Universität Bielefeld/IMW

Working Papers
Institute of Mathematical Economics

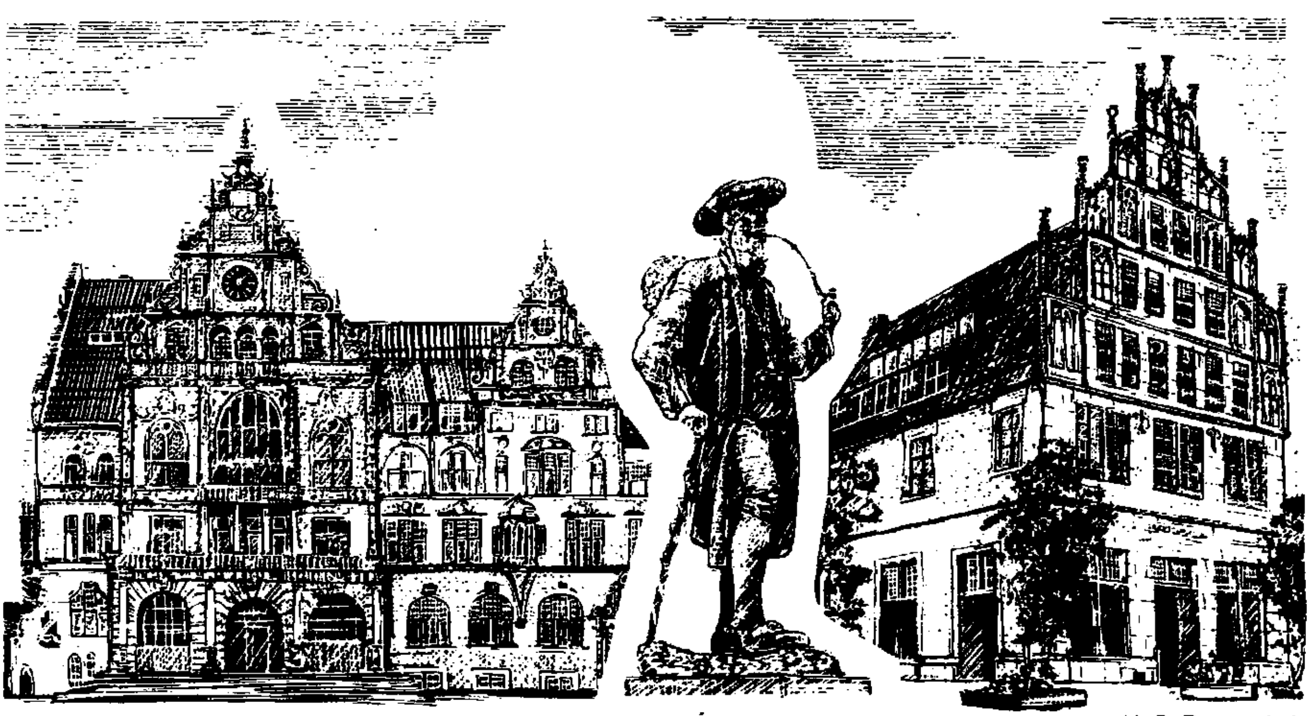
Arbeiten aus dem
Institut für Mathematische Wirtschaftsforschung

Nr. 87

Hans-W. Gottinger

An Information-Theoretic Approach to Large Organizations

September 1979



H. G. Bergenthal

Institut für Mathematische Wirtschaftsforschung
an der
Universität Bielefeld
Adresse/Address:
Universitätsstraße
4800 Bielefeld 1
Bundesrepublik Deutschland
Federal Republic of Germany

An Information-Theoretic Approach to Large Organizations

Hans W. Gottinger

1. INTRODUCTION

Organization theory is an elusive subject, being of interest to diverse disciplines: sociology, psychology, human engineering and economics. But past work has been mostly qualitative, and the originator of classical organization theory is usually considered to be MAX WEBER/1922/. Modern approaches to organization notably MARCH and SIMON/1958/, distinguish a traditional and a modern form of organization theory. The traditional version appears to be characterized by the fact that it treats organization members as fully cooperative instruments while the latter accords to them preferences and aspirations that may not coincide with those of the organization.

