

Note: Cowles Foundation Discussion Papers are preliminary materials circulated privately to stimulate private discussion and critical comment. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

A Flow Model of Communication - -
Towards an Economic Theory of Information.*

Martin J. Beckmann**

November 30, 1956

Was man nicht weiss, das eben
brauchte man und was man weiss,
kann man nicht brauchen! -- Faust

* Research undertaken by the Cowles Commission for Research in Economics under Contract Nonr-358(01), NR 047-006 with the Office of Naval Research.

** Many thanks are due to J. Marschak, who has discussed the first section with me repeatedly and has made numerous critical and terminological suggestions which resulted in important changes in the paper, and to R. Radner who gave further advice.

Information has some interesting properties whose implications for economic theory are worth exploring. Like other "commodities" it can be stored, can be made available in other locations, and can become an input into production -- the production of decisions. Its use is limited by its availability. It can have a marginal*

* In activity analysis the right and left hand derivatives may be different and "marginal productivity" must be understood in this broader sense.

productivity and hence a value.-- Unlike other commodities, it is not exhausted by use and not removed by transportation. In particular the cost of its dissemination may be independent of the number of destinations it is sent to. Since most problems of interest center around its transportation i.e., the processes of communication, it is natural to turn to the facts of commodity transportation for guidance by analogy. The question arises, to what extent the theory of transportation in networks can serve to illuminate problems of communications in organizations, that is, of the efficient flow of information between many users.

One notices at once that the fundamental equation of continuity for commodity flows

$$\text{outflow} = \text{inflow} - \text{net absorption}$$

is inapplicable. What can take its place? Is it possible to formulate communications problems in terms of linear models? Can the theorem on efficiency prices be applied and give insight into the structure of the solution?

1. Information: Concept, Cost, Utility

In these preliminary considerations, the emphasis is on the relation of cost to amount of information. However, by symmetry they apply equally well to the utility of information.

1.1 Information will be defined in terms of subsets $\omega \in \Omega$ where Ω is the universal set.

A partition ξ of Ω , written $\{\omega\} = \Omega_\xi$

will define (subject to later refinement) the kind of information one is interested in.

In a partition ξ let information ω be realized with frequency $\pi(\omega)$. Let $\varphi_\xi(\omega)$ be the cost of information ω under the partition ξ . Then the expected cost of information for a partition ξ equals

$$\sum_{\omega \in \Omega_\xi} \pi(\omega) \varphi_\xi(\omega)$$

Assumption 1.1 If for every element ω of a partition ξ there exists an element σ_ω of a partition η such that $\omega \subseteq \sigma_\omega$, then, for every ω

$$\varphi_\xi(\omega) \geq \varphi_\eta(\sigma_\omega)$$

In other words cost (and utility) of information increases with the fineness of the partition.

1.2 For greater concreteness assume now that there is a well defined concept of "size" of the information sets valid for all partitions of a certain class E , such that both cost and utility of information depend only on the size of information sets so defined. Mathematically this means

that we have postulated the existence of a measure $\alpha(\omega)$

$$\text{such that } \varphi_{\xi}(\omega) = \hat{\phi}_E[\alpha(\omega)] \quad (\text{cost})$$

$$\psi_{\xi}(\omega) = \hat{\psi}_E[\alpha(\omega)] \quad (\text{utility})$$

for all $\xi \in E$.

It turns out to be more convenient to use instead the negative logarithm of α as the argument.

$$\varphi_{\xi}(\omega) = \phi_E[-\log \alpha(\omega)] \quad \psi_{\xi} = \psi_E[-\log \alpha(\omega)].$$

This assumption implies 1.1 but is not implied by it. Without great loss of generality we shall assume $\alpha(\Omega) = 1$.

