in industrialized countries show that much more than one-third of employees work full time in the management of firms, performing activities not directly connected with the production process such as processing and communicating information, monitoring actions of the other members of the firm, analyzing the market, planning, training employees, making decisions, etc. (see Radner, 1992, for a detailed analysis of this issue). The common feature of all these activities is information-processing, i.e., collecting and aggregating information, transforming data, presenting them in the appropriate form, etc. The majority of the activities associated with data-processing in the firm is carried out by managers with the help of staff, secretaries, or clerks using computation and telecommunication equipment, buildings, electricity, etc. Thus, the management (or decision-making) sector of the firm is understood to be a system which takes signals from the environment and uses the capital and labor to transform them into decisions. The quality of these decisions is inversely related to the deviation from the best possible decisions. The difference between decisions required and decisions made depends, in general, on errors in the computational process and delays in data-processing. If all members of the decision-making team have no incentives to report false information and do not make mistakes (i.e., errors are not possible), then the quality of the decisions depends solely on the delay in data-processing (see, e.g., Radner and Van Zandt, 1992).

[2] We can consider only technological improvements, because in a world with perfect information, new technology that reduces output per unit of each input would never be adopted.

The simplest example of the transformation of data from the environment into decisions which involves computation of associative operations is a linear decision rule (see Section 1). Therefore, without loss of generality, we can focus exclusively on this decision-making paradigm. To simplify the analysis, assume that conversion to a common unit is not required, i.e., $c_i=1$, for $i=1,2,...,N$. Following Radner (1992, 1993), represent the computational process in the decision-making sector of the firm as an idealized parallel computer. That is, let each computational center in the firm be a processor which contains both a memory for data storage (called a buffer) and a register where summations are made. Each processor can read a single item of data from its memory and add the value to the contents of the register resetting it equal to the resulting sum. Loading and adding a single datum to the contents of the register is called an operation. The time of operations is assumed to be the same 1) whatever the values of data added are or 2) when a datum is added to the cleared register (i.e., to zero). Moreover, a processor can send the contents of its register to an output or to the buffer of any other processor (through a communication channel) in zero time (see Radner and Van Zandt, 1992, for details).

Each processor has a limited capacity in that there is a maximum number of operations it can compute per unit of time. However, the speed of computation in each individual processor depends upon the capital and labor allocated to it. The relationship between the resources allocated to a single processor and the number of operations it can compute in a unit of time is determined by the technology of information-processing and is given in functional form as an information-processing function (F(k,l)). This function is understood to be a "production function" in information-processing and specifies the number of operations per unit of time that can be made in a single processor to which capital (k) and labor (l) are allocated. Similarly to the ordinary production function, the information-processing function is assumed to be technologically efficient in the sense that it is not possible to get the same output (compute a given number of operations in a unit of time) using less of one input and no more of the other. The properties of the information-processing function (such as, for example, returns to scale or homogeneity) can be defined analogously as in production theory.

The duration of a single operation (d) can be determined as 1/F(k,l). It is also a function of the capital (k) and labor (l) employed in the processor considered (d(k,l)=1/F(k,l)). If all processors in the structure are identical (i.e., the structure is homogeneous) then the duration of each individual operation can be specified as

$$d\left(\frac{K}{P},\frac{L}{P}\right) = \frac{1}{F\left(\frac{K}{P},\frac{L}{P}\right)}, \qquad (1)$$

where K and L denote capital and labor allocated to information-processing, respectively, and P is the number of the processors in the structure.

Assuming that data items are not costly (Radner, 1992 and 1993), the total cost of the computational process is determined by the costs of the resources (i.e., labor and capital) involved in the computation.

Each processor adds data items in a serial fashion. Thus, to speed up the computational process, data-processing can be done in parallel using more than one processor (i.e., in a decentralized computational structure). However, in the decentralized structure, the fixed amount of the resources allocated to information-processing have to be distributed among all the processors. Consequently, the processing power of each individual processor decreases. It implies that, on the one hand, parallelization reduces the length of the longest sequence of operations needed for the computation of the result but, on the other hand, increases the duration of each individual operation (decreases the computational power of the processors). Therefore, the delay in computation in the decentralized structure would not necessarily be smaller than in the original one (i.e., the decentralized computational structure would not necessarily be better than the original structure), and, consequently, organizational forms of data-processing have to analyzed in an economic context.

3. Efficient Organization of Information-Processing in Decision-Making

Following Radner (1992 and 1993), assume that processors in the computational structure cannot make errors in data analysis. In this case, the value of the computational service[3] is determined solely by the quality of the result computed which is inversely related to the delay in information-processing (see Radner and Van Zandt, 1992). The delay in the computation of the cohort of N items of data (D_N) in any homogeneous structure with P processors is proportional to the duration of individual operations, d(K/P,L/P), and, consequently, is a decreasing function of the resources allocated to the computational structure, i.e., $\partial D_N(K,L)/\partial K<0$ and $\partial D_N(K,L)/\partial L<0$. Thus, in the efficient information-processing structures the resources allocated to data-processing should be minimized for a

[3] The value of the computational service in decision-making is understood to be the difference between the value of the decisions based on the computational service and the value of the decisions without the service (Radner and Van Zandt, 1992).

given delay in computational process and the number of data items processed (or vice versa, i.e., for a given information workload and amount of resources the delay in the computational process should be minimized).