This paper here also treats organization members as being fully cooperative, and therefore may be considered a quantitative version of the traditional theory. Most other quantitative theories of organization, as those using hierarchical forms , assume the structure of the organization to be given at the outset . A notable exception is the theory of teams by MARSCHAK and RADNER/1972/, where the structure is determined by the requirement for an expected minimal cost of communication, or, equivalently for an expected maximal payoff in terms of the teams' goals. Following some earlier work on "computable" organizations, GOTTINGER/1973/, we suggest that an organization reveals very much of the structure of a sequential machine, and for that purpose, in this approach, the structure of the organization is very much determined by the capabilities and limitations of the organization members which weakens the rigorous normative set-up of other organization theories.

2. ORGANIZATIONAL STRUCTURE AND PERFORMANCE

We deal here with the major building blocks of the kind of organization which has been described as "task-oriented". The purpose of such organizations appears to be best modeled as a decision-making task, so we start with an extremely simple decision-making or "computational" model.

We assume that

(i) the organization receives an input drawn from a certain set of inputs and is required to respond to it with an output,

(ii) it is rewarded for such a response with a certain payoff which in general depends on the response, as well as the input,

(iii) the goal of the organization is to maximize the expected or "average" payoff.

It will be useful to consider the inputs to come from "sources" and the outputs to be delivered to "destinations", both outside the organization. In information-theoretic terms we may speak of all inputs and outputs as being "signals", the same term may also be applied to transmissions among organization members.

A *signal* will be understood to be a sequence of symbols: a symbol may be a single letter, a memo, a certain number of resource costs, a job specification, a price quotation, etc. The collection of different symbols that are used in a signal are its "alphabet". All alphabets are assumed here to be finite but possibly quite large.

The members of the organization will be assumed to operate in the same way as the organization as a whole. They will acquire symbols as inputs, either from outside sources or from other members, and will dispatch them after suitable processing to certain destinations inside or outside of the organization. The members can be thought of as persons or machines. For simplicity, we designate them with impersonal letters such as alpha, beta, gamma, etc. If the organization is of a centralized nature we will have to add a centralized unit, cu.

Suppose now that x_j is an input symbol received by the organization at a certain point of time, either from one source or jointly from several. It is assumed that if x_j is dispatched the symbol will be received with a known "symbol probability" P_j and that one such symbol is acquired per unit time. The incidence of symbols in different time units, as they subsequently come in, are assumed to be *statistically independent.*

The organization is required to respond to the input symbol x_j with an output symbol y_k It need not do so within the same time interval as the acquisition of x_j. Thus delays are permitted, provided they do not pile up. That is, the average delay between the receipt of an input symbol and the delivery of an output symbol must not exceed one time unit.

Under these general conditions, if the organization respond to x_j with y_k it receives a reward R_{jk}. Of course, it would be desirable as a normative postulate, to respond in such a way that the reward is maximized.

However, under the assumptions to be made here concerning the individual organization members, in situations in which they have difficulties in coping with the schedule, the optimal response will rarely be realized.

Rather, when the input symbol x_j is received, every output symbol y_k is a possible response, and in fact will be the response with a certain conditional probability $P(y_k|x_j)$. The performance of the organization will be given by the average payoff, i.e. the quantity

$$(*) \quad ER = \sum_j P_j \sum_k R_{jk} P(y_k|x_j)$$

in which j and k range over the input and output alphabets, respectively.

Optimal performance is, of course, a special case of (*), i.e. it is the largest expected pay-off that the organization can obtain if all messages will be responded properly, when

$$k = k(j).$$

This appears to be the case, if

$$(**)\quad P(y_k|x_j) = 1, \text{ if } k = k(j)$$
$$= 0, \text{ otherwise.}$$

Completing the definition of the organizational goal we have a decision-making problem of a conventional type, with the restriction on the organization to stay on schedule. The theory of teams in particular uses a very similar set-up. The problem itself, as formulated, appears too simple, and if any claim to realism is to be made, some more restrictions have to be imposed.

(a) *Stochastic Dependence*
Successive input symbols should be allowed to be *dependent*, or, the input arrivals should be allowed to be a specific random process,

(b) *Uncertainty*
Complete knowledge of the input alphabets and probabilities, as well as the payoff matrix R_{jk} should be dropped.

(c) *Complexity*
Cognitive and computational limits in handling vast amounts of messages and in choosing among responses should be explicitly recognized.

(d) *Penalty*
The organization should be penalized for delays.

We will not treat these generalizations in a detailed fashion, but instead show what the conceptual building blocks are that provide for such generalizations.

3. A TWO - MACHINE MODEL

As mentioned in the previous section, the organization member "processes" x_j into y_k. This processing may be deterministic or stochastic. In the first instance, the organization member produces the same output symbol y_k whenever he receives x_j and transmits it to the same destination, in the second he does so only with a given probability.

