EFFICIENT STRUCTURES OF THE HIERARCHICAL ORGANIZATION OF MANAGEMENT IN THE FIRM

ANTON ABDULBASAH KAMIL
School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.
e-mail: anton@cs.usm.my

Abstract. In this paper, economic explanation of the hierarchical organization of management in the firm is provided based on the analysis of organizational aspects of data processing in decision making. The paper shows that decreasing returns to scale in information processing is a necessary condition for hierarchical organization of management in business firms. However, decentralization of decision-making is desirable only if an additional (sufficient) condition is satisfied, i.e., if the information workload of the decision-making sector is sufficiently large. Moreover, the paper shows that specific features of human information processing in the firm such as disagreement about the goals of data analysis or the possibility of random errors do not imply a hierarchical organization of management, but could change forms of the efficient decision-making structures. Finally, contrary to the results recently presented in economic literature, the paper shows that in the firm, unlike in computer systems, there is no unique architecture for the efficient information-processing structures, but a number of various efficient forms can be observed.

Keywords: Information-processing, hierarchy, organization of the firm, decentralization,

1. INTRODUCTION

In classical microeconomic theory, the firm is viewed as a 'black box' transforming inputs into outputs according to a rule described by a production function. In fact, firms have complex multidivisional, usually hierarchical, structures which determine their economic performance. Especially large corporations are widely perceived to be organized hierarchically.

The purpose of the present paper is to provide an economic justification of the hierarchical organization of information processing in the management of the firm, to explain the reasons why information processing in the firm differs from that in an idealized computer, and to show how these differences affect the efficient organizational forms of information-processing structures.

The simplest example of the transformation of data from the environment into decisions is a linear decision rule, where the value of the linear function:

$$\sum_{i=1}^{N} c_i x_i ,$$

is computed (c_i is a coefficient of conversion to a common unit, and x_i is a numerical data item, $i = 1, 2, \dots, N$). 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 (Aoki, 1986; Radner and Van Zandt, 1992 and 1993). Based on the analysis of computational processes for the purpose of predicting demand in the firm, Radner and Van Zandt (1992) show that the values of the decisions made according to the linear decision rule depend on the quality of the result computed, measured by the delay in information processing.

2. THE MODEL OF INFORMATION PROCESSING IN DECISION MAKING

Consider the decision-making sector in the firm in which decisions are made based on data analysis. The value of the decisions, and, consequently, the value of the computational service provided depend on how good the resulting decisions are compared to how good they would be without the service. Assuming that the linear decision rule is used in the decision-making process, the value of the computational service depends only upon the delay in data analysis (more precisely they are inversely proportional to the delay in information processing).

To simplify the analysis of the delay in the computational process, assume that the linear decision rule requires summations of cohorts of N items of data (conversion to a common unit is not required, i.e., $c_i = 1$, for $i = 1, 2, \ldots, N$), and that the decision-making system considered works in a one-shot regime, i.e., delays\between subsequent cohorts of data coming into the system are greater (or at least equal) to the time of a single cohort processing (it ensures that queues of data in the information-processing structure cannot arise).

The computational process in the decision-making sector of the firm as in idealized computer, i.e., assume that each processing element (a computational center) is modeled as a processor which contains an infinite memory where data are stored (called a buffer) and a register where summations are made. Each processor has a limited capacity (i.e. a maximum computational power), in that there is a maximum number of operations it can compute per unit of time. In business firms, however, the computational power of each processing element depends upon the capital and labor allocated to it. The relationship between the resources allocated to the processing element and the number of operations it can compute in a unit of time is determined by the existing technology of information processing, and can be written in functional form as information-processing function $F(k,l)$: $R_+ x R_+ \rightarrow R_+$, where $F(k,l)$ is continuous, twice differentiable and strictly concave in k and l. Consequently, the duration of a single operation (d) is also a function of the capital (k) and labor (l) employed in the processing element ($d(k,l) = 1/(F(k,l))$. In any information-processing structure, the delay in the summation of N items of data (D_N) is proportional to the duration of individual operations, and, consequently, is a decreasing function of the resources allocated to the computational structure, i.e., $\delta D_N(K,L)/\delta K < 0$ and $\delta D_N(K,L)/\delta L < 0$.