Definition 1: *Efficient Structures.* The information-processing structure is said to be organized in an efficient way if, for a given number of data items processed (N), it is not possible to get the same delay in information-processing (D_N) using less of one input to information-processing (i.e., capital or labor) and no more of the other.

In the model under study, the duration of individual operations is not constant but is endogenously determined in the model, i.e., it depends on the resources allocated to information-processing and the number of processors in the structure. Nevertheless, the architecture of the efficient structures remains the same as in the original dynamic parallel-processing model of associative computation with an exogenous duration of operations.

Proposition 1

For any given number of data items processed (N) *and any feasible combination of the resources* (K,L)*, the minimum delay in information-processing in any computational structure with* P *processors is given by the following formula:*

$$D_N(K,L) = \frac{\left\lfloor \frac{N}{P} \right\rfloor + \lceil \log_2(P+N \bmod P) \rceil}{F(\frac{K}{P},\frac{L}{P})}\ , \tag{2}$$

where brackets ⌊ ⌋ *and* ⌈ ⌉ *denote rounding down and up to the nearest integer, respectively.*

Proof:

Gibbons and Rytter (1988) show that the minimum delay in the summation of N data items in the structure with P identical processors (with fixed processing power and a duration of individual operations d=1) is determined by the time of computation of $C_N(P)$ operations, where $C_N(P)$ is given as follows:

$$C_N(P) = \left\lfloor \frac{N}{P} \right\rfloor + \lceil \log_2(P+N \bmod P) \rceil\ , \tag{3}$$

In the model under study, the duration of each individual operation (d) is a function of the resources allocated to information-processing and the size of the structure (i.e., d=d(K/P,L/P)). Consequently, for a given number of data items (N) and the resources (K,L), the minimum delay $D_N(K,L)$ that can be attained by any structure with P identical processors can be determined as $D_N(K,L)=C_N(P)d(K/P,L/P)$. Taking into account (3) and using d(K/P,L/P)=1/F(K/P,L/P) yields (2).

Q.E.D.

Radner (1992 and 1993) shows that the minimum time needed to add N items of data with the help of P processors is attained by so-called "skip-level reporting" structures with processors loaded as equally as possible (if $1<P\leq\lfloor N/2\rfloor$),[4] or by a fully centralized structure (if P=1). The term "skip-level reporting" refers to the practice in an organization whereby a processor in level[5] X sends reports (partial results) to a processor in level X+L ($L\geq 1$). That is, the processor in level X can skip one or several levels in reporting to its direct hierarchical superior. An example of the skip-level reporting structure (with P=8 processors, designed for the summation of N=40 items of data) is presented in Figure 3.1. In this network each of the processors receives five numbers. All the processors spend periods 1 through 5 adding numbers. At this point, four of the processors send their total to the other four, with each processor receiving one number. This is added to the processor's previous total in period 6. At the end of this period, two of the processors send their partial results to the other two. These numbers are added to previous totals in period 7, after which one processor sends its total to the other. Finally, the result is computed in period 8. The time diagram describing this computational process is shown in Figure 3.2.

Note that the denominator in (2) always decreases with P while the numerator is a non-increasing function of P. It follows that the structure of the size P, such that

---

[4] The number of processors (P) in any skip-level reporting structure is limited ($P\leq\lfloor N/2\rfloor$) because at least two data items have to be assigned to each of them.

[5] The processor belongs to the level

$$X = \begin{cases} 0, & \text{if it does not have any subordinate processors,} \\ x+1, & \text{otherwise,} \end{cases}$$

where x denotes the highest level of the hierarchy to which one of its immediate subordinate processors belongs (see Figure 3.1 for an example).

Figure 3.1
The Skip-Level Reporting Structure (P=8, N=40, Processors Are Represented As Circles, Triangles Represent Their Information Workload)

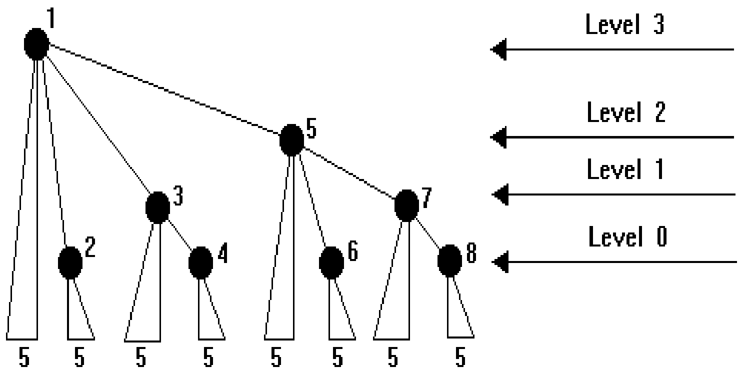
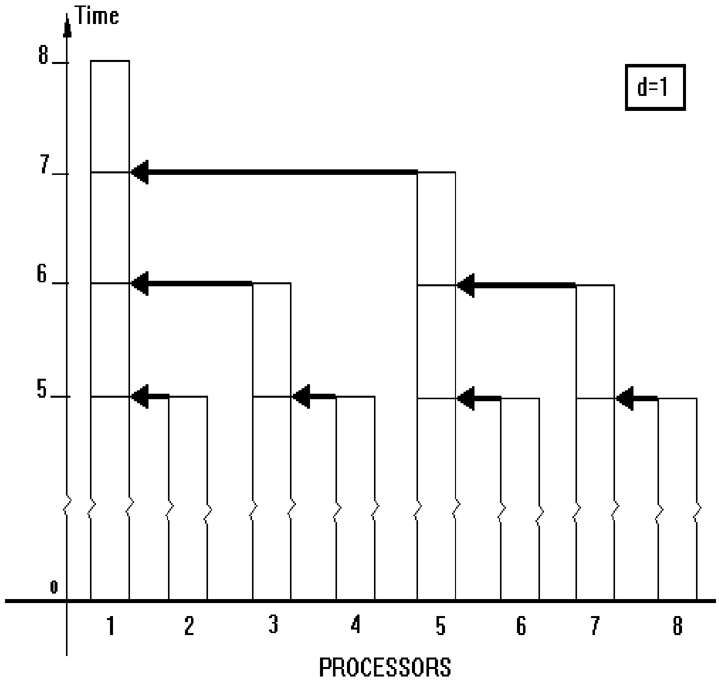


