NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

REMARKS ON INFORMATION, ENERGY, AND ECONOMICS

H.L. Resnikoff

January 1982
CP-82-1

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

At the beginning of 1981 a new research task was initiated in the Management and Technology (MMT) Area of the International Institute for Applied Systems Analysis (IIASA). It was named "Impacts of Information Technology." The plan of operations for this first year was deliberately fuzzy. In addition to longer-term research staff, more than a dozen people versed in different aspects of Information Technology visited IIASA for periods of up to eight weeks. Each was asked to produce a paper and a research proposal on any topic in this field that interested him (or her). And everyone did so.

One most distinguished visitor who participated in this "plan" was Dr. Howard Resnikoff, the former Director of the Information Services Program of the National Science Foundation in Washington, D.C. and now at Harvard. During his short visit, Dr. Resnikoff completed the present, thoughtful paper. It deals, inter alia, with an issue of consuming interest to some of us and which we expect to enlarge upon in future--information economics. It is our hope that Dr. Resnikoff will be an active and leading contributor to these endeavors.

Alec M. Lee
Area Chairman
Management and Technology
Area

CONTENTS

Remarks on Information, Energy, and Economics

H.L Resnikoff

1. INFORMATION AND ENERGY

All physical phenomena, including those of life, involve the exchange of energy. In a complex system, energy exchanges are taking place all the time. If we focus our attention upon one part of the system, then some of the energy exchanges will be internal, that is, will involve only constituents of that part. Regarding the others, the part of the system commanding our attention will either absorb energy which originated elsewhere, or it will emit emergy which thereafter is, temporarily or permanently, lost to it; these exchanges may be called external. Often, the external exchanges will involve another part of system which can be conceived as a coherent subsystem upon which the first part acts or which acts upon it. In these situations there are, at a given moment, a number of different energy transfers which may be of interest: (a) the internal energy of each part; (b) the external energy emitted by one part and absorbed by the other; (c) the external energy emitted by one part and not absorbed by the other; and (d) the external energy absorbed by one part which was not emitted by the other. This last may be called background energy; it is the energy considered as noise in many applications.

The human interpretation of energy exchanges is conditioned by their relative rather than their absolute magnitudes. It probably depends on psychophysical properties of the organs of sensation. In general, we may say that exchanges will involve, in a given period of time, the absorption of a quantity of energy E_{abs} by a part of the system having internal energy exchange E_{int} will be associated with or interpreted as <u>information</u> if $E_{abs} << E_{int}$; otherwise the exchange is normally interpreted quite differently, usually being construed as or associated as an action. From this point of view actions and information are distinct and complementary. Thus the merely quantitative distinction--whether $E_{abs} << E_{int}$ or not--emerges at the level of human psychology as a qualitative distinction. This has potentially important consequences for it suggests that information can be treated on the same footing as commodities in economic theory and especially in considerations which pertain to productivity despite its apparent public good aspects.

Some examples will clarify the intent of these remarks and also make the quantitative relationship between E_{abs} and E_{int} more precise for the case of information.

Consider the parts of interest to be an atomic nucleus and a collection of laboratory instruments which includes a human observer. The instruments may also include a particle accelerator which may fire a high-energy electron or proton at the target nucleus. Under the proper conditions, the latter will be transformed under the impact and release a shower of other particles which may be detected by the measuring instruments and ultimately appear as tracks on photographic emulsions or numbers on a sheet of paper presented for the experimenter's visual inspection. From the point of view of the measuring apparatus, the radiated emissions have very small relative energy and are therefore considered as informational by the experimenter. From the point of view of the target nucleus (without ascribing anthropomophic properties to physical objects), the absorbed energy initiates a catastrophically disruptive action which disorganizes it.

These contrasting alternatives are typical and occur at every level of complexity and absolute energy exchange. For instance, visible light at normal daylight intensity when incident on the retinal receptors conveys information; at higher intensities it can cause irreversible damage to the retina, an action. But the same level of energy exchange may be informational when it occurs in a beamed telecommunication channel.

As another example, consider the relationship of a pilot to the airliner he directs. A person expends about 1.4×10^{21} ev (electron volts) per second (2.88×10^{6} calories per day) to maintain metabolic integrity whereas a large commercial airliner requires about 3.7×10^{26} ev per second (about 50,000 horsepower) for flight. Although the power consumed by the pilot is insignificant compared with that consumed by the aircraft, that small amount is principally responsible for the control and guidance--the informational--functions.

2. THERMODYNAMICS AND INFORMATION

That there is a formal analogy between information and entropy was recognized very early, and many authors have argued that the connection is more than merely formal. Here we will not repeat the arguments but refer the reader to the extensive literature.