Assuming that data items are not costly, the total cost of the computational process, $C(K,L)$, is determined as

$$C(K,L) = r\,K + w\,L,$$

where,

$w\,L$ = cost of labor involved in the computation (w denotes the price of labor);

$r\,K$ = cost of capital (r is the price of capital).

The consideration above implies that the objective of the firm in decision making is to maximize the difference between the value of the computational service, $V(D_N(K,L))$, and the cost of the resources used in computation, $C(K,L)$.

In the model analyzed in this paper, neither the processors nor the computational centers, but the capital and labor allocated to them are considered as inputs to information processing. Thus, in the framework of the model under study, one can say that if the capital (K) and labor (L) allocated to data processing as well as the information workload (N) are multiplied by the same constant, say $\alpha > 1$ (or $0 < \beta < 1$), then the information-processing system faces:

1. increasing returns to scale, if the quality of the result computed increases, i.e., $D_N(K,L) > D_{\alpha N}(\alpha K, \alpha L)$, (correspondingly, decreases, if $0 < \beta < 1$, i.e., $D_N(K,L) < D_{\beta N}(\beta K, \beta L)$);
2. constant returns to scale, if the quality of the result computed does not change, i.e., $D_N(K,L) = D_{\alpha N}(\alpha K, \alpha L)$, (correspondingly, $D_N(K,L) = D_{\beta N}(\beta K, \beta L)$);
3. decreasing returns to scale, if the quality of the result computed decreases, i.e., $D_N(K,L) < D_{\alpha N}(\alpha K, \alpha L)$, (correspondingly, increases, if $0 < \beta < 1$, i.e., $D_N(K,L) > D_{\beta N}(\beta K, \beta L)$).

3. THE DECENTRALIZATION OF INFORMATION-PROCESSING IN THE FIRM

Assume that the amounts of resources allocated to information processing are fixed (the cost of computation is fixed as well). In this case the objective of the firm is to organize data analysis in decision making in a\way in which it maximizes the values of the decisions made based on the computational service ($V(D_N)$), or in other words, in which it minimizes the delay in information processing (D_N).

Suppose, in the beginning, that all the resources used in decision-making are allocated to a single computational center (($P = 1 = 2^0$). If the technology of information processing is such that $D_N(K,L) > D_{\beta N}(\beta K, \beta L)$, i.e., if the firm faces decreasing returns to scale in information processing, then a better quality result can be computed if the inputs to information processing (K and L), as well as the information workload (N), are reduced. Thus, sums of $N/2$ data items can be computed in two separate in two separate computational centers (with equally divided resources, $\beta = 1/2$ with less delay than the sum of N data items in the original structure. For simplicity it is assumed that the number of data items processed (N) is such that ($N \bmod 2^m$) = 0, for $m = 0,1, \ldots, \log_2(N/2)$. However, a linear decision rule requires the sum of N data items. Consequently, the computational centers have to be connected and one additional operation has to be made in order to summarize the partial results computed. Therefore, the decentralization of information-processing is desirable only if $D_N(K,L) > D_{N/2}(K/2, L/2) + d(K/2, L/2)$, where $d(K/2, L/2)$ is the duration of the last operation, i.e., if

$$Nd(K,L) = \frac{N}{F(K,L)} > \frac{N}{2} d\left(\frac{K}{2}, \frac{L}{2}\right) + d\left(\frac{K}{2}, \frac{L}{2}\right) = \frac{\frac{N}{2}+1}{F\left(\frac{K}{2}, \frac{L}{2}\right)}$$

The inequality above is satisfied if

$$N > \frac{F(K,L)}{F\left(\frac{K}{2}, \frac{L}{2}\right) - \frac{1}{2} F(K,L)}$$