Figure 3.2
The Time Diagram Corresponding to the Information-Processing Structure Presented in Figure 3.1



$$\lfloor N/(P-1)\rfloor + \lceil \log_2(P-1+N \bmod (P-1))\rceil = \lfloor N/P\rfloor + \lceil \log_2(P+N \bmod P)\rceil , \qquad (4)$$

is surely inefficient—i.e., for any given amount of the resources, a smaller delay can be attained in the structure with P−1 processors. Consequently, the size of the efficient structure is in the set $P_N = \{1,2,...,\lfloor N/2\rfloor\} \setminus P_{N-}$, where P_{N-} is the set of P for which the equality (4) is satisfied.

Thus, for any information workload (N), the efficiency frontier (in the model with an endogenous duration of individual operations) is characterized by the following expression

$$D_N(K,L) = \underset{P}{Min} \frac{\left\lfloor \frac{N}{P} \right\rfloor + \lceil \log_2(P+N \bmod P)\rceil}{F(\frac{K}{P},\frac{L}{P})} , \qquad (5)$$

where $P \in P_N$.

Consequently, for a given information workload (N), the size of the efficient structures (which minimize the delay in data analysis, $D_N(K,L)$, for a given combination of the resources to information-processing (K,L)), P^*, is determined solely by the properties of the information-processing function.

4. The Decentralization of Information-Processing in Decision-Making

Although the size of the efficient structure (for a given number of data items processed) depends on both the amount of resources used and the properties of the information-processing function, in the case when the information-processing function faces non-decreasing returns to scale, the centralized structure (P=1) minimizes delay in data-processing for any feasible combination of the resources (i.e., is always efficient). This property is formally established by the Proposition below.

Proposition 2

If the information-processing function exhibits non-decreasing returns to scale (i.e., constant or increasing), then the efficient information-processing structure is centralized.

Proof

For any given number of data items processed (N) and any feasible combination of the resources (K,L), the minimum delay in information-processing in any computational structure with P processors is given by the expression (2). If the information-processing function faces non-decreasing returns to scale, then the denominator in (2) decreases at least proportionally with P. At the same time, the numerator in the expression (2) decreases with P less than proportionally. This implies that the delay in information-processing increases with the number of processors (P). Consequently, the efficient structure is the smallest one, i.e., $P^*=1$.

Q.E.D.

Proposition 2 can be illustrated by a simple example. Assume that the information-processing function has a general Cobb-Douglas form: $F(K,L)=K^\alpha L^\beta$, where $\alpha+\beta\geq 1$. For any N ($N\geq 2$) and any feasible combination of the resources (K,L), the delay in information-processing in the centralized structure equals

$$D_{N,P=1}(K,L) = N/(K^\alpha L^\beta).$$

The delay in information-processing in the structure with two processors equals

$$D_{N,P=2}(K,L)= 2^{\alpha+\beta}[\lfloor N/2\rfloor + \lceil \log_2(2+N \bmod 2)\rceil]/(K^\alpha L^\beta)$$

and is always greater than $D_{N,P=1}$ if $\alpha+\beta\geq 1$. Moreover, it can be verified that, for the information-processing function considered, the delay in information-processing in any structure with $P>1$ ($P\in P_N$) is greater than in the centralized one.

Note that Proposition 2 says nothing about the conditions under which the decentralization of the information-processing structures would be desirable. However, it shows that decreasing returns to scale should be considered as a necessary condition for the decentralization of data-processing.

The next result specifies an additional condition (a sufficient condition) which has to be satisfied in order to expand (i.e., decentralize) the information-processing structure.