We are emphasizing that the organization is 'task-oriented' and that every member is coping with the difficulty of properly processing input into output symbols, depending on the nature of the task as well as on his competence for its execution. As a proxy measure for the level of difficulty we could use the (average) processing time (or the number of computational steps in time) which he needs to perform a given task.

It will be convenient to visualize an organization member, a machine, as being composed of an input machine and an output machine - hooked in serial connection as illustrated in Fig. 1. The reason for this distinction is that an organization member may be called upon to do two quite different tasks, and may have difficulties of different kinds with performing them. One is the acquisition and sorting out of input symbols, the other the production and dispatching of output symbols.

The processes which are needed for one usually differ significantly from those for the other, and their processing times accordingly depend on quite different parameters.

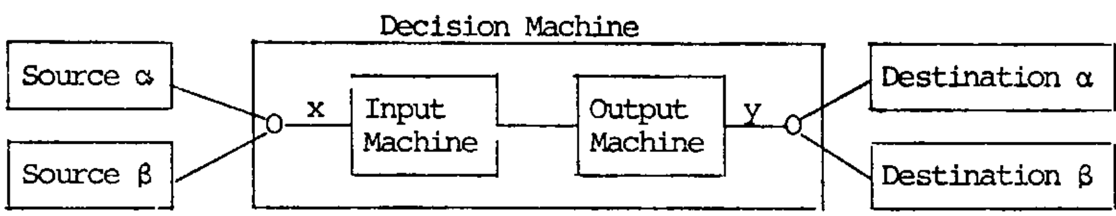


Fig. 1

The decision machine is then characterized by two sets of processing times. Thus, $t_j^{(i)}$, i for input, is the processing time of the input machine for symbol x_j, and $t_k^{(o)}$ is the corresponding time of the output machine for the symbol y_k.

The performance of the two machines can be measured by their average processing times

$\tau^{(i)}$ and $\tau^{(o)}$, given by

$$\tau^{(i)} = \sum_i P_j^{(i)} t_j^{(i)} \quad , \tau^{(o)} = \sum_k P_k^{(o)} t_k^{(o)} \tag{1}$$

where P_j and P_k are the probabilities with which the processing of x_j and y_k is called for.

The *processing load* of the organization member is measured by the sum of $\tau^{(i)}$ and $\tau^{(o)}$, i.e. by

$$\tau = \tau^{(i)} + \tau^{(o)} \tag{2}$$

The individual processing times $t_j^{(i)}$ and $t_k^{(o)}$ may depend only on the symbols x_j or y_k that are being processed, but they may depend also on other parameters, for instance, length or number of symbols, not explicitly considered here.

All these parameters may reflect the *complexity* or difficulty of a processing task for a particular organization member. This will be developed further in the next two sections. By the assumption that an organization is required to meet its schedule we see immediately that the average processing time τ of each member must not exceed the average time between the arrivals of input symbols, namely one time unit. Therefore,

$$\tau = \tau^{(i)} + \tau^{(o)} \leq 1 \tag{3}$$

must be assured for every member of the organization. If this bound is violated for some member, he will be considered "overloaded", in an intuitive and technical sense.

4. *THE OUTPUT MACHINE*

We set out to measure the relative difficulty of the task by the processing times t_k which the output machine needs to carry them out. In conventional information theory, as in SHANNON's model /1949/, the processing times are assumed constants, i.e., would depend only on the symbol y_k. It will become apparent below that the machine outputs of one organization member often need to be transmitted to several destinations according to certain more or less involved procedural rules. It is then too simple an assumption to suppose that the processing time t_k will be the same for all such rules.

Suppose, for instance, that y_k is a directive which is produced by the output machine of an organization member in response to some input x_j. The procedural rule may then specify that y_k be sent to the same organization member, α say, whenever it is called for but that other directives be addressed to other members. The rule may, on the other hand, be rotational. Thus, if the directives have r potentional destinations, α, $\beta, \ldots, \mu$ the rule may specify that α receives every r-th directive. One may also consider a stochastic rule which would prescribe the probabilities

$$P_{\alpha k}, P_{\beta k}, \ldots, P_{\mu k} \text{ with which}$$

y_k be transmitted to those m members, such that

(*) $P_{rk} \geq 0 \quad (r = \alpha, \beta, \ldots, \mu \; ; \; k = 1, 2, \ldots, n)$ and

(**) $\sum_r P_{rk} = 1.$

An output machine operating in this way will be said to process its symbols "alternatively" and the probabilities P_{rk} will be called the "assignment probabilities".