If the decentralized computational system (with $P = 2^1$ processing elements) faces decreasing returns to scale in information processing, then the delay in the summation of $N/2$ data items in the structure, in which the resources $K/2$ and $L/2$ are allocated to $P = 2^1$ processing elements, is smaller than the delay in the computation of the sum of N data items in the structure with $P = 2^1$ processing elements and the entire resources. If the information workload (N) and the resources (K, L) are divided in two equal parts, and sums of $N/2$ items of data are computed in two identical structures, then the top-level computational centers of these structures have to be connected and one additional operation has to be made in order to add the partial sums. The duration of this operation equals $1/F(K/2^2, L/2^2)$. Consequently, the decentralization of the structure with $P = 2^1$ processing elements is desirable only if

$$\frac{\frac{N}{2}+1}{F\left(\frac{K}{2}, \frac{L}{2}\right)} > \frac{\frac{N}{2^2}+1}{F\left(\frac{K}{2^2}, \frac{L}{2^2}\right)} + \frac{1}{F\left(\frac{K}{2^2}, \frac{L}{2^2}\right)} = \frac{\frac{N}{2^2}+2}{F\left(\frac{K}{2^2}, \frac{L}{2^2}\right)},$$

i.e., when

$$N > \frac{2\left[2F\left(\frac{K}{2}, \frac{L}{2}\right) - F\left(\frac{K}{2^2}, \frac{L}{2^2}\right)\right]}{F\left(\frac{K}{2^2}, \frac{L}{2^2}\right) - \frac{1}{2} F\left(\frac{K}{2}, \frac{L}{2}\right)}.$$

The first two steps of the decentralization process are presented in fig. 1

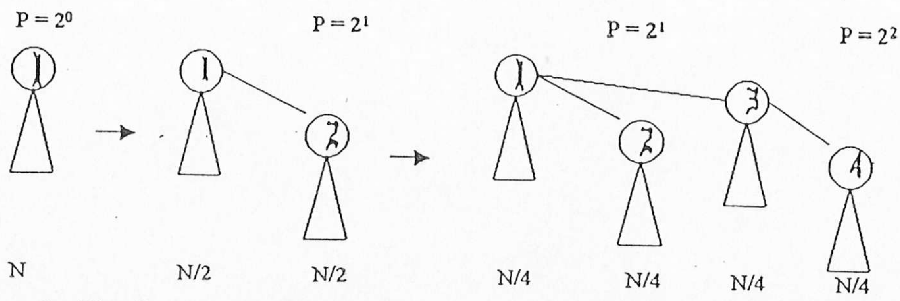

Figure 1 : The first steps of the decentralization process (circles denote computational centers, and triangles represent the information workload).

Decentralization can be continued (if the corresponding conditions are satisfied) until the number of the processing elements in the structure equals P = *N*/2 (P is bounded because at least two data items have to be assigned to each computational center).

One can see that the decentralization process produces so-called 'skip-level reporting' structures. Hierarchical forms of such kind contain P (where P is a power of 2) processing elements organized in hierarchical (multilevel) formations where each computational center has one immediate subordinate at every lower level of the hierarchy. 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.

After m $(m \le \log_2(N/2))$ steps of the decentralization process, the skip-level reporting structure contains $P = 2^m$ processing elements, and $N/2^m$ data items are assigned to each of them. The delay in the summation of N data items in such a structure (if N is a multiple of P) is determined as

$$D_N(K,L) = \left(\frac{N}{P} + \log_2 P\right) d\left(\frac{K}{P}, \frac{L}{P}\right) = \frac{\frac{N}{2^m} + m}{F\left(\frac{K}{2^m}, \frac{L}{2^m}\right)}$$

Therefore, if the system with $P = 2^m \, (m < \log_2(N/2))$ processing elements faces decreasing returns to scale in information processing (it is a necessary condition), and if

$$\frac{\frac{N}{2^m} + m}{F\left(\frac{K}{2^m}, \frac{L}{2^m}\right)} > \frac{\frac{N}{2^{m+1}} + m}{F\left(\frac{K}{2^{m+1}}, \frac{L}{2^{m+1}}\right)} + \frac{1}{F\left(\frac{K}{2^{m+1}}, \frac{L}{2^{m+1}}\right)} = \frac{\frac{N}{2^{m+1}} + m + 1}{F\left(\frac{K}{2^{m+1}}, \frac{L}{2^{m+1}}\right)},$$

then it is desirable to expand the information-processing sector and allocate the resources to $P = 2^{(m+1)}$ computational centers.