Proposition 3

If the information-processing function exhibits decreasing returns to scale, then the decentralization of the information-processing structure with

$P < \lfloor N/2 \rfloor$ *processors up to* P' *processors* ($[\lfloor N/2 \rfloor \geq P' > P$) *is desirable only if the information workload* (N) *is large enough, i.e.,*

$$V > \frac{P \; \{ \; F(\frac{K}{P},\frac{L}{P})[(Y_N(P')-1] \; - \; F(\frac{K}{P'},\frac{L}{P'})[Y_N(P)-1]\}}{F(\frac{K}{P'},\frac{L}{P'}) \; - \; \frac{P}{P'}F(\frac{K}{P},\frac{L}{P})} \tag{6}$$

where

$$Y_N(P) = \left\lfloor \frac{N}{P} \right\rfloor - (\frac{N}{P} - 1) + \lceil \log_2(P + N \bmod P) \rceil \;, \tag{7}$$

and $Y_N(P)$[6] *is determined for* $N \geq 2$, $P \in \{1,2,...,\lfloor N/2 \rfloor\}$.

Proof

It follows from Proposition 1 that the minimum delay in the computation of N data items in the structure with P ($P \in \{1,2,...,\lfloor N/2 \rfloor\}$) processors is determined as

$$D_N(K,L) = \frac{\frac{N}{P} + Y_N(P) - 1}{F(\frac{K}{P},\frac{L}{P})} \;. \tag{8}$$

Thus, for any given information workload (N), it is desirable to decentralize a structure with P ($P \in \{1,2,...,\lfloor N/2 \rfloor - 1\}$) processors if, and only if, there exists $P' \in \{1,2,...,\lfloor N/2 \rfloor\}$, such that $P' > P$ and $D_{N,P'}(K,L) < D_{N,P}(K,L)$, i.e.,

$$\frac{\frac{N}{P'} + Y_N(P') - 1}{F(\frac{K}{P'},\frac{L}{P'})} < \frac{\frac{N}{P} + Y_N(P) - 1}{F(\frac{K}{P},\frac{L}{P})} \;. \tag{9}$$

Rearranging the inequality above yields

[6] Function $Y_N(P)$ has the following properties: (a) $Y_N(P) \geq 1$, for each $P \in \{1,2,\ldots,\lfloor N/2 \rfloor\}$, and (b) values of $Y_N(P)$ oscillate around certain values (but are bounded) if N increases and P is fixed (i.e., for any N, $0 < \lfloor N/P \rfloor - (N/P - 1) \leq 1$, and $\lceil \log_2 P \rceil \leq \lceil \log_2(P + N \bmod P) \rceil \leq \lceil \log_2(2P-1) \rceil$).

$$\frac{N}{P}[F(\frac{K}{P'},\frac{L}{P'}) - \frac{P}{P'}F(\frac{K}{P},\frac{L}{P})] >$$
$$> F(\frac{K}{P},\frac{L}{P})[(Y_N(P')-1] - F(\frac{K}{P'},\frac{L}{P'})[Y_N(P)-1] \ . \qquad (10)$$

Assuming that the information-processing function faces decreasing returns to scale, the expression

$$\frac{1}{P}[F(\frac{K}{P'},\frac{L}{P'}) - \frac{P}{P'}F(\frac{K}{P},\frac{L}{P})] \ , \qquad (11)$$

is positive. Dividing (10) by (11) gives (6).

Q.E.D.

It follows that decreasing returns to scale of the information-processing function is a necessary (but not sufficient) condition for the decentralization of data-processing in the decision-making systems. Once this condition is satisfied, the condition for the information workload of the decision-making sector (i.e., a large number of data items) is sufficient for the decentralized organization of data-processing. Moreover, Proposition 3 implies that if the information-processing function exhibits decreasing returns to scale, the size of the efficient information-processing structures increases with the information workload (N), i.e., it is desirable to decentralize any information-processing structure with P ($P \in P_N \setminus \{\lfloor N/2 \rfloor\}$) processors if N is sufficiently large.[7]

Thus, if the information-processing function faces non-decreasing returns to scale, then the efficient information-processing structure is always centralized (following from Proposition 2). On the other hand, if the information-processing function exhibits decreasing returns, then (following from Proposition 3) the size of the efficient structure (P^*) is determined by the number of data items processed (N) and the resources allocated to information-processing. Note, however, that in general (if the information-processing function is non-homogeneous), for any fixed N, the size of the efficient structure (P^*) could be different for various combinations of the resources allocated to information-processing.

---

[7] Note that if N increases, the left-hand side of inequality (6) increases, but the right-hand side is bounded (see footnote 6, property (b) of the function $Y_N(P)$). Thus, for any P' ($P < P' \leq \lfloor N/2 \rfloor$) and the decreasing–returns-to-scale information-processing function, there exists such N which satisfies the inequality (6).

To simplify the analysis, we restrict our considerations to homogeneous of degree h information-processing functions—i.e., we assume that for the information-processing function under study the following is true

$$\beta^h F(K,L) = F(\beta K, \beta L) \ , \tag{12}$$

where $0<\beta<1$ and $0<h<\infty$. In this case, necessary and sufficient conditions for the decentralization of information-processing structures can be formulated as follows:

Proposition 4

If the information-processing function is homogeneous of degree h*, then (for any given information workload,* N*) it is desirable to decentralize the information-processing structure with* P ($P<\lfloor N/2 \rfloor$) *processors if, and only if,*

(a) *the information processing function exhibits homogeneity of a degree less than* 1*, i.e.,* $0<h<1$*;*

(b) *there exists integer* $P' \in [P+1, \lfloor N/2 \rfloor]$*, such that*

$$N > \frac{P[Y_N(P') - 1 - (\frac{P}{P'})^h (Y_N(P) - 1)]}{(\frac{P}{P'})^h - \frac{P}{P'}} \ , \tag{13}$$

where $Y_N(P)$ *is given by* (7).