The procedural rule for an output machine can however be more involved than the ones that have just been described. The greater complexity occurs when y_k is actually an m-tuple of subsymbols $y_k^1, y_k^2, \ldots, y_k^m$, say, each of which can be transmitted to α or β or... μ. The rule may then specify first a certain permutation p_{uk} of the superscripts $\{1,2,\ldots m\}$
i.e.

$$\{\sigma_1, \sigma_2, \ldots, \sigma_m\} = p_{uk}\{1,2,\ldots,m\} \quad \text{such that}$$

$$y_k^{\sigma_1} \rightarrow \alpha \; , \quad y_k^{\sigma_2} \rightarrow \beta, \ldots$$

The rule would further specify how many permutations p_{uk} are allowed, and it would finally specify the probability P_{uk} with which each of those is to be used. The P_{uk} would then be the assignment probabilities under this rule and would satisfy

$$P_{uk} \geq 0, \quad \sum_u P_{uk} = 1 \quad (k = 1,2,\ldots,m)$$

with u ranging over all allowed permutations. An output machine operating under a procedural rule such as this will be said to be doing its processing "in parallel". One might expect parallel processing to represent an even more complex task than alternative processing, especially when the number u of allowed permutations is large, and hence to require even longer processing times t_k. The discussion of two possible processing procedures indicates that the output processing times t_k should be allowed to depend on certain parameters, for instance, the number m of distributions or the number u of allowed permutations, and perhaps, also on the assignment probabilities P_{rk} or P_{uk}. The dependence should in some way reflect the difficulty of the processing jobs.

Let us make these notions more precise:

DEFINITION 1. An *alternatively* processing output machine will be called "load-dependent" if it has the following properties:

(a) Its processing times t_k are functions of the number m and of the assignment probabilities

$$P_{rk}, \; r = \alpha, \beta, \ldots, \mu \quad \text{, i.e.}$$

$$t_k = t_k(m, P),$$

defined for all $m \geq 1$ and all vectors P with components $P_{\alpha k}$, $P_{\beta k}, \ldots, P_{\mu k}$, satisfying the constraints (*) and (**).

(b) Suppose m is fixed and one of the assignment probabilities, $P_{\mu k}$, say, is reduced to $P'_{\mu k} < P_{\mu k}$ but in such a way that

$P'_{rk} \geq P_{rk}$ for all $r \neq \mu$. The processing time t_k of y_k is then not decreased, and we have

$$t_k(m;P') \geq t_k\ (m;P)$$

(c) Suppose one of the assignment probabilities, $P_{\mu k}$ say, is zero. In that case

$$t_k\big((m-1);\ P'\big) \leq t_k(m,P)$$

where P' is an (m-1) vector whose components $P_{\alpha k}, P_{\beta k}, \ldots, P_{1k}$ are the same as the first (m-1) components of P.

DEFINITION 2. An output machine processing in parallel will be called "load dependent" if its processing times t_k are functions of the number m of the destinations, of the number u of allowed permutations p_{uk}, i.e.

$$t_k = t_k(m;u;P)$$

provided that the t_k have the properties (b) and (c) in u and P when m is fixed, and

$$t_k\ (m-1;u;P) \leq t_k(m;u;P)$$

when u and P are fixed, provided $u \leq (m-1)$

THEOREM 1: Let an alternatively processing output machine be load dependent in the sense of *DEF.1*. For fixed m, its processing times t_k are continuous concave functions of the assignment probabilities P_{rk}, and each assumes its minimum when all but one of the P_{rk} vanish. This minimum is a monotone non-decreasing function of m.

PROOF. Consider first alternative processing with m destinations, and some fixed y_k. Suppose that the assignment probabilities are changed to

$$P'_{rk},\ r = \alpha,\beta,\ldots,\mu, \text{ in accordance with (b) of } \mathit{DEF.1}:$$

$$P'_{\alpha k} \geq P_{\alpha k}, \; P'_{\beta k} \geq P_{\beta k}, \ldots, \; P'_{1k} \geq P_{1k},$$

$$P'_{\mu k} < P_{\mu k}.$$

These inequalities imply

$$P_{\alpha k} + P_{\beta k} + \ldots + P_{1k} \leq P'_{\alpha k} + P'_{\beta k} + \ldots + P'_{1k},$$

$$P_{\alpha k} + P_{\beta k} + \ldots + P_{1k} + P_{\mu k} = P'_{\alpha k} + P'_{\beta k} + \ldots + P'_{1k} + P'_{\mu k} = 1.$$