In a thermodynamic system in equilibrium, small changes ΔE in energy correspond to small changes in entropy according to the relation

$$\Delta S = \Delta E/kT$$

where

$$k = 1.38041 \times 10^{-4} \text{ ev/degree Kelvin}$$

is Boltzmann's constant and T is the temperature of the system measured in degrees Kelvin. Entropy increments are related to information decrements by

$$\Delta I = -c\Delta S$$

where c is a constant of proportionality selected to measure information in convenient units. If information is measured in bits, then

$$\Delta S = -9.568 \times 10^{-5}\ \Delta I\ .$$

Thus substantial changes in the quantity of information correspond to small entropy changes. From the above relations we find

$$\Delta I = \frac{-1}{k \log 2}\left(\frac{\Delta E}{kT}\right)$$

where log denotes the natural logarithm. At ordinary temperatures, $kT \sim 4x10^{-2}$ ev, whence each energy decrement of 1 ev corresponds to about $2x10^{5}$ bits of information. This shows that a small change in energy can correspond to a great deal of information; conversely, a great amount of information counts for very little from the standpoint of energy-intensive activities, for 1 ev is approximately the amount of energy required to excite a bound electron in an atom without causing ionization. This is one reason why information was not considered an important matter in pre-electronic times.

3. INFORMATION AND THE PROCESS OF MEASUREMENT

Information enables one to distinguish amongst alternatives. This concept can be reduced to the selection, perhaps by measurement, of one category from a collection of alternatives. In the framework of measurement, we may think of a linear scale, say the interval [0,1], subdivided according to positional notation relative to, say, base 2. Then each point in [0,1] corresponds to a binary expansion

$$a = 0.a_1a_2\ldots a_n\ldots$$

where each a_i is 0 or 1 (and the expansion is non-terminating, i.e., is not ultimately the sequence all of whose digits are 0).

A measurement consists of the specification of a finite number, say N, of the initial digits of the expansion of a:

$$M_N(a) = 0.a_1a_2...a_N \quad ,$$

and such a measurement may be said to be accurate to within one part in 2^N. That is, one amongst 2^N alternatives has been identified by the measurement, where the alternatives are the 2^N distinct sequences of N binary digits. These numerical alternatives are conceived as corresponding to distinct physical states capable of being measured by the apparatus.

Since the extension of the binary expansion of a from N to N+1 binary digits amounts to identifying one new bit, we may say that the expression $0.a_1a_2...a_N$ contains N bits of information. Thus the amount of information provided by the N-bit expression $0.a_1a_2...a_N$ is just

$$N = \log_2 2^N = \log_2(\text{Number of alternatives}) \quad .$$

We have implicitly assumed that the alternatives are equally likely. In this case, the probability of obtaining the expression $0.a_1a_2...a_N$ as the result of the measurement process is $p = 1/2^N$ whence the amount I of information it yields can be computed as

$$I = -\log_2 p \quad .$$

If the different N-bit expressions are not equally likely, then let p_k , $k = 1,2,...,2^N$ denote the probability of occurrence of the k-th expression in some fixed ordering. Then the information provided by the measurement is the weighted average of the $\log_2 p_k$, i.e.,

$$I = -\sum_k p_k \log_2 p_k = -\frac{1}{\log 2} \sum p_k \log p_k \quad ,$$

where the sum runs over all possible states. This is the measure of information introduced into communication theory by Shannon. We see that it is related to the concept of measurement in a basic and intrinsic way.

4. INFORMATION ASSOCIATED WITH A PARTITION

We will generalize the assignment of a quantity of information to a measurement process described above so that it can be applied to more varied situations, restricting our attention to discrete countable collections of states.

Let S denote a set of states, which may be states of a physical system such as an atom, or a collection of weakly interacting molecules which constitute a gas, or states of an organization or other societal system, etc. Let K be some index set, and for each $k \in K$ let S_k be a subset of S. Suppose that the S_k are disjoint ($S_k \cap S_l \neq \phi$ implies k=l) and that they exhaust S: $S = \bigcup_k S_k$. Let p denote a probability measure on the subsets of S such that the S_k are measurable and define the information associated with the partition $\{S_k\}$ to be

$$I = \frac{-1}{\log 2} \sum_{k \in K} p(X_k) \log p(X_k) \quad ;$$

the factor 1/log 2 insures that I is measured in bits. Thus each partition and probability distribution correspond to a certain quantity of information. Hereafter we will suppose that the probability distribution is given; usually it will be the one derived from counting measure.

5. CONSTRUCTION OF PARTITIONS

One important means for constructing partitions makes use of functions. Let V denote a set and $f: X \to V$ a function. For $v \in V$ denote by $f^{-1}(v)$ the set of all $x \in X$ such that $f(x) = v$; if v does not belong to the image of f, then put $f^{-1}(v) = \phi$. The non-empty sets amongst $\{f^{-1}(v): v \in V\}$ form a partition of X; we will often write $X_v = f^{-1}(v)$. If vol (read "volume") denotes a measure on X (e.g., counting measure), then define

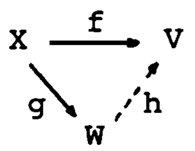
$$p(X_v) = p(f^{-1}(v)) = \frac{\text{vol } f^{-1}(v)}{\text{vol } X} = \frac{\text{vol } X_v}{\text{vol } X} \quad .$$

so the information associated with the function f is

$$I = I_f = -\frac{1}{\log 2} \sum p(X_v) \log p(X_v)$$

where the sum runs over those $v \in V$ which can be written as $f(x)$ for some $x \in X$.