Proof

The information-processing function, F(K,L), faces decreasing returns to scale if the degree of homogeneity belongs to the interval (0,1). Thus, following from Proposition 3, if $h \in (0,1)$, then the decentralization of the structure with P processing elements is desirable if there exists an integer P' ($P' \in [P+1, \lfloor N/2 \rfloor]$) such that inequality (6) is satisfied for a given number of data processed N. Taking into account condition (12), plugging $P'=P/\beta$ ($0<\beta<1$) into (6), and rearranging yields (13).

Q.E.D.

One important implication of Proposition 4 is that, for any given information workload (N), the size of the efficient information-processing structure (P^*) depends solely upon the degree of homogeneity (h) of the information-processing function and is the same for any feasible inputs combination. Therefore, for any given number of data items processed (N), the size of the efficient structure (P*) could vary only in response to changes in the degree of homogeneity (h) of the information-processing function.

5. Organizational Restructuring of the Efficient Structures

In our simple model, the degree of homogeneity (h) of the information-processing function reflects the current status of the information-processing technology, and, consequently, it may change in response to technological progress (note, however, that not every improvement in information-processing technology changes the degree of homogeneity of the information-processing function). Therefore, the size of the efficient information-processing structures can be affected only by those technological changes in information-processing which affect returns to scale of the information-processing function (i.e., the degree of homogeneity, h).

The next result explains the pattern of changes in the size of the efficient information-processing structures of decision-making systems in response to changes in returns to scale of the information-processing function.

Proposition 5

The size of the efficient information-processing structures (P*) *of decision-making systems decreases or doesn't change (increases or doesn't change) if the degree of homogeneity of the information-processing function increases (correspondingly, decreases).*

Proof

The minimum delay in the summation of N data items that can be attained by the structure with P processors (assuming that the duration of each individual operation equals 1) can be specified as[8]

$$C_N(P) = \left\lfloor \frac{N}{P} \right\rfloor + \lceil \log_2(P + N \bmod P) \rceil \, , \tag{14}$$

Thus, for any given combination of inputs (K,L), the size of the efficient structures can be derived from the following optimization problem:

$$\underset{P}{Min} \frac{C_N(P)}{F(\frac{K}{P}, \frac{L}{P})}, \quad P \in \{1, 2, \ldots, \lfloor N/2 \rfloor\}.$$

[8] See Radner (1992 and 1993) or Gibbons and Rytter (1988).

(15)

If the information-processing function is homogeneous of degree h (h>0), i.e., $(1/P)^h F(K,L) = F(K/P,L/P)$, then the optimization problem above is equivalent to

$$\min_P \; P^h C_N(P), \quad P \in \{1,2,\ldots,\lfloor N/2 \rfloor\}, \tag{16}$$

and, consequently, for any given N, the size of the efficient structure (P^*) depends only upon the degree of homogeneity of the information-processing function and doesn't depend on the combination of inputs to information-processing.

To show the pattern of changes in the size of the efficient structures (P^*) in response to changes in the degree of homogeneity (h) of the information-processing function, suppose that $C_N(P)$ can be approximated by the continuous, twice differentiable function $\hat{C}_N(P)$[9], where $P \in [1,\lfloor N/2 \rfloor]$, such that $\hat{C}_N(P)>0$, $d\hat{C}_N(P)/dP<0$, and $d^2\hat{C}_N(P)/dP^2>0$, for each $P \in [1,\lfloor N/2 \rfloor]$. As an example, the function $C_{N=500}(P)$ and its continuous approximation $\hat{C}_{N=500}(P)=N/P+\log_2 P$ are presented in Figure 5.1.

Define continuous function $R_N(P,h)=P^h\hat{C}_N(P)$, where $P \in [1,\lfloor N/2 \rfloor]$, $h \in (0,\infty)$, $R_N \in (\hat{C}_N(\lfloor N/2 \rfloor),\infty)$, and consider the following optimization problem:[10]

$$\min_P \; R_N(P,h=h^\circ), \quad P \in [1,\lfloor N/2 \rfloor], \tag{17}$$

where h° ($h^\circ \in (0,\infty)$) is a parameter. The first order condition can be represented as follows:

$$P^{h^\circ - 1} h^\circ \hat{C}_N(P) + p^{h^\circ} \hat{C}'_N(P) = 0, \tag{18}$$