It is known that two vectors P and P' whose components P_{rk} and P'_{rk} satisfy inequalities such as these, are related by an equation of the form

$$P = QP'$$

in which Q is a doubly stochastic matrix, i.e., one whose elements Q_{rs} satisfy

$$Q_{rs} \geq 0, \; \sum_r Q_{rs} = \sum_s Q_{rs} = 1$$

$$(r,s = \alpha, \beta, \ldots, \mu).$$

Suppose that P', in particular, is specified as the vector with the component

$$P'_{\alpha k} = 1$$

and all others zero. Then P will have the components $Q_{r\alpha}, Q_{r\beta}, \ldots Q_{r\mu}$, and since Q can be any stochastic matrix, all vectors P of assignment probabilities can be generated from this P' by suitable choices of Q. These vectors can be, however, also be generated by

$$P = \sum_j \lambda_j P^{(j)} \; (j = 1,2,\ldots,m; \; \sum_j \lambda_j = 1)$$

where P' is the vector that has just been defined, P'' similarily has the component

$$P''_{\beta k} = 1$$

and allothers zero, and $P^{(m)}$ has

$$P^{(m)}_{\mu k} = 1$$

and all others zero. Therefore, by the inequality in property (b) of *DEF.1* it follows that

$$t_k(m;P) \geq \max_j \; t_k(m;P^{(j)}) \geq \sum_j \lambda_j t_k(m,P^{(j)}),$$

which proves the concavity of $t_k(m,P)$ on the polyhedron defined by the assignment probabilities, i.e

$$P_{rk} \geq 0, \; \sum_r P_{rk} = 1.$$

Its continuity on the interior of this polyhedron follows from a wellkown fact regarding concave functions, see e.g. KARLIN/

In the present case, however, it is permissible to include also the boundary of the polyhedron on which one or more of the P_{rk} vanish because any discontinuity in t_k there can be absorbed into its behavior as a function of m. Thus, t_k is concave on a closed bounded convex set, namely the polyhedron. According to another wellknownfact concerning concave functions, t_k assumes its minimum at extremepoints of such a set, in the present case at one of the vertices $P^{(j)}$.

Thus, the individual processing times t_k of a load dependent output machine have the properties as stated in the theorem.

COROLLARY 1. Consider an output machine, processing alternatively to m destinations. Suppose that its processing times t_k are concave functions of the assignment probabilities. Then

$$t_k(m;P) \geq t_k(m;P')$$

for any two vectors P and P' related to each other as in the proof of *THEOREM* 1.

5. *THE INPUT MACHINE*

The input machine collects certain inputs x_j from sources outside or inside the organization, and converts them in a one-to-one fashion into a form that the output machine can process them. The collection and conversion of x_j will in general require a certain processing time also, for which the notation $t_j^{(i)}$ was introduced. (The superscript (i) will henceforth be omitted, for notational simplicity.)

The allowances for the complexity of an input processing task are different from those of an output machine, in fact, they appear to have no couterpart on the output machine.

According to a large volume of psychometric data, the processing time ('the reaction time') for an input symbol x_j varies with the probability with which the symbol arrives. The input machine in other words, somehow quickly accumulates statistical evidence concerning the relative frequency with which the various x_j are received and then adapts its processing times accordingly. Symbols that occur rarely are processed more slowly and those that come up frequently are disposed of quickly. There are, in fact, indications that the variation of t_j with the probability p_j is roughly logarithmic, i.e.

$$(*) t_j = t_{oj} - c_j \log P_j$$

but this observation does not seem to be uniformly accepted by experimental psychologists.

Under these circumstances it may be appropriate to define load dependence for input machines in a way that is roughly analogous to *DEF.1* for output machines, but includes (*) as a special possibility. In such a case the analogy should further make plausible allowance for the complexity of alternative and parallel processing tasks. The size of an input alphabet can be quite large. It may be appropriate to associate the notion of the complexity of the task of input processing with the numbers of input symbols, and the probability of their occurrence in roughly the same way in which this notion was associated with the number of destinations (or of permutations) for the output machine. The qualitative analogy that suggests itself here would then be this: an input processing task would be the easier, the smaller the number n of symbols in the output alphabet, and if n remains the same the task should become easier if the frequency of the processing is increased.

DEFINITION 1. An input machine will be called "load dependent" if its mean processing time τ has the following properties:

(a) It is a function of the number n of symbols in the input alphabet and of symbol probabilities

$$P_j, j = 1,2,\ldots,n :$$

$$\tau = \tau(n;P) = \sum_j P_j t_j(n;P)$$

defined for all $n \geq 1$ and all vectors P with components $P_1, P_2, \ldots, P_n$.