The inequality above is satisfied if the number of data items processed (N) is such that

$$N > \frac{2^m\left[(m+1)F\left(\frac{K}{2^m}, \frac{L}{2^m}\right) - mF\left(\frac{K}{2^{m+1}}, \frac{L}{2^{m+1}}\right)\right]}{F\left(\frac{K}{2^{m+1}}, \frac{L}{2^{m+1}}\right) - \frac{1}{2}F\left(\frac{K}{2^m}, \frac{L}{2^m}\right)}.$$

The expression above describes a sufficient condition for decentralization of the skip-level reporting structure with $P = 2^m$ processing elements (m = 0, 1, ..., log_2 (N/2)-1).

As already emphasized, despite the formal proof of efficiency, it is difficult to find such organizational forms of data processing in real firms. The architecture of information-processing structures in enterprises differs from the one described above, because

1. in real firms computations are usually much more complicated than a simple addition of numerical data, and
2. in the real firms, unlike in the computer systems, people (not electronic processors) form the information-processing structures.

4. IMPLICATIONS OF HUMAN INFORMATION-PROCESSING ON ORGANIZATIONAL FORMS OF DECISION MAKING IN THE FIRM.

To clarify the statement above, consider the process of selecting the best project (out of N projects submitted) in a team of P decision makers. Without loss of generality, assume that project n ($n = 1, 2, \ldots, N$) is fully characterized by the value of one (aggregated) numerical parameter, Q_n, determined based on the analysis of the entire project. Therefore, one of the projects considered, say n^*, should be selected as the best one if $n^* = \arg\{\min |Q^*-Q_n|\}$, where $Q_n|$ is a numerical characteristic of the project n ($n = 1, 2, \ldots, N$), and Q^* is an attribute of the project wanted by the entrepreneur (the best for the firm).

The values of the decisions decrease with the delay in information processing. Consequently, in order to reduce the time of data analysis, the process of selecting the best project can be decentralized. However, if each processing element of the decentralized structure represents a single member of the decision-making team, then each individual decision maker, p ($p = 1, 2, \ldots, P$), computes (and compares) the absolute values of the differences: $|Q_p^*-Q_{p,n}|$, where $Q_{p,n}$ denotes his subjective evaluation of the project n ($n \in N_p, N_p$ is a set of projects analyzed by member p), and Q_p^* is an attribute of the project considered as the best by decision-maker p. Divergences between decision makers in attitudes, in perceptual abilities, or in their ability to concentrate, and also some random factors such as, for instance, emotions, frustrations or stresses, imply that subjective evaluations of the same project, say n, by different members of the team, $Q_{p,n}$ ($p = 1, 2, \ldots, P$), could not be the same, i.e., $Q_{1,n} \neq Q_{2,n} \neq \ldots \neq Q_{P,n} \neq Q_n$. Moreover, the possibility of misinterpreting the target of data analysis (i.e., of the goal of the entrepreneur, Q^*) and divergences among the members' individual goals in information processing imply that the understanding of which is the best project could be different for each individual decision maker, i.e., $Q_1^* \neq Q_2^* \neq \ldots \neq Q_P^* \neq Q^*$. Consequently, if all the projects submitted would be considered by all the members of the decision-making team, then each decision maker could choose a different project (also different from the project that would be selected by the entrepreneur). Therefore, the decentralization of the process of project selection in the firm implies that the result of data analysis could be determined with error, measured by the absolute value of the difference between the numerical characteristics of the project wanted by the entrepreneur (Q^*) and the project selected (Q_n^*), $|Q^*-Q_n^*|$.

To represent the divergences between the members of the information-processing team, and to describe the possible variability in subjective evaluations of the information analyzed in the framework of the dynamic parallel processing model of associate computation, assume that decision-makers (i.e., the processing elements) do not make errors in the evaluations of the projects analyzed (i.e., $Q_{1,n} = Q_{2,n} = \ldots = Q_{P,n} = Q_n, n = 1,2,\ldots,N$), but each decision maker p ($p = 1, 2, \ldots, P$) computes results according to his individual understanding of the goal of the analysis, Q_p^* (the possibilities of random mistakes in evaluations of projects can be represented as random shifts in Q_p^*).