Example 1

Suppose that the cost of information is the same for all sets ω of a given partition ξ . Then if n_{ξ} is the number of sets ω in ξ

$$\alpha(\omega) = \frac{1}{n_{\xi}} \sum_{\omega} \alpha(\omega) = \frac{1}{n_{\xi}} \alpha(\Omega) = \frac{1}{n_{\xi}}$$

$$\varphi_{\xi}(\omega) = \phi_E\left(-\log \frac{1}{n_{\xi}}\right) = \phi_E(\log n_{\xi})$$

and
$$\sum_{\omega \in \Omega \xi} \pi(\omega) \varphi_{\xi}(\omega) = \varphi_{\xi}(\omega) = \phi_E(\log n_{\xi})$$

Now $\log n$ is proportional to the number of steps required to identify one out of n possibilities if at each step k alternatives can be distinguished, the number of steps required being

$$k \log n = \frac{\log n}{\log k}$$

If therefore cost of information is a linear function of the number of steps, the expected value of cost equal

$$a \log n_{\xi} + b$$

where a and b are constants.

Example 2

Suppose that the cost of information is a linear function of the number of steps required for its identification, that k alternatives can be distinguished at each step, and that the information sets ω and their probabilities $\pi(\omega)$ are given. What is the partition which minimizes the expected cost?

It is intuitive that at each step the alternatives distinguished should be as nearly as possible of equal probability. Then an information set of probability $(1/k)^n$ requires n steps. An information set of probability π requires therefore approximately $\log_k \pi$ steps. This result is in accordance with optimal coding as considered in information theory.* Now the cost of information was assumed to be a linear function

* Shannon, C.E. and W. Weaver: The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1949, pp. 18-32.

of $-\log \alpha$. We see that in terms of our model an optimal partition is characterized by the fact that $\alpha(\omega)$ equals* (approximately) $\pi(\omega)$, the

* Equality, and not just proportionality, results from the normalization $\alpha(\Omega) = 1$.

probability of the occurrence of ω .

1.3 Returning to the case of a general information measure $\alpha(\omega)$ let us consider the implications of disaggregating information into several independent and independently observed variables. In terms of the concepts introduced, this means that

Assumption 1.3a Ω is a (finite dimensional) vector space and the sets ω are product sets

$$\omega = \omega_1 \times \omega_2 \times \dots \times \omega_m \dots \times \omega_M$$

Assumption 1.3b The costs of information are additive for different components.

$$\phi_{E,m}(\omega) = \sum_{m=1}^M \phi_{E,m}(\omega_m)$$

Combining these assumption with 1.2 we have that

$$\phi_E[-\log \alpha(\omega)] = \sum_m \phi_{E,m}[-\log \alpha_m(\omega_m)]$$

where the $\phi_{E,m}$ are monotonically non-decreasing functions.

From now on we shall consider only partitions for which the set of variables (the co-ordinate system) E is fixed. We shall say that we regard the kind of information considered as fixed. The terms

$$-\log \alpha_m(\omega_m)$$

for which we use the shorter notation x_m will be said to express the precision of the partition within the framework E of the kind of information. We shall suppress the index E in many expressions that follow.

Both cost and utility of information are now functions of the vector $x = x_m$ of precision for the various variables. We shall write

$$= \sum_{\omega_m \in \Omega_m} -\log \alpha_m(\omega_m) \pi(\omega_m)$$

$$f(x) = \sum_{\omega, m} \phi_{E, m} [x_m(\omega)] \pi(\omega) = \sum_m f_m(x_m) \quad (\text{cost})$$

$$g(x) = \sum_{\omega, m} \psi_{E, m} [x_m(\omega)] \pi(\omega) \quad (\text{utility})$$

Example 3: Assume that the cost of information on the different variables is linear and independent of m

$$\begin{aligned} \phi_{E, m}(x_m) &= a x_m + b \\ &= - a \log \alpha_m(\omega_m) + b \end{aligned}$$

then

$$\begin{aligned} \phi_E(-\log \alpha) &= - a \sum_m \log \alpha_m + b \\ &= - a \log \prod_m \alpha_m + b \\ &= \phi_E(-\log \prod_m \alpha_m) \end{aligned}$$

from which we conclude that

$$(1.3.1) \quad \alpha(\omega) = \prod_m \alpha_m(\omega_m) \quad \text{and}$$

$$(1.3.2) \quad \phi_E(-\log \alpha) = - a \log \alpha + b$$

In words, the total cost of information is also linear, and the measure α is the product measure $\alpha(\omega) = \prod_m \alpha_m(\omega_m)$.