(b) Suppose n is fixed and one of the symbol probabilities P_n say, is reduced to $P_n' < P_n$ but in such a way that $P_j' \geq P_j$ for all $j \neq n$. The mean processing time is then not decreased, i.e.

$$\tau(n;P') \geq \tau(n;P)$$

(c) Suppose that one symbolprobability, P_n for instance, is zero. In that case,

$$\tau(n-1;P') \leq \tau(n;P)$$

where P' is an (n-1)-vector whose components $P_1, P_2, \ldots, P_{n-1}$ are the same as the first (n-1) components of P.

Notice that this definition bases load dependence of an input machine on the mean processing time τ. This is in contrast to its counterpart in Sect. 4 which is based on the individual symbol processing time $t_k^{(o)}$.

The reason for this lies in *THEOREM 1*, Sect. 4, which shows that the $t_k^{(o)}$ are concave functions of the assignment probabilities P_{rk} and P_{nk}. If the definitions here were based on the individual symbol processing times, t_j, also, these would be found concave in the symbol probabilities P_j, and that would be incompatible with (*) : t_j in that equation is convex in P_j. *DEF.1*, however, implies properties for the mean processing time which are the same as those of its counterpart of the output machine.

THEOREM 1: Let an input machine be load dependent in the sense of *DEF.1*. For fixed n, its mean processing time is a continuous concave function of the symbol probabilities P_j which assumes its minimum when all but one of the P_j vanish. This minimum is a monotone non-decreasing function of n.

COROLLARY 1. Suppose that the mean processing time τ of an input machine is a concave function of the symbol probabilities. Then

$$\tau(n;P) \geq \tau\ (n;P')$$

where P and P' are two vectors of symbol probabilities related by P = QP' with Q being a doubly stochastic matrix.

THEOREM 1 shows that the mean processing time τ of an input machine can be concave in the symbol probabilities even in cases in which the individual processing times t_j are convex. In particular, if the P_j are given in logarithmic form, then

$$\tau = \sum_j P_j t_{oj} - \sum_j c_j P_j \log P_j$$

is concave, and the same is true when

$$t_j = c(1 - P_j^{s-1}) \ , \ (s > 1)$$

which is convex for $s < 2$, but

$$\tau = c\left(1 - \sum_j p_j^s\right) \qquad (s > 1)$$

is again concave.

Information theory supplies a remarkable characterization of the mean processing times for signals that can be processed by certain channels.

These channels are of the same kind as the "input machines" introduced here, but they must be load independent. The signals are characterized by their "entropies"

$$H = -\sum_j p_j \log p_j .$$

The corresponding quantity for the channel is its "channel capacity" C which can be defined in two ways, i.e., either by

$$C = \log \alpha_o \tag{a}$$

where α_o is the largest root of the equation

$$\sum_j \alpha^{-t_j} = 1, \tag{b}$$

or as the limit

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T} \tag{c}$$

where N is the number of sequences of input symbols that can be processed by the channel in period T. It can be shown (see e.g. SHANNON and WEAVER /1949/, p. 37) that (b) always has a positive root and further that the limit in (c) coincides with (b) and (c).

The above mentioned characterization of signals can then be stated as follows, for the present purposes.

THEOREM 2: Let a load independent machine be given which has the processing times t_j for the symbols x_j, $j = 1,2,...,n$. The input signal x^* with these symbols and with the entropy H which has the smallest mean processing time τ^* is then one with the symbol probabilities P_j^* defined by

$$- \log P_j^* = Ct_j \quad (j = 1,2, ...,n)$$

C being the capacity of the machine. The mean processing time for x^* is

$$\tau^* = H|C.$$

PROOF. This proposition is a rephrasing and specialization of some of SHANNON's results, namely of his theorems 1,8 and 9 combined, and the proof is therefore omitted. It can be shown that results analogous to *THEOREM* 2 hold also for load dependent input machines.

6. *ORGANIZATIONAL MALFUNCTIONING*

An organization member is considered *malfunctioning* if he (she) cannot keep up with the schedule on which the organization works but falls progressively further behind. This will be the case when

$$(*) \quad \tau = \tau^{(i)} + \tau^{(o)} > 1$$

where $\tau^{(i)}$, $\tau^{(o)}$ are the mean processing times of his input and output machines. The only way in which an overloaded member can react to this predicament is to make mistakes.