Suppose that $f:X \to V$ and $g:X \to W$ are two functions and that the associated partitions of X are $X(f) = \{X_v\} = \{f^{-1}(v):v \in V\}$ and $X(g) = \{X_w\} = \{g^{-1}(w):w \in W\}$. The corresponding measures of information I(f) and I(g) (where the probability measure is the same for both) are equal if the partitions coincide, and the partitions coincide if and only if there is a 1-1 mapping $h:W \to V$ such that the following diagram is commutative:



,

i.e., $h(g(x)) = f(x)$. From the standpoint of partitions and their associated information measures, the functions f and g are indistinguishable, but we will see in the next section that there are ways in which the distinctions between functions are reflected in the analysis of the properties of the corresponding informational structures.

6. CONSTRAINTS

Having more information is usually thought of as a good thing. This suggests that one should seek to maximize the information measure I, which is the same thing as seeking a function which produces a partition of X corresponding to the maximum information. This point of view is unrealistic because there is no mechanism provided to account for the costs or constraints associated with different partitions. The purpose of this section is to formulate such a mechanism analytically.

We will assume that $f:X \to V$ assumes numerical values in the field of real or complex numbers. In this case the k-th moment of f is defined by

$$m_k(f) = \sum_{v \in V} p(v) v^k \quad , \qquad k = 0,1,2...$$

where p is the fixed probability measure associated with X.

Note that

$$m_0(f) = \sum p(v) = 1$$

since p is a probability measure. $m_1(f) = \sum p(v)v$ is the average value of f which we will also denote $\langle f \rangle$.

Now let K denote some subset of the set of positive integers and suppose that for $k \in K$, the values $m_k(f)$ are fixed. We will consider the family of functions f for which these constraints hold, and seek to maximize I(f) for functions in this family, i.e., subject to the constraints $m_k(f) = c_k$, $k \in K$, where the c_k are constants. Recall that in all cases, $m_0(f) = 1$ holds.

The method of Lagrange multipliers applies and calls for the maximization of

$$-\frac{1}{\log 2} \sum_v p(v) \log p(v) - (\lambda_0 - \frac{1}{\log 2}) \sum_v p(v) - \sum_{k \in K} \lambda_k \sum_v p(v) v^k \quad ,$$

where the λ_k are the Lagrange multipliers. Then

$$\log p(v) = - \log 2 \{\lambda_0 + \sum_{k \in K} \lambda_k v^k\} \quad ,$$

$$p(v) = \exp(- \log 2 \{\lambda_0 + \sum_{k \in K} \lambda_k v^k\}) \quad .$$

The λ_k are determined by the constraints $m_0 = 1$, $m_k = c_k$.

7. EXAMPLES

(1) Suppose X denotes a collection of weakly interacting molecules of one type which forms a gas, and that $f(x) = E(x)$ denotes the energy of the molecule labelled x. Further suppose that $m_1(E) = \langle E \rangle$ is fixed; this is the average energy of the molecules and, if the number of molecules is held constant, it is also a measure of the total quantity of energy available for partition amongst the molecules. The set X_E of the partition is the set of molecules of energy E and p(v) is the probability that a molecule has energy v.

Information is maximized subject to the constraint $m_1(E) = \langle E \rangle$ if the distribution of molecules amongst energy states is given by the formula

$$p(v) = \exp(-\{\lambda_0 + \lambda_1 v\}) = A \exp(-\lambda_1 v) \quad ;$$

this is the Boltzmann distribution of statistical thermodynamics. The constants are obtained from the relations

$$1 = \sum_v p(v) = A \sum_v \exp(-\lambda_1 v)$$

and

$$\langle E \rangle = \sum p(v) v = A \sum_v v \exp(-\lambda_1 v) \quad .$$

(2) A related example exhibits some interesting features. Consider the same gas confined to a thin cylindrical tube so that motion is one-dimensional and let $v(x)$ denote the velocity of the molecule x. We will suppose that the gas is contained in a fixed container and that its center of mass remains fixed. This means that the average value v is 0: $\langle v \rangle = 0$. The function $v:X \rightarrow V$ defines a partition but of great interest is the partition defined by composing v with the function $v \rightarrow 1/2mv^2$ where m denotes the mass of a molecule; $1/2mv^2$ is the energy of the molecule. If we assume that the total energy and hence the average energy is fixed, then $m_2(v) = \langle E \rangle$ is constant. We also have $m_1(v) = 0$ and $m_0(v) = 1$. Hence the distribution which maximizes information subject to these three constraints is

$$p(v) = \exp(-\lambda_0 - \lambda_2 v^2) \quad ;$$

for an appropriate choice of parameters λ_0 and λ_i this is Maxwell's velocity distribution of statistical thermodynamics.

