Toward a Political Economy of Information Capital

by

Lawrence Arie Hirschhorn

B. A., Brandeis University

(1967)

SUBMITTED IN PARTIAL FULFULLMENT

OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF

PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

September, 1971

Signature of Author. .....
Department of Economics, September, 1971

Certified by..... .....
Thesis Advisor

Accepted by..... .....
Chairman, Departmental Committee on Graduate Students

Archives
MASS. INST. TECH.
SEP 10 1971
LIBRARIES

ACKNOWLEDGMENTS

I would like to thank Professor Kindleberger for his receptivity to what were initially broad and unstructured ideas. His ability to cut to the core of a concept gave me much encouragement and help. I would like to thank Professor Foley for his continued insistence on rigor. I would like to thank Mrs. Karolyi who made this entire manuscript readable.

Special thanks and deep gratitude are due Professor Chandler H. Stevens. Hiw own work first led me into the field of "information economics". He encouraged me in my work even when my initial proposals were most ambiguous, and gave me the opportunity to participate in a lively seminar on communications in the Fall of 1970.

"No social order ever perishes before all the productive forces for which there is room in it have developed; and new, higher relations of production never appear before the material conditions of their existence have matured in the womb of the old society itself.

Therefore mankind always sets before itself only tasks as it can solve; since, looking at the matter more closely, it will always be found that the task itself arises only when the material conditions for its solution already exist or at least are in the process of formation."

Karl Marx, Preface to the Critique of Political Economy

"In conveying the essential departure of cybernetics from traditional physics, Norbert Wiener once noted:

> Here there emerges a very interesting distinction between the physics of our grandfathers and that of the present day. In nineteenth century physics, it seemed to cost nothing to get information.

In context, the reference was to Maxwell's Demon. In its economic reincarnation as Walras' auctioneer, the demon has not yet been exorcised."

Axel Leijonhufvud, On Keynesian Economics and the Economics of Keynes

TABLE OF CONTENTS

Page

ABSTRACT 2

ACKNOWLEDGMENT 3

INTRODUCTION 7

INFORMATION CAPITAL -- An Introduction from an Anecdotal Perspective

The Hardware 7
The Software 12

CHAPTER ONE

INFORMATION CAPITAL -- from the Perspective of Market Equilibrium

THE MARKET AS AN INFORMATION SYSTEM 16

CHAPTER TWO

INFORMATION CAPITAL -- from the Perspective of Organizational Growth

CHANNEL CAPACITY, ORGANIZATIONAL GROWTH AND THE EMERGENCE OF INFORMATION CAPITAL 29

Introduction 29
§1. The Organization as an Information System 31
§2. Channel Capacity as Minimum Average Cost 44
§3. Growth Through Structural Differentiation 57
§4. Growth Through Restructuring and the Emergence of Information Capital 68

APPENDIX 78

CHAPTER THREE

INFORMATION CAPITAL -- from the Perspective of Software

THE MODEL AS A CAPITAL GOOD 84

Introduction 84
§1. An Intuitive View of Structure and Transmission 86
§2. The Sampling Process, Rank Value and the Cost of Perfect Information 95
§3. Transmission with Error 107
§4. Structure and Rank Value 117
§5. Some Special Properties of Information Capital 127

Page

CHAPTER FOUR

INFORMATION CAPITAL -- from the Perspective of Conflict

ON GROWTH AND CONFLICT 131

ADDENDUM 145

CHAPTER FIVE

INFORMATION CAPITAL -- from a Radical Perspective

INFORMATION AND THE SOCIALIZATION OF THE MEANS OF PRODUCTION 147

§1. The Demise of the Market 147
§2. The End of Economics 157

APPENDIX: THE BRIEFEST INTRODUCTION TO INFORMATION THEORY 166

BIBLIOGRAPHY 172

BIOGRAPHICAL NOTE 177

INFORMATION CAPITAL

An Introduction from an Anecdotal Perspective

INTRODUCTION

The Hardware

It is common knowledge to the Wall Street buff that computer manufacturing and data processing are the fastest growing economic activities in the private sector. From 1958 to 1968 the rate of growth in sales of the computing and office equipment industry far exceeded that of any other industrial activity.[1] From 1955 to 1963 the number of computers in use grew from 500 to 16,000[2] (10% operated by the Federal Government), while from 1960 to 1969 the sales of IBM grew from two to almost eight billion dollars.[3]