It does not matter for the moment how an organization member makes his errors, i.e., whether he does so deterministically, stochastically or in some other , less well defined manner. All that matters here is that on occasion an overloaded organization member will generate an output other than the one that is called for, and as a result, will jeopardize the *expected payoff* to the organization. What are the conditions under which such an over-

overload occurs?

The issue that will matter most is the intuitive notion that overload should develop in most cases of interest when the complexity of processing tasks for either or both of the machines of an organization member increases beyond some critical point.

DEFINITIONS 1 (Secs. 4,5) in fact characterize the effect of measuring complexity but neither makes overload inevitable even if the increase is unbounded. To see this, consider first an input processing task. While there are several ways of increasing its complexity, according to *DEF*.1 (Sec.5), the only way to do so without bound is to let the size n of the input alphabet grow to infinity. For output machines, the same can be achieved only by letting the number m of destinations go to infinity, according to *DEF*. 1 (Sec.4), since μ cannot increase indefinitely without m doing the same. However, both definitions merely imply monotonicity of the mean processing times in n or m, which need not lead to (*). Malfunctioning or overload will, however, be inevitable if, for instance

$$(**)\quad \lim_{n \to \infty} \tau^{(i)}(n) = \infty \text{ or } \lim_{m \to \infty} \tau^{(o)}(m) = \infty,$$

disregarding here the possible dependence of $\tau^{(o)}$ on μ. In cases in which (**) does not apply, overload will be inevitable at least for certain procedural rules, if

$$(***)\quad \lim_{n \to \infty} \max_P \tau_P^{(i)}(n,P) = \infty \text{ or}$$

$$\lim_{m \to \infty} \max_P \tau_P^{(o)}(m,P) = \infty .$$

In practice, unboundedness of the mean processing times, in the sense of (**) or (***) or both is likely to be encountered more often than not. It will therefore be useful to establish two results here which are pertinent in such cases.

THEOREM 1. Suppose that a load dependent output machine has the second property of (**). For any given input load, $\tau^{(i)}$, there is a maximum number $\bar{m}$ of destinations which the organization member can serve without being overloaded, and which is achieved by a deterministic assignment rule. An analogous statement holds for the input machine of an organization member with fixed output load $\tau^{(o)}$.

PROOF: Consider an alternatively processing output machine first. Since mean processing time $\tau^{(i)}$ of the input machine is assumed fixed, then overload will be avoided by the member if,

$$\tau^{(o)}(m,P) \leq 1 - \tau^{(i)}.$$

Now, for every m, $\tau^{(o)}(m)$ is the minimum of the l.h.s. above. Therefore, the largest value of m for which this inequality holds is

$$\tau^{(o)}(m) = \min_P \tau^{(o)}(m,P) \leq 1 - \tau^{(i)}.$$

The minimum, however, occurs for a vector P with all but one of its components zero, according to *THEOREM* 1 (Sec. 4) Thus, for a fixed $\tau^{(i)}$, the maximum m is achieved by a nonstochastic assignment rule, as asserted in this theorem. The argument clearly applies also to output machines doing their processing in parallel.

It is conceivable that situations exist in practice in which (**) fails. It is more specifically possible that

$$\lim_{n\to\infty} \tau^{(o)}(m,P) = T < 1$$

for an output machine and that no overload develops if an assignment rule is used that minimizes the mean output processing time. In such cases, (***) may hold and overload will develop for at least non-minimizing assignment rules. One then has a result that is a counter-part to the preceding, namely this.

THEOREM 2. Suppose that a load dependent output machine has the property (***). For any given input load $\tau^{(i)}$, there is a maximum number $\bar{m}$ of destinations which the organization member can serve, using a stochastic assignment rule and without being overloaded. An analogous statment holds for the input channel of an organization member with fixed output load $\tau^{(o)}$.

PROOF. Consider again an alternatively processing machine. As m increases, the vectors of assignment probabilities make up a sequence of convex polyhedra $\overline{CP}^{(m)}$ of increasing dimensions. Suppose that the proposition is false and there exists a vector $\bar{p}$ which lies in all of these polyhedra and for which

$$\tau^{(o)}(m,\bar{P}) \leq 1-\tau^{(i)}(n)$$

Let $P^{(m*)}$ be the maximum of $\tau^{(o)}(m,P)$ on $CP^{(m)}$. One can then construct a convex sub-polyhedron of $CP^{(m)}$, $CP^{(m*)}$ say, which has as its vertices some of those of $CP^{(m)}$ and $P^{(m*)}$, which contains $\bar{P}$. That is