(3) Functions $f:X \rightarrow \underline{\underline{R}}^+$ which are real valued and correspond to physical or conceptual quantities which are inherently positive offer the possibility of further transformation to spread the function values across the whole real number line and thereby provide the opportunity to approximate the function

by Gaussians by means of Hermite expansions, to apply the standard central limit thorem, etc. This can be achieved by composing f with the logarithm function, thus,

$$X \xrightarrow{f} \underline{\underline{R}}^{+} \xrightarrow{\log} \underline{\underline{R}} \quad , \qquad g = \log \circ f$$

$$g(x) = \log f(x) \quad .$$

If constraints are more naturally expressed in terms of the moments of g rather than the moments of f, the distributions which optimize I(g) will often be related to lognormal probability distributions.

These circumstances apply to the case of ensembles of documents of books where x denotes, say, a document and f(x) denotes its size measured in text character equivalents, but in this situation the constraints are most naturally expressed in terms of moments of the distribution of the logarithm of the size of the documents. Here one finds that $m_2(g) = m_2(\log f)$ is fixed as well as $m_0(g)$ and $m_1(g)$; the distribution which maximizes I is then lognormal, in agreement with observations. It seems that the logarithm of size is an approximation to the cost of using a document to retrieve information, although the reason why this should be so is not yet entirely understood; cp. section 10.

8. PATTERNED STRUCTURES

We have classified energy exchanges into categories: those which are relatively large compared with internal energy transactions were interpreted as related to actions while those which are relatively small as related to information. The communications engineering point of view treats all signals in a common way, not distinguishing their utility or value beyond the coarse estimation of their degree of organization which is provided by Shannon's information measure. Yet, for a given individual, some signals, i.e., some information, have a subjectively great value whereas other information appears to have little value. The value or utility of information certainly depends on the properties of the receiving individual as well as on the properties of the transmitted signal. Consequently we may adopt two extreme viewpoints: one is to study utility and value of information in the context of the properties of a given individual; the other is to average over the properties of all individuals and study the averaged utility and value of information.

The former approach is most suitable for studies of the psychophysical internal processing of information transmitted to the brain by the sensory organs, and analyses of how the brain operates on sensed information to coordinate it with, and incorporate it into, the internal knowledge state of the brain. Some of the recent work in the cognitive psychology and artificial intelligence branches of information science proceeds along this path.

The latter approach is generally more suitable for studies of the economic properties of information and of information systems which are intended to interact with large numbers of users, for which design optimization can only be approached from the standpoint of optimizing average rather than individual performance. It is this essentially statistical theory which can call upon related methods and results from the engineers' statistical theory of communication and from statistical thermodynamics.

The first of the two approaches includes problems related to patterned structures. Intuition leads us to believe that the recognition and classification of patterns in a signal are essential elements in utilizing information. From the signal processing point of view, rare structures correspond to the largest measures of information; of course, the universe from which the signals are considered as having been drawn plays a role. Nevertheless, even random sequences of information can be valuable as, for example, when they are used as seeds for Monte Carlo calculations or to provide a basis for hypothesis testing. The RAND Corporation publication of one million random digits in the 1950's was a costly and valuable example of the utility of unpatterned information. This shows that noise information cannot be considered as without utility in a comprehensive treatment of information and that the goals of an information system designer may not always include elimination of noise.

The creation of a theory of patterned structures remains one of the key and most elusive problems in information science. It may ultimately be conceptually reduced to a problem of signal coding in the sense that to assert the relatedness of two patterns may be the same as asserting that they contain certain primitive or "atomic" constituent patterns in common which can form part of the "code" for both. But at the present time it cannot be maintained that there is a viable theory of pattern structure at this level of abstraction, although progress has been made in identifying primitive "atoms" for human vision.

The process of learning something new evidently involves detection of pattern relatedness, for it is not possible to learn something which is entirely disjoint from one's pre-existing fund of knowledge. Thus the development of a theory of patterned strucutres and pattern classification is fundamental for gaining a deeper understanding of the process of learning, and for far-reaching applicationsof the new information technology to education.

9. HIERARCHICAL STRUCTURES

Amongst the patterned structures hierarchical structures appear to play a special role. Their advantageous properties for the design of manufacturing and other organizational systems has been set forth in a cogent way by Herbert Simon. Their role in information systems <u>per se</u> is equally important. Indeed, one may adopt the view that the creation of complex manufactured objects requires a certain quantity of information; the product can be conceived as consisting in part of embodied information, and the systematic organization of the manufacturing process can be thought of as providing an efficient means for embodying that information in the physical product.