dividing on P^{h° yields

$$h^\circ \frac{\hat{C}_N(P)}{P} - (-\hat{C}'_N(P)) = 0 . \tag{19}$$

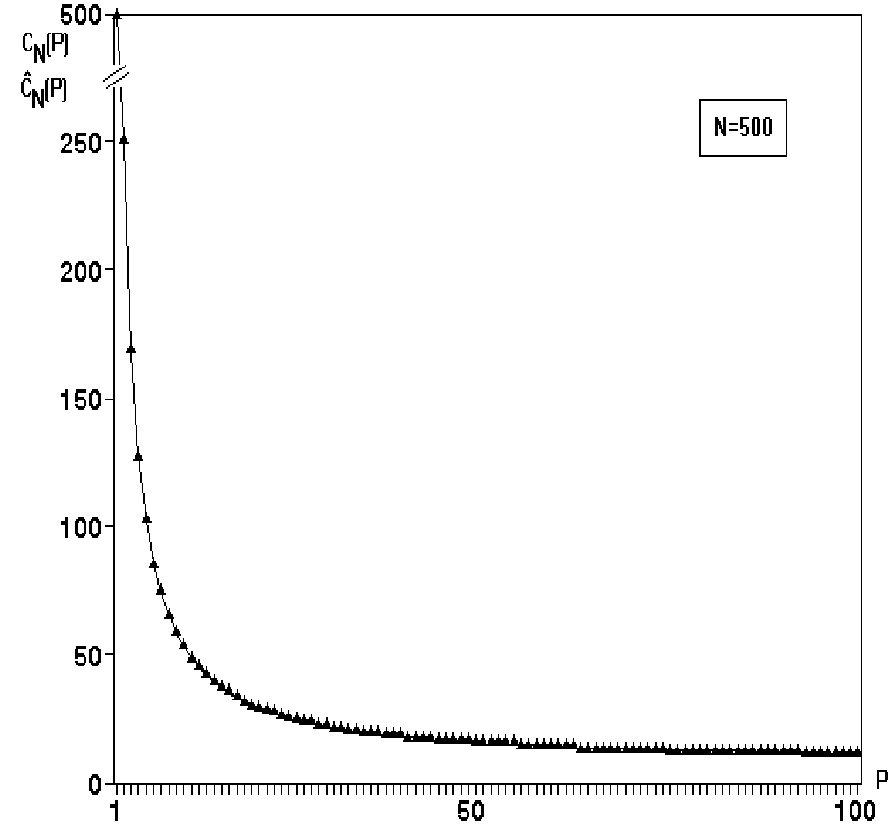
---

[9] In particular, Radner (1993) argued that the function of the form $\hat{C}_N(P) = N/P + \log_2 P$, gives a good approximation of $C_N(P)$, if N, P, and N/P are all large.

[10] Note that $R_N(P,h)$ is continuously differentiable in P and h.

Note that the equation above does not necessarily have a solution in $P \in (1, \lfloor N/2 \rfloor)$, and, consequently, the interior extremum does not necessarily exist. If there is no such P in $(1, \lfloor N/2 \rfloor)$ which satisfies the equation (19), then

Figure 5.1
The Minimum Delay $C_N(P)$ and Its Continuous Approximation $\hat{C}_N(P)$
(Information Workload: N=500, Duration of a Single Operation: d=1)



the solution to the optimization problem (17) is in one of the corners of the interval $[1,\lfloor N/2\rfloor]$, that is, $P^* = \arg\{\min R_N(P)$, where $P\in\{1,\lfloor N/2\rfloor\}\}$. On the other hand, if the function $\hat{C}_N(P)$ can be represented as $\hat{C}_N(P)\equiv S(N)P^{-h}$ (where S(N) denotes any function of N), then the minimum value of the objective function does not depend on P, i.e., the objective function is constant in P. In all other cases, if the solution to (19) exists in the interval $(1,\lfloor N/2\rfloor)$, i.e., $P^*\in(1,\lfloor N/2\rfloor)$, then it corresponds to the minimum of the objective function, $R_N(P)$, if the second derivative of $R_N(P)$ with respect to P at $P=P^*$ is positive (i.e., $d^2R_N(P,h=h°)/dP^2|_{P=P^*}>0$).[11] To show the pattern

[11] Note that $h\hat{C}_N(P)/P$ and $-\hat{C}_N'(P)$ in (19) are decreasing and strictly convex functions of P. Consequently, these functions (a) cannot cross each other (in this case neither interior minimum nor interior maximum exist), (b) can cross once, or (c) can cross at most twice. In the case (b), the objective function has a maximum or a minimum in the point of the intersection. In the case (c), in one intersection point there is a maximum and in the second one the minimum of the objective function. However, in any case, i.e., (a) or (b), if the interior

of changes in the size of the efficient structures, in response to changes in parameter h°, represent the first order condition as G(h°,P^*(h°))=0, and define function

$$,P^*(h^\circ)) = \frac{G(h^\circ,P^*(h^\circ))}{P^{*\ h^\circ}} = \frac{h^\circ \hat{C}_N(P^*)}{P^*} + \hat{C}'_N(P^*) \qquad (20)$$

Function Q(h°,P^*(h°)) is continuously differentiable with respect to h°∈(0,1) and P^*∈(1,⌊N/2⌋). Consequently, by the implicit function theorem, the first derivative of P^*(h°) with respect to h° is

$$\frac{dP^*}{dh^\circ} = -\frac{\frac{\partial Q}{\partial h^\circ}}{\frac{\partial Q}{\partial P^*}} . \qquad (21)$$