1.4 In terms of the model 1.3 the optimal use of information by a single decision-maker may be considered after the fashion of determining optimal factor inputs into production. For, in the context of organization theory, information is viewed instrumentally as the raw material from which decisions are produced and not as an end in itself. With the usual assumption that the decision-maker seeks to maximize the expected payoff

subject to the limitations of capacity, we have a well defined (decision) production function, expressing the maximal expected payoff as a function of the precision with which the different variables are known.

$$\text{Max}_x [g(x) - \sum_m f_m(x_m)]$$

If $g(x)$ and the $f_m(x_m)$ are differentiable we obtain the marginal conditions

$$(1.4.1) \quad \frac{\partial g}{\partial x_m} \begin{cases} = \\ < \\ \leq \end{cases} \frac{df_m}{dx_m} \quad \text{according as} \quad x_m \begin{cases} > \\ = \end{cases} 0,$$

stating that the marginal cost of precision for each variable must equal its marginal productivity and that a variable must remain unobserved when its marginal cost of precision always exceeds its marginal productivity.

Not much is changed in this simple model when we admit, that the decision makers' time is limited and the decision production function (the marker's payoff) depends on both the precision with which the variables are known and on the time w available to the decision maker, $g(x;w)$. The latter excludes of course the time $s(x)$ consumed in information handling. If c is his total working capacity, then

$$w = c - s(x)$$

and the decision maker seeks to maximize

$$g[x; c - s(x)] - \sum_m f_m(x_m)$$

The Equations (1) are then replaced by the conditions

$$(1.4.2) \quad \frac{\partial g}{\partial x_m} \begin{cases} = \\ < \\ \leq \end{cases} \frac{\partial g}{\partial w} \frac{ds}{dx_m} + \frac{\partial f_m}{\partial x_m} \quad \text{according as} \quad x_m \begin{cases} > \\ = \end{cases} 0,$$

which take into consideration the value of time.

2. An Almost Linear Model of Communication

The more interesting aspects of information are brought out when we consider the generation and transmission of information in an organization consisting of several decision-makers or offices, as we shall say.

From now on we identify "cost" of information with "time consumed" and disregard money cost, which do not add anything important.

Basic Notation:

m variables constituting information

t age of information

i offices

ij channels

Our basic variables are two flows: observation x_{im}^t and transmission x_{ijm}^t of information; and one stock: availability of information u_{im}^t . Their levels are measured in terms of precision as elaborated in the preceding section. With the help of these variables we can admit the following activities:

<u>activity</u>	<u>level</u>
observation	x_{im}^0
amplification	x_{im}^t
transmission	x_{ijm}^t
reception	x_{jim}^t
omission	$u_{im}^t - x_{ijm}^t$

We consider a stationary state in which information of varying age is current.

In a stationary state all activities of one period are repeated in the next. Storage of information* is therefore not called for,

* other than the rules of the organization which are considered ageless.

since wherever information is available now, more recent information will be known one period hence.

Up to date information is available only where it is observed

$$(2.1a) \quad u_{im}^0 = x_{im}^0 .$$

We assume that the transmission of information requires one unit of time.

Information aged one period or more may therefore be received in messages from other offices. By nature of precision, and since we measure information x_m in terms of precision, the information contained in all messages on variable m received at an office equals the maximum information in any such message. Observation (research) may add further information of that type and age. The total information available at an office thus equals

$$(2.1b) \quad u_{im}^t = x_{im}^t + \text{Max}_j x_{jim}^t \quad t > 0.$$

The availability of information limits the amount that can be transmitted

$$(2.2) \quad x_{ijm}^t \leq u_{im}^{t-1}$$

Observation may be restricted by the availability of sources

$$(2.3) \quad x_{im}^t \leq e_{im}^t .$$

A final constraint refers to the amount of work that can be performed on information in each office. Following information theory we assume that time consumed in handling information (observation, reception, dissemination) is proportional to quantity of information (precision) x_m as defined in section 1.