Hierarchical systems occur frequently because they possess a certain optimality property with regard to the retrieval of information, and this property can be transformed into time savings and quality improvement, and hence lower costs, in production of complex products. The optimality property can be summarily described as follows.

By a <u>hierarchical structure</u> we mean a tree. Suppose the tree consists of n+1 levels, with the level of the root labelled 0 and the level constituted by the leaves (i.e., terminal nodes) labelled n. Nodes of the tree will be labelled by k-tuples of numbers, where $k>0$ denotes the level of the node. Thus $S(i_1,...,i_k)$ will stand for the k-th level node belonging to the sub-tree whose root is the (k-1)th level node $S(i_1,...,i_{k-1})$ is the i_k-th node with parent $S(i_1,...,i_k)$ in some ordering, with $1 \leq i_k \leq N_k$; the root of the tree will be denoted S_0. Thus N_k is the member of nodes of level $k \geq 1$. We have $N_0=1$ and N_n = total number of leaves.

Each node will be thought of as corresponding to a data object. A data object at level k will be thought of as containing aggregated descriptive information, or descriptors of, the data objects corresponding to nodes of level k+1 which belong to the subtree determined by the given node. The data base itself will be identified with the leaves of the tree. An information retrieval task will be identified with a search that begins with S_0 and terminates with a leaf $S(i_1,...,i_n)$. The search proceeds inductively by sequentially searching nodes at level k until the appropriate data descriptor is found which switches the search to the subtree determined by that node, whereupon the level (k+1)-nodes belonging to that subtree are sequentially search. The search begins with S_0.

The optimal number n of levels and nodes N_k at level k, i.e., the arrangement which minimizes average search time, can be determined for a given number of leaves in terms of the size of the data objects (which is proportional to the required search time) and the search speed, which is assumed to be a function of only the level.

First suppose that the number n of levels is fixed. Then the solution shows that N_{k+1}/N_k is a constant which depends only on the relative search speeds. If all speeds are equal, then

$$N_k = (N_n)^{k/n} \quad , \qquad k=0,1,...,n.$$

Note that the information provided by the nodes at level k will be proportional to k, with the factor of proportionality being the product of the average size of a data object of level k by the factor log N_n/n.

The number n+1 of levels can also be optimized but we will not discuss this question further than to observe that it involves a straightforward minimization problem of differential calculus.

10. CONSTRAINTS ASSOCIATED WITH HIERARCHICAL STRUCTURES

The average time, or number of operations, required to retrieve a data object from a data base containing N objects will be proportional to N if the search passes sequentially through the objects in the data base. Access systems which classify or categorize the data objects can significantly reduce the average retrieval search time. For the hierarchical access systems described in the previous section, the average search time varies approximately as the product of number n of levels and the n-th root, $N^{1/n}$, of the number of terminal nodes. If n log N, then search time will be proportional to n, i.e., essentially to the number of levels. Suppose constraints are expressed in terms of search time costs. Since these consts are largely proportional to search time, it follows that for hierarchical systems of the type described they will be proportional to the logarithm of the size of the data base (measured, i.e., in bits) since the latter is proportional to the number N of terminal nodes for data objects of roughly comparable size. Thus we are led to measure constraints as functions of the logarithm of the size of the data objects.

Writing s(x) for the size (in bits) of the data object x, constraints will be expressed in terms of the moments $m_k(\log s)$.

The condition that a fixed and limited time is available for search is equivalent to $m_1(\log s)$ = constant. Since the data objects must be of comparable size for the hierarchical access system model to apply, we may also suppose that $m_2(\log s)$ is essentially constant. These constraints produce the lognormal distribution for the size function $s:X \rightarrow \underline{R}^+$ for each level of a hierarchical information retrieval system.

11. INFORMATION AND ECONOMICS

We are now prepared to consider the relationship of information and economics. There are several important themes which have been discussed by numerous investigators. They fall into two main classes: (1) the role of information in economic theory, and (2) the role of information as an economic good.

The importance of the first of these arises in the following ways. To the extent that decision making by economic actors involves risk and uncertainty it involves, equivalently, lack of information. Thus measures of the quantity of information are simultaneously measures of the degree of uncertainty or risk, and strategies which minimize risk can be construed as those which maximize information. Since information can normally be obtained by increasing the resources devoted to its acquisition, the desire or necessity to acquire information will compete with the desire or necessity to acquire commodities (other than information) or make investments. In this way natural constraints are associated with the acquisition of information. These constraints can usually be expressed in terms of a cost function. It may happen that the information needed in order to make a decision is implicitly available but that the calculations which must be performed in order to make practical use of it are too complex to be performed in the time available. This is another type of constraint which can be ameliorated by advances in the techniques and technology of computation, but in any event constraints of this type are more likely to be expressed as time constraints rather than cost constraints. This is an essential distinction, because although in many circumstances time is money, money may not be time if it is not known how to do what must be done within the allotted period. From a more general standpoint, it may be reasonable to interpret this as an information deficiency as well.