Thus, a numerical characteristic of the project considered as the best by the members of the team $(Q_1^* \neq Q_2^* \neq \ldots \neq Q_P^* \neq Q^*)$ can be described by the random variables distributed around a numerical characteristic of the project wanted by the entrepreneur, Q^*. For the sake of simplicity assume that this distribution is normal, with mean Q^* and variance σ^{*2}. Assuming that all the projects submitted are not identical, but all of them satisfy more or less the expectations of the entrepreneur, we can presume that their characteristics, Q_n ($n = 1, 2,, \ldots, N$), are distributed around the numerical characteristic of the project wanted by the entrepreneur (Q^*). For the sake of simplicity, assume that this distribution is normal, with mean Q^* and variance σ^2. The random factors in data analysis imply that the selection process should be organized in the decentralized structure (in order to minimize the delay in information processing) which minimizes the expected value of the error in data analysis, $E = \mathrm{E}(|Q^* - Q_n^*|)$, where E denotes the operator of expectation.

If Q (n = 1, 2, ..., N) and Q_p^* (p = 1, 2, ..., P) are normally distributed random variables, with mean Q^* and variances σ^2 and σ^{*2}, respectively, then the random variable characterizing the project selected (Q_n^*) in the arbitrary decision-making structure can be represented as follows:

$$Q_{n^*} = \sum_{n=1}^{N} Q_n P_n,$$

where P_n denotes the probability that the project n will be selected (n = 1, 2, ..., N). This implies that, in an arbitrary information-processing structure, the random variable Q_{n^*} is distributed normally, with mean Q^* and variance $\sigma^2_{n^*}$. Moreover, the random variable $Q^* - Q_{n^*}$ is normally distributed, with zero mean and variance $\sigma^2_{n^*}$. Consequently, the expected value of the error in data analysis can be determined as

$$E(|Q^* - Q_{n^*}|) = \int_0^\infty x f_{|Q^* - Q_{n^*}|}(x)dx = \int_0^\infty \frac{2x}{\sigma_{n^*}\sqrt{2\pi}} e^{-\frac{x^2}{\sigma^2_{n^*}}} dx = -\frac{\sigma_{n^*}}{\sqrt{2\pi}} e^{-\frac{x^2}{\sigma^2_{n^*}}} \Bigg|_0^\infty = \frac{\sigma_{n^*}}{\sqrt{2\pi}}.$$

This means that the information-processing structure which minimizes the expected value of the error in data analysis also minimizes the variance of the random variable Q_{n^*}. In an arbitrary information-processing structure, this variance can be computed as

$$\sigma^2_{n^*} = \sigma^2 \sum_{n=1}^{N} P_n^2,$$

where P_n denotes the probability that the project n (n = 1, 2, ..., N) will be selected.

The values of the probabilities P_n^* (n = 1, 2, ..., N) minimizing variance $\sigma^2_{n^*}$, and, consequently, minimizing the expected value of the error in data analysis, can be determined by finding the solution to the following optimization problem:

$$\min_{P_1, P_2, \ldots, P_N} \sigma^2 \sum_{n=1}^{N} P_n^2$$

such that

$$\sum_{n=1}^{N} P_n = 1.$$

The probabilities P_n^* (n = 1, 2, ... , N) equal $P_1^* = P_2^* = \ldots = P_N^* = 1/N$ and consequently, the minimum expected value of the error in data analysis equals $E_{\min} = \sigma/(2\pi N)^{1/2}$.

REFERENCES

Aoki, M., 1986, Horizontal vs. Vertical Information Structure of the Firm, *American Economic Review*, 76, 971-983.

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

Radner, R., 1993, The Organization of Decentralized Information Processing, *Econometrica*, 61, 1109-1146.

Radner, R., and T. Van Zandt., 1992, Information Processing in Firms and Return to Scale, *Annales d'Economie et de Statistique*, 25/26, 265-298.

Radner, R., and T. Van Zandt., 1993, *Information Processing and Returns to Scale in Statistical Decision Problems*, Murray Hill, New Jersey: AT&T Bell Laboratories.