The time available in an office after all handling of information will be denoted by w_i , the total working time available by c_i . [Thus if all personnel in an office is of equal skill, c_i is the algebraic product of working time (per person) and the number of personnel]. With the appropriate proportionality factors we have then a condition

$$(2.4) \quad \sum_{mt} a_{im} x_{im}^t + \sum_{jmt} b_{jm} x_{jim}^t + \sum_{jmt} d_{ijm}^t x_{ijm}^t + w_i \leq c_i$$

The restriction expressed by (2.4) hinges on the fact that all variables must be non-negative. -- If all offices are equally efficient and accessible the coefficients reduce to

$$a_{im} = 1 \quad b_{im} = b_m \quad d_{ijm}^t = a_m^t$$

The decision production function will be assumed in the form

$$g(\dots u_{im}^t \dots, w_i \dots)$$

non-decreasing with respect to all its variables. For simplicity we assume g_i to be also differentiable and denote

$$\frac{\partial g_i}{\partial u_{im}^t} = g_{im}^t \quad \frac{\partial g_i}{\partial w_i} = g_{iw}$$

We shall assume furthermore that the law of diminishing returns holds, i.e. that the functions g_i are concave.

The object is to maximize the sum of payoffs over all offices within the capacity limitations of the organizations as expressed by the constraints (1) ... (4).

3. Enter Lagrange Multipliers

This is a straightforward programming problem, linear in the constraints with one exception: the maximum operation in equation (2.1) is not linear or convex. In writing out the efficiency conditions we must assume it to be known by which of the variables the maximum is realized. Consequently the efficiency conditions are merely necessary, not sufficient; in particular they leave undecided which of the j 's are the best maximizers in (2.1).

Let the Lagrange Multipliers associated with the equations be denoted, respectively,

$$(1) \lambda_{im}^t$$

$$(2) \beta_{ijm}^t$$

$$(3) \gamma_{im}^t$$

$$(4) \mu_i$$

It is understood that the β , γ and μ are non-negative, and equal to zero when in the corresponding constraints the $<$ sign applies.*

* Actually the λ and μ are also non-negative, since the solution of the problem is not changed by relaxing the equations (1) and (4) to inequalities of the \leq type.

We also use the notation

$$\delta_{jim} = \begin{cases} 1 \\ 0 \end{cases} \text{ if } x_{jim}^t \begin{cases} = \\ < \end{cases} \text{Max}_k x_{kim}^t$$

The efficiency conditions [H.W. Kuhn and A.W. Tucker, Nonlinear Programming, Second Berkeley Symposium 1951, pp. 481-492, in particular theorem 3 p. 486.]

are now as follows:

$$\begin{aligned}
 (3.1) \quad u_{im}^t & \left\{ \begin{array}{l} \geq \\ = \\ \leq \end{array} \right\} 0 \text{ according as } \varepsilon_{im}^t + \sum_j \beta_{ijm}^{t+1} & \left\{ \begin{array}{l} = \\ < \end{array} \right\} \lambda_{im}^t \\
 (3.2) \quad x_{im}^t & \left\{ \begin{array}{l} \geq \\ = \\ \leq \end{array} \right\} 0 \text{ according as } \lambda_{im}^t - a_{im} \mu_i & \left\{ \begin{array}{l} = \\ < \end{array} \right\} \gamma_{im}^t \\
 (3.3) \quad x_{ijm}^t & \left\{ \begin{array}{l} \geq \\ = \\ \leq \end{array} \right\} 0 \text{ according as } \delta_{ijm}^t \lambda_{jm}^t - b_{ijm} \mu_j & \left\{ \begin{array}{l} = \\ < \end{array} \right\} \beta_{ijm}^t \\
 (3.4) \quad w_i & \left\{ \begin{array}{l} \geq \\ = \\ \leq \end{array} \right\} 0 \text{ according as } \varepsilon_{iw} & \left\{ \begin{array}{l} = \\ < \end{array} \right\} \mu_i
 \end{aligned}$$

4. Discussion in Economic Terms

These conditions become intelligible with the following economic interpretations of the Lagrange multipliers.

λ_{im}^t is the value of information of type m and age t at office i

β_{ijm}^t is the value of a message sent from i to j

γ_{im}^t is the value of a source of information at office i

μ_i is the value of time in office i.

(3.1) states: information m should be available at office i to the point where its value equals its marginal utility in the decisions made at that office plus the value of messages m sent to other offices.