$$\bar{P} = \sum_j \lambda_j^{(m)} P^{(j)} + \lambda^{(m*)} P^{(m*)}, \text{ where}$$

the sum ranges over the vertices of $CP^{(m*)}$, other than $P^{(m*)}$, and where

$$\sum_j \lambda_j^{(m)} + \lambda^{(m*)} = 1.$$

Because of the concavity of the mean processing time

$$\tau^{(o)}(m,\bar{P}) \geq \sum_j \lambda_j \tau^{(o)}\left(m, P^{(j)}\right) + \lambda^{(m*)} \tau^{(o)}\left(m, P^{(m*)}\right).$$

By assumption, $\tau^{(o)}\left(m, P^{(m*)}\right)$ approaches infinity as m does. Therefore, so does $\tau^{(o)}(m,\bar{P})$ unless $\lambda^{(m*)} = 0$ from some $m \geq \bar{m}$. But this means that the components of $\bar{P}$ beyond the $\bar{m}$-th vanish, as asserted. The same argument again applies to parallel processing output machines, as well as to input machines.

6. *CONCLUSIONS*

A model of an individual organization member should first of all, be consistent with experimental evidence. And experimental evi-

dence suggests that principles of bounded rationality and complexity considerations impose restrictions on behavior rules, even more so in an organizational context (see e.g. GOTTINGER/1979/, and SIMON/ 1978/).
Beyond that, however, the model should be as general as possible. We attempt to achieve this by making the assumptions as intuitively plausible as possible. The machine model of an organization and its members was chosen here mainly on grounds of intuitive plausibility. The task of acquiring and interpreting inputs (the input machine) is often so different from that of producing and delivering outputs (the output machine) that it appears to be justified to represent them by different mathematical models. The notion of a load dependent machine is also based on intuitive grounds. It is supported by observing executions in "large" organizations, in particular the existence of interdependence in such organizations, facilitating bottleneck multipliers, adds support to "load dependence".

On the other hand, the formalization of load dependence leans on SHANNON's ideas, and on the observation that the concepts of task complexity in this study and of signal uncertainty in information theory have quite similar properties.

The definition of load dependence for input machines, by contrast to that of the output machine, is based on the evidence of experimental psychology (H.W.GARNER /1962/). The resulting machine model seems to be one of the most general that are compatible with this evidence, and with the idea of the complexity of a processing task. The fact that human input machines often react to overload by making errors is also well established experimentally. Some improvements of this model appear to be desirable and can be pursued in the following direction.

(i) an input machine, as defined here, is an especially simple kind of stochastic service (i.e. waiting line) system, e.g. in which the service times are known constants rather than random variables and in which problems, such as the distributions of waiting times and the numbers of customers waiting in line are ignored. Similar problems should presumably be considered in connection with the machine models developed here.

(ii) one might also wish to make "load" dependent on effects other than the parameters that are now included. For input machines the number of sources seems a particularly useful additional parameter.

The disadvantage of keeping the basic concepts as general as possible are that the statements that can be made concerning them are inevitably rather diffuse and weak. A case in point is the fact that the definitions of load dependence in Secs. 4 and 3 do not necessarily lead to overload, as task complexity increases. This is certainly contrary to intuition and experience. The additional assumptions made in Sec.5 try to specify the results and put them in a proper perspective.

REFERENCES

GARNER, H.W., Uncertainty and Structure as Psychological Concepts, New York: Wiley 1962.

GOTTINGER, H.W., 'Computable Organizations - Representation by Sequential Machine Theory: Annals of Systems Research 3, 1973, 81 - 108. 'Complexity, Bounded Rationality and Problem Solving', Proc.- Third Wittgenstein Symp., Kirchberg/Austria, Aug. 1978, Vienna: Hölder 1979.

KARLIN, S., Mathematical Methods and Theory of Games, Programming and Economics. Reading (Mass.): Addison-Wesley 1959.

MARCH, J.G. and H. SIMON, Organizations, New York: Wiley 1958.

MARSCHAK, J. and R. RADNER, The Economic Theory of Teams, New Haven: Yale Univ. Press 1972.

SHANNON, C.E. and W. WEAVER, The Mathematical Theory of Communication, Urbane: Univ. of Illinois Press 1949.

SIMON, H., 'How to Decide to Do What', Bell Jour. of Economics and Management Science 1978.

WEBER, M., Wirtschaft und Gesellschaft, Bd. II, Tübingen: Mohr 1922.