The commodity aspect of information has already shown itself in the brief description given above of the connection between information and economic theory, but it involves enough thorney issues only indirectly related to its role in economic decisions that it deserves independent attention.

Information is the control that directs all the activities of the individual as well as of society. At the level of the individual, muscles without the guidance of a brain will not accomplish much; nor will the guidance a brain provides be of value unless it can effectuate physical actions in some way. There is an inherent assymetry in the relation of muscle- to mental-power: we are willing to invest effort to develop prosthetic replacements for injured action-elements of the human body, barring none, but failure of the brain has not yet attracted a similar effort, although this may change as the performance characteristics and energy and volume requirements of microelectronic components become competitive with and surpass those of the brain's neuronal constituents.

Be that as it may, the guidance and control functions of life and civilization rely on information, and on the evaluation and analysis of information, and this provides information with a value, which will nevertheless be a function of the particular circumstances. We have already seen that a small amount of energy, properly organized, will correspond to a large amount of information. It follows that a society which concentrates on the energy-intensive production and distribution of physical goods will find the cost of the information needed for the fabricatation, distribution, acquisition, or use of an energy-intensive physical commodity to be a very small fraction of the cost of the energy associated with the commodity in this sense. For this reason, the consideration of information as a commodity has been relatively recent and even today there is debate on this question. In summary, however, we may say that the relative cost of information has been generally viewed as trivial by comparison with other costs, except in special cases.

There is another, more abstract, difficulty associated with considering information as a commodity. Unlike most other commodities, when one individual communicates information to another, the result is that <u>both</u> posses the information. It does not work that way with energy-intensive physical objects like automobiles or food. If you ask me for the time, I will normally tell it to you; but asking for my watch is another matter.

If an item of information (such as the time) has a positive value, by telling it to you, it would appear that I have created value from nothing whatever. Were this true, a simple and convenient scheme for increasing the gross domestic product would be to have everyone in the nation shout out the time or, since that is a perishable commodity, the enduring multiplication table.

The absurdity of these remarks coupled with the recognition that energy dominates information in physical objects suggest that items of information have a positive if small value and can be tallied in a general accounting of an individual's or firm's or nation's economic activities. From this standpoint one may interpret the teller of time as having made a small investment rather than having distributed a free good. Moreover, if the cost of the information item--the cost of the telling, so to speak--is small enough, then its value compared with the value of energy-intensive activities will be so small that repetition of the offering will not increase gross domestic product in a significant way. Thus, the cost of individual information transfers can be so small as to make them appear "free" or "public" goods, but in fact the cost is positive and can be subjected to normal accounting procedures as part of the investment strategy of the individual or firm.

12. INFORMATION PRICES

Everyone pays for information but its price is generally not well established. This has also been a source of confusion which has obstructed the integration of information as a commodity in economic theory and the formulation of public policy concerned with information in education, in regulated information-industries such as telecommunications, and within government itself.

One reason that the price of an item of information may have a widely fluctuating or indeterminate value is that its utility is to a greater degree a function of the purchaser's or user's internal state of knowledge than in the corresponding case for physical commodities. By and large, physical commodities, such as food or tools, can be used by large segments of the population in known and largely similar ways, and markets for such products will be broad and statistically stable. Consequently the substitutional equivalence of such commodities as reflected by the price function is relatively clearly understood and competition or regulation will insure that the price spread will be relatively narrow. It is not the same with information. Whereas we all know what to do with an auto or a cup of coffee, we do not all know what to do with a description of the genetic code or with the equations of general relativity or with the proof of the classification of finite groups. And whereas competitive suppliers of automobiles and coffee stand ready to tell us, through ubiquitons advertising, what to do with their products should be not already know, no-one stands ready to broaden our horizons about the uses of the classification of finite groups. So there is inherently a small market for most information goods, especially for those products whose use is idiosyncratic, whose validity may not be easily verified, and whose value is very much tied to the internal knowledge-state of the prospective purchaser. Thus one should expect wide price fluctuations or price indeterminancy for certain information products, e.g., scholarly papers, consulting services, and stock market tips. But the price range of a newspaper or admission to a motion picture theater is relatively narrow and stable regardless of their content although special circumstances arise even in these cases.*

Information products have value not only intrinsically but also in connection with their intended, or most probable, mode of application. Thus the information accumulated through experience by the aircraft pilot or the surgeon is probably available at relatively low cost in books, but high salaries and fees are paid when that information is applied in connection with risky situations which involve costly energy-intensive

* The price of the final edition of the recently defunct Washington Star rapidly rose to the hundreds of dollars.

products and consequences, and where competitive information sellers are not available when needed. Because the immediate energy-intensive risks and opportunities are much less in their realm of activity, and there is substitutibility of sellers, primary and secondary school teachers are paid rather less, but even here the most important points to consider are that their task essentially consists of replicating information and modifying internal knowledge states, that the curriculum--i.e., the information content--is highly standardized, and that it is not distributed as a free good although it is widely known and available. In fact, teachers are paid precisely to insure that a certain body of information is widely replicated. This is a situation in which the information value presumably increases as its scarcity decreases, in distinction to most commodities. Value as a function of scarcity may approach an asymptotic value, or even have a maximum, for very small measures of scarcity. The foundation of an "information accounting" analysis might be found in the educational area.