The partial derivative of Q(h°,P^*(h°)) with respect to h° equals

$$\frac{\partial Q}{\partial h^\circ} = \frac{\hat{C}(P^*)}{P^*} , \qquad (22)$$

and it is positive for all P^*∈(1,⌊N/2⌋). The partial derivative of Q(h°,P^*(h°)) with respect to P^* equals

$$\frac{\partial Q}{\partial P^*} = \frac{\frac{\partial G(h^\circ,P^*(h^\circ))}{\partial P^*}}{P^{*\ h^\circ}} - h^\circ P^{-(1+h^\circ)} G(h^\circ,P^*(h^\circ)) . \qquad (23)$$

The second term in the expression above equals zero (because G(h°,P^*(h°))=0). Moreover, ∂G(h°,P^*(h°))/∂P^* = d^2R_N(P,h=h°)/$dP^2|_{P=P^*}$ and is positive if P^* is the interior minimum. Thus, the second term disappears and the first term in (23) is greater than zero. Consequently, dP^*(h°)/dh°<0, i.e., the size of the efficient information-processing structure (P^*) decreases if the degree of homogeneity (h) of the information-processing function increases.

This would be a general pattern of changes in the size of the efficient structures in response to changes in the degree of homogeneity of the information-processing function, if C(P) were a continuous function of P. However, C(P) is determined only for integer values of P (i.e., for P∈{1,2,...,⌊N/2⌋}). In this case, small changes in h could not affect the size

---

minimum exists, then it is unique.

of the efficient structures (P^*). To show it, assume that for a certain information workload (N) and the degree of homogeneity of the information-processing function $h=h_0$ ($h_0 \in (0,1)$), the structure with $\lfloor N/2 \rfloor$ processors is efficient, i.e., for any $P \in \{2,...,\lfloor N/2 \rfloor - 1\}$ there exists $P' = \lfloor N/2 \rfloor$ such that condition (13) is satisfied.[12] Note that the right-hand side of the inequality (13) increases with h when $h \in (0,1)$. Thus, if $h=h_0$ rises a bit, say to h_1 ($h_1 > h_0$, but the difference $h_1 - h_0$ is close to zero), then inequality (13) could still be satisfied for all $P \in \{1,...,\lfloor N/2 \rfloor - 1\}$, and the structure with $P' = \lfloor N/2 \rfloor$ processors would still be efficient. However, if the increase in h is significant, say to h_2 close to 1, then the right-hand side of inequality (13) rises considerably,[13] and this inequality is not satisfied for at least one $P < P' = \lfloor N/2 \rfloor$, i.e., there exists a structure with $P < P' = \lfloor N/2 \rfloor$ processing elements that gives a better result than a fully decentralized one. Consequently, a substantial increase in the degree of homogeneity of the information-processing function reduces the size of the efficient structure.

Consider now the case when, for given values of N and $h=h_0$, the size of the efficient structure P^* equals P_x ($P_x \in \{2,...,\lfloor N/2 \rfloor - 1\}$)). To show the pattern of changes of P^* in response to changes in h, define the function

$$_V(P,h) = \underset{P'=P+1,...,\lfloor N/2 \rfloor}{Min} \frac{P[Y_N(P') - 1 - (\frac{P}{P'})^h (Y_N(P) - 1)]}{(\frac{P}{P'})^h - \frac{P}{P'}} \tag{24}$$

where $P \in \{1,2,...,\lfloor N/2 \rfloor - 1\}$, $h \in (0,1)$, and $Y_N(P)$ is given by (7).

Proposition 3 implies that if P_x is the size of the efficient structure, then the following conditions have to be satisfied:

$$N \le V_N(P_x, h_0), \tag{25}$$

and

$$N > V_N(P_x - s, h_0), \tag{26}$$

where $s = 1,2,...,P_x - 1$.

---

[12] See Proposition 4.

[13] In the extreme case, if h goes to 1, then the right-hand side of the inequality (13) goes to infinity.

It means that it is always desirable to expand all structures with processors less than P_x (up to P_x), but it is surely not desirable to expand the structure with P_x processors (i.e., the delay in information-processing attained by any bigger structure is greater than the delay attained by the structure with P_x processors). Note that function $V_N(P,h)$ increases with h when $h\in(0,1)$. Thus, if the degree of homogeneity of the information-processing function increases, then the values of V_N also increase for each particular P ($P\in\{1,2,...,\lfloor N/2\rfloor-1\}$). It implies that if h rises (i.e., departures from h_0), then conditions (25) and (26) could be violated. However, if the increase in h is relatively small, say to h_1 ($h_1>h_0$, but $\Delta h_1=h_1-h_0$ is close to zero), then the size of the efficient structure remains unchanged, i.e., the following inequalities are still satisfied: $N\leq V_N(P_x,h_0+\Delta h_1)$ and $N>V_N(P_x-s,h_0+\Delta h_1)$, where $s=1,2,...,P_x-1$. On the other hand, if h increases considerably, say to h_2 ($h_2>h_0$, but $\Delta h_2=h_2-h_0$ is significant) then the condition $N>V_N(P_x-s,h_0+\Delta h_2)$, $s=1,2,...,P_x-1$, would be violated at least for one s (i.e., P_x would no longer be the size of the smallest structure which should not be expanded). It implies that only a significant increment in the degree of homogeneity (h) of the information-processing function reduces the size (P^*) of the efficient structures.

Q.E.D.