No information m should be available at office i when its marginal productivity in decisions and messages does not attain the level of its value. Value, it will be noticed, is here constructed from the cost side rather than derived from the utility side. But whenever information is actually present, the two amount to the same thing. This is another way of stating the equation part of (3.1).

(3.2) says: information m should be generated at i with the degree of detail for which the value of the information becomes equal to the time cost of producing it plus the cost imputed to using the source.

(3.3) expresses that a message m aged t should be sent from i to j if, and in such detail that, the value of the message equals its value at its destination minus the cost of transmission. In particular no message will be sent unless $\delta_{ijm}^t = 1$, i.e., unless this is the message received at j with the greatest detail for m and t . Hence at most one message about m and t will be received at every office j .

(3.4) demands that the office staff should have as much time for decisions apart from time spent on information so as to make the value of time equal to its marginal utility in decision making.

If we disregard the unequal availability of sources as one cause, the main reason for communication would seem to rest in the economy of using in several other places information generated in one office. For the cost of communication is typically much less than that of observation. Of course the cost of transmission must not exceed the usefulness of the information at its destination. Condition (3.3) implies that

$$\delta_{ijm}^t \lambda_{jm}^t \geq a_{ijm} \mu_i^t + b_{ijm} \mu_j$$

Any surplus of the left hand side over the right is imputed to the message, giving it a value β_{ijm} . This message value in turn contributes [according to equation (3.1)] to the value of information λ_{im} at the sending office i . In other words some of the cost of observation of information at i could be borne by the recipients j of communications from i . Notice that the μ_i must maintain certain proportion to the λ 's (and by implication to each other) in order to permit communication to be economical at all (equation 3.3). This implies a certain degree of equalization in the burdens of obtaining and forwarding information among the offices. In summary, the efficiency conditions (3.1)...3.4) supplied by "convex" activity analysis achieve an evaluation of the various types of information at the different offices, where they occur possibly in varying degrees of precision, and of the communications between them. They do not suffice, however, for an unambiguous answer to the question of where each type of information should be observed and to whom it should be communicated. They merely rule out certain combinations of observation and communication as involving inefficient duplication or unequal utilization of time in the various offices.

Apparently the solution of the communications problem as formulated here rests on combinatorial considerations of an essentially deeper nature than the marginal analysis attempted here.

The efficiency conditions are sufficient only in the case when each office can receive information of a given type from at most one other office. This happens when:

- 1) there are no circuitous connections in the communications network (the network is a graph of the tree family) and information of a given type

can originate in at most one office (while additions may be made everywhere).

2) Each office is the terminal of at most one unilateral channel -- while being the origin of possible several such channels, circuitry not excluded.

In each case the problem reduces to the simple one of finding the optimal strength of flow, possibly zero, along a system of pre-established paths.

5. Value and Amount of Information Related.

We have not dealt with an activity that plays a central part in the mathematical theory of information, namely coding. A model in which all messages are coded in the most efficient manner differs from the preceding one by the addition of a coding cost (value of coding time) to the costs of transmission. Detail of information may now be equated with amount of information, as defined in information theory: the minimum number of bits required to express the information in terms of an optimal code, assuming the absence of noise.

To obtain a picture of the relationship between value and amount of information in that case, let us consider a radically simple situation, that of one source disseminating information to all decision makers simultaneously. The payoffs in different offices may then be compounded into a function

$$G_{\phi} = G_{\phi}(u_1, u_2, \dots, u_M, w)$$

where u_M is the amount of information m disseminated and w the total working time. We assume that information is available to the source at zero (or a fixed) cost, and that the cost of transmission is proportional to the amount of information sent, or what is the same, to the length of time that the channels are used. Let d be the unit cost of transmission, b the time

required to receive and decode a unit amount of information and a number of offices. Then the conditions of optimality are

$$(5.1) \quad \frac{\partial g}{\partial u_m} = n \cdot \frac{\partial g}{\partial w} + d$$

In other words the marginal productivity of information is equalized for all types of information. Total value and amount of information are thus proportional for all variables. It will be observed that the relation between value and amount of information is none other than that between value and labor input in a one factor economy.