13. PRODUCTIVITY AND INFORMATION

In recent years the United States has experienced declining productivity increases, and absolute productivity decreases. This has occurred during a period of rapid expansion in the use of information technology throughout all aspects of the American economy. Numerous analyses and discussions have been published but the relationship of information technology to productivity, and the problems associated with the measurement of productivity changes which accompany adoption of one or another information processing machine have not been adequately explored.

Productivity is the numerical ratio of some measure of economic output to a measure, in comparable units, of some factor or factors of input required to produce the output. The traditional factors of input are labor, capital, and land. Land is required for all forms of production, either directly as in traditional agriculture or indirectly as a surrogate for physical resources (including energy and materials), and also for situating the production labor and capital. Capital is conceived by some economists as stored labor. If this view is adopted, we see that the three factors of input are all expressible in terms of energy and physical resources, and hence ultimately in terms of energy. But the fabrication of economic goods involves more than materials and energy; it also requires information to control and direct the process of organizing material resources by the application of energy into desirable configurations. The quantity of information required to produce a product may be large or small in an absolute sense, and relatively large or small compared with the energy and material content, but it cannot be done without. One of the characteristic features of contemporary industrial civilization is that the information content of manufactured items is increasing rapidly compared with the energy and materials content. Information technology itself provides striking examples of this assertion: the information necessary to fabricate an integrated circuit microprocessor is very large compared with the energy and materials consumed by the manufacturing process. Thus the labor and materials intensive mechanical calculators of fourscore years ago have yielded place to the information intensive pocket calculator of vastly superior capability and strikingly lower price.

Now that information-intensive products are forming an increasingly important part of the economic mix, it seems timely to recognize the role of information in productivity by explicitly identifying it as an input factor of production. In many traditional areas of the economy the productivity contribution of this factor may be negligible but in others it will be an important term.

No one yet knows how to directly measure the effects of information on output, although it may be obvious that the effects are large. Indeed, most segments of the economy of the United States which show a positive export trade balance are those for which the products have a high information input and a relatively low labor and materials input: commercial aircraft; computers, telecommunications, and other information technology; selected military equipment; agricultural products incorporating chemical fertilizer and plant genetics information; and biologicals for health care. Yet it is correct to claim that economic theory has no solution for the problem of measuring the quantity of information incorporated in these products nor of specifying the degree of investment in information appropriate for product improvement or innovation. I have no solution to propose either. But this is a topic which would benefit from the attention of economic theorists.

Let us return to consider productivity ratios more carefully. If the measured output of product y is denoted OUT(y) and the input due to factor x is denoted IN(x), we may write

$$P(y:x) = OUT(y)/IN(x)$$

for the productivity of factor x in product y. Thus P(y:labor) will denote labor productivity associated with product y. Tricky questions remain before productivity ratios can be used: shall labor be measured by the time consumed, e.g., labor hours, or by the money value of the labor? Shall output be measured in number of units produced, or market money value of the units, or production cost of the units? These matters must be considered when a productivity measurement system is created, and they have been treated in detail by others; I do not propose to consider them further.

Suppose that $\{y_i : i=1,...M\}$ are the various products produced by an economy and $\{x_{ij} : i=1,...,N\}$ are the various inputs. Then $P(y_i : x_{ij})$ is the productivity of the j-th input factor in producing the i-th output product. Suppose that all measures of output quantities are expressed in comparable units, and that all measures of input factors are also measured in comparable but possibly distinct units. It will be convenient for us to think of both measures as money equivalents so that the productivity ratios are pure numbers.

The total output of the economy will be $\sum_{i=1}^{M} OUT(y_i)$ and the total input required to create this output is $\sum_{i=1}^{M} \sum_{j=1}^{N} IN(x_{ij})$.

Thus the productivity of the economy as a whole will be

$$P = \sum OUT(y_i) / \sum \sum IN(x_{ij}) \quad .$$

P may of course differ significantly from some of the partial ratios $P(y_i:x_{ij})$ or from ratios which combine inputs corresponding to a given output such as $P(y_i:X)$ where X is some subset of the inputs x_{ij}, =1,...,N.