It should be clear that the number of levels in the efficient processing network ("skip-level reporting") is proportional to its size, P^* (note, however, that the number of levels could be the same for different numbers of processors). Consequently, if the number of processors in the efficient structure decreases (increases), then the depth of the hierarchy either decreases or does not change (correspondingly, increases or does not change). Since the size of the efficient structures changes only in response to technological improvements that change the degree of homogeneity of the information-processing function, the flattening hierarchical information-processing structures can be explained only by those technological changes which increase the degree of homogeneity of the information-processing function.

For example, if the information-processing function is of the form $F(K,L)=A(K^{\alpha}L^{\beta})^{\mu}$, where A, α, β, and μ are all positive parameters (the function is homogeneous with the degree of homogeneity $h=(\alpha+\beta)\mu$), then a neutral technological change could be represented by an increase in A or in μ. An increase in A leaves the returns to scale unchanged while an increase in μ increases those returns (raises the degree of homogeneity, h). Consequently, an increase in A does not affect the form of the efficient hierarchical structures, whereas an increase in μ might result in a reduction (and flattening) of efficient information-processing structures.

Note, however, that changes in information-processing technology which affect returns to scale do not automatically imply changes in the size (and the depth) of the efficient hierarchies. To reduce efficient information-processing structures, changes in the degree of homogeneity of the information-processing function have to be large enough.

6. Conclusion

The analysis of data-processing in decision-making presented in this paper shows that organizational forms of the efficient information-processing structures are determined by properties of the information-processing function and the information workload of the system. In particular, it has been shown that the decentralized (hierarchical) organization of information-processing in decision-making can be observed if, and only if, the information-processing function exhibits decreasing returns to scale and the information workload of the decision-making system is large enough. If the information-processing function is homogeneous, then the size of the efficient structures is determined solely by the degree of homogeneity of the information-processing function and the information workload of the decision-making system. Furthermore, if the information workload remains unchanged, then the size of the efficient structures is inversely related to the degree of homogeneity of the information-processing function and, consequently, efficient information-processing structures can be reduced only by those technological improvements which affect economies of scale in information-processing, i.e., those which increase the degree of homogeneity of the information-processing function. It should be emphasized however, that small changes in the degree of homogeneity of the information-processing function might not alter the size of the efficient processing networks.

The results above have been derived based on the set of simplifying assumptions concerning information-processing in the decision-making sector of the firm. The analysis has been restricted to the simplest case of associative operations in homogeneous, one-shot structures with perfectly rational processors (it has been assumed that processors always perform according to specification, i.e., cannot make errors). However, similar results could be obtained using various modifications of the dynamic parallel-processing model of associative computation in which, for example, structures with heterogeneous processors are considered (Cukrowski, 1995a), the possibility of errors in associative operations is introduced (Cukrowski, 1995b), or computational processes in repeating structures are studied (Radner and Van Zandt, 1992). These models make the analysis more difficult, but the general result concerning the pattern of organizational restructuring in response to changes in information-processing technology remains the same.

It should be also clear that although the particular subject of the analysis presented in the paper was the decision-making system inside the firm, the model developed concerns organizations that may be different from firms. It is relevant for decision-making systems in large corporations and nonprofit organizations as well as in public administration. Consequently, organizational changes in these systems can be justified exclusively by those technological improvements that affect returns to scale of the information-processing function.

REFERENCES

Aoki, M. 1986. "Horizontal vs. Vertical Information Structure of the Firm," *American Economic Review* 76: 971–83.

Bolton, P. and M. Dewatripont. 1994. "The Firm as a Communication Network," *The Quarterly Journal of Economics* Vol. CIX, No. 4: 809–40.

Business Week. 1989. "Special Report: A New Era for Management" (April 25): 50–86.

Chandler, A. 1966. *Strategy and Structure*. New York: Doubleday.

Cukrowski, J.A. 1995a. "An Economic Analysis of Numerical Data Processing in the Firm," *CERGE-EI Working Paper* 79: 31.

———. 1995b. "The Hierarchical Organization of Management in the Firm: Economic Reasons and Efficient Structures," *CERGE-EI Working Paper* 81: 29.

Gibbons, A. and W. Rytter. 1988. *Efficient Parallel Algorithms*. Cambridge: Cambridge University Press.

Kennedy, P. W., 1994. "Information Processing and Organization Design," *Journal of Economic Behavior & Organization* 25: 37–51.

Marschak, T. and R. Radner. 1972. *Economics Theory of Teams*. New Haven, CT: Yale University Press.

Milgrom, P. and J. Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *American Economic Review* 80,3: 511–28.

Radner, R. 1992. "Hierarchy: The Economics of Managing," *Journal of Economic Literature* 30: 1382–1415.

———. 1993. "The Organization of Decentralized Information Processing," *Econometrica* 61: 1109–46.

Radner R. and T. Van Zandt. 1992. "Information Processing in Firms and Return to Scale," *Annales d'Economie et de Statistique* 25/26: 265–98.

———. 1993. "Information Processing and Returns to Scale in Statistical Decision Problems," Murray Hill, NJ: AT&T Bell Laboratories.

Schein, E.H. 1989. "Reassessing the Divine Rights of Managers," *Sloan Management Review* (Winter): 36–43.