To increase the productivity of an economy, or of a sector of an economy, there are two and only two independent strategies: (1) increase the numerator of the corresponding productivity ratio, and (2) decrease its denominator. Although economic planners normally desire to increase the productivity P of the economy as a whole, what in fact occurs is that the sectoral productivity ratios $P(y_i:x_{ij})$ are the ones which can be modified by changes in policy, management, organization, and technology, but normally not all these ratios can be increased, and some of them are adversely affected by improvements in others, and there are "side effects" which must also be taken into account.

For instance, let y_{max} denote that product whose labor productivity ratio P_{max} is greatest, let y' denote the combination of all other products, and let OUT(y')/IN(labor) = P' be the corresponding labor productivity ratio. One strategy for increasing the productivity of an economy is to eliminate production of all the products y'; the aggregate productivity of the economy, $P = \frac{OUT(y_{max}) + OUT(y')}{IN(labor)}$ will then rise to the value P_{max}. This of course implies that the labor previously devoted to producing y' must either be reallocated to the production of produced in sufficient quantity to meet demand, which we may suppose from the outset, then the labor previously devoted to y' will indeed become unemployed. Thus strategies which seek to increase productivity without any constraints concerning the deployment of labor freed by disinvestments in low-productivity industries will necessarily lead to high unemployment and low total output as well as high productivity. We can account for this possibility by thinking of unemployed labor as labor which is devoted to the production of the "zero" product y_0, that is, a product whose contribution to the output numerator of the productivity ratio is 0. Thus P(y_0:labor) = 0, but unemployed labor will increase the denominator of the productivity ratio P corresponding to the economy as a whole and will consequently reduce this aggregate productivity measure.

In this context we must also note that introduction of labor producing the zero product into our calculations of productivity also forces us to consider how to account for labor that is not employed to the maximum extent, and whether reductions in the standard work week should be compensated in calculations of productivity by incorporating fractional amounts of labor devoted to the zero product corresponding to the difference between actual hours worked and a hypothetical maximum number of hours (per week, say) which could have been worked. Were this done, productivity ratios would decline, corresponding to the product that went unproduced during leisure hours. This "adjustment" may not satisfy our intuitive judgment that only hours devoted to actual production should figure in the productivity ratio, but then the option of maximizing productivity by creating unemployment in low productivity industries reappears. This delemma suggests that the productivity measure does not fully capture the relationship between the efficient production of commodities and the deployment of the labor force.

These difficulties are further exacerbated in the information sector and others which are heavily dependent upon it because the measurement of output is hard to define. Thus the productivity of a copy machine can be measured by the ratio of the number of copies per minutes that it can produce to the cost per copy. The latter will incorporate the cost of paper, which we will assume fixed, and the amortized cost of the machine per page. Productivity will rise as the speed of the machine rises relative to its cost. The labor productivity of the office copy machine user will also increase as the speed of the machine increases, if productivity is measured as the ratio of pages copied to labor hours required. But this accounting scheme makes no provision for assessing the value of the copied pages, that is, for assessing their contribution to the labor costs of the marketed products of the organization. And the availability of this piece of information technology may encourage the distribution of copies of documents to persons in the firm or other organization who do not have need for them for the fulfillment of their own responsibilities and for whom, consequently, this information may be functionally equivalent to noise and actually reduce their productivity.

In this regard it is important to realize that machines which process information are, to a much greater degree than others, neutral with respect to productivity improvement. Whereas a new chemical fertilizer or machine tool will only be introduced if its contribution to productivity exceeds that of the productive means it replaces, in most cases information technology can be used either to increase or to decrease productivity. Productivity improvement is not inherent in the equipment itself; advances in information technology only assure more cost effective processing of information without regard to its productive utility. In particular, the new information technology has the potential to encourage expansion of the administrative and audit functions of large bureaucracies, governmental and industrial, without the creation of an identifiable product that can appear in the numerator of the productivity ratio. Thus the use of a word processor to increase the number of memoranda that can be typed by a secretary by a factor of as much as three does not imply the necessity for that many additional memoranda, although the tendency is to increase this type of output rather than to decrease secretarial employment. One reason for this trend and recent rapid growth in the information sector of the economies of the advanced industrial nations is undoubtedly that information work in organizations is easily expanded so that, as productivity in the agricultural and industrial sectors increases, labor is transferred to the service sector, to the information sector, and to the ranks of the unemployed. Insofar as the utility of information products generated within an organization cannot be identified as increasing productivity of its distributed output, labor used to generate those products can be thought as a surrogate for unemployment, with the depressing effect on productivity described above.

We have said that information technology is essentially neutral with regard to productivity; it can be as easily used to decrease as to increase it. The rapid advance of this technology makes it imperative that means be found to adapt the structures of organizations to encourage the productive, and constrain the non-productive, application of information technology. Underlying an effective policy for the proper application of information technology will be a theory, and measures, of the productivity of information technology and of the value of information relative to the context wherein it is used.