As of October 1969, the telecommunications industry was selling its services at a twenty billion annual rate, and were growing at 20% per year.[4] This growth in both hardware and software led to a growth rate in business data transmission of about 25% a year such that by 1980 it is estimated that 10% of all the economy's network capacity (including telephone and cable T.V.) will be shuffling data alone.[5]

In 1968 the sales of information services tied to data processing stood at two billion and was growing at a 25% rate.[6] It is estimated

---

[1] "Who has Ridden the Growth Curve?", Business Week, Jan. 4, 1969, p. 20.

[2] The Diebold Group, Automation, Impact and Implications, Communications Workers of America, AFL-CIO, April, 1965.

[3] "Two Gee Shiz Giants Go at Each Other", Business Week, June 13, 1970, 70.

[4] "Growing Wildly Without Blueprint", Business Week, Dec. 6, 1969, 192.

[5] Op. cit., 192.

[6] "A New Industry's Wild Ride", Business Week, May 24, 1969, 64.

that a $260 billion ten-year investment is required to develop the information systems required for the 1980's.[7]

From one perspective these facts may be indicative only of another "growth industry", simply another sign of technological change in the economy. But from another and more fruitful perspective, we can view them as part of the fast development "information infrastructure" for the burgeoning "communications revolution". Just as the industrial revolution was proceeded by infrastructure investment in canals, railroads, and new energy sources, so does the communications revolution rest on the development of "society's channel", on its ability to process and model increasing flows of information through its information networks. In this context the growth in computer sales and data processing services can be analyzed in tandem with the developments in the communications, media, and electronic processing services.

Perhaps most symbolic of these developments has been the growth in cable T.V. There are now 2,400 cable systems in the U.S. serving 12 million viewers or about 6% of the television audience.[8] Industry representatives estimate that by 1980 cable will serve close to 100 million viewers generating an annual revenue of three billion and employing 750,000 men and women.[9] The power and significance of cable lies, of course, in its channel capacity. Present rates of

---

[7]"Growing Wildly Without Blueprint", Business Week, Dec. 6, 1969, 192.
[8]Ralph Lee Smith, "The Wired Nation", The Nation, May 18, 1970, 583.
[9]Radical Software, no. 2, p. 2.

credit), real-estate men can get instant readings on available properties across the country, cattle farmers can get instance price and production data several times a day through a computer based information service, and state and local governments have developed prototype urban information systems. The development of a domestic satellite system along the lines of Intelstat (the present global system) will permit an even greater integration of national and international information networks. Banks may turn out to be the central institution in the gradual integration of corporate information systems as their own systems of credit card payment, counseling services, payroll compilations (for corporate depositors), economic research and trust fund supervision become increasingly interwoven.

The information technologies have also had a decisive impact on the technological structure of the industrial sector of the economy. The simplest indicator of this impact is the steady shift to white collar employment in manufacturing enterprises. Thus, from 1947 to 1969 the proportion of white collar in the manufacturing labor force rose from 16.4% to 26.8% or 60%.[12] In effect, the proportion engaged in information "production" rises while the proportion engaged in commodity production falls. This reflects not only an increase in information processing activities but also the impact of cybernetic (i.e., information-feedback-control) modes of physical production. In fact, from 1955-1969 nine basic manufacturing industries

---

[12] The Manpower Report of the President, 1970, Table C-5.

experienced absolute declines in blue collar employment though sizeable increases in output.[13]

The following table (a daring attempt at quanitification by Richard Meier)[14] captures the notion of a developing information infrastructure. (The vertical vectors superimposed on the graph represent rates of growth.) As can be seen the greatest number and fastest growing of "transactions per interval" lies in the region of telecommunications. As the graph indicates, the growth in telecommunications promise a decrease in the average cost of a socio-economic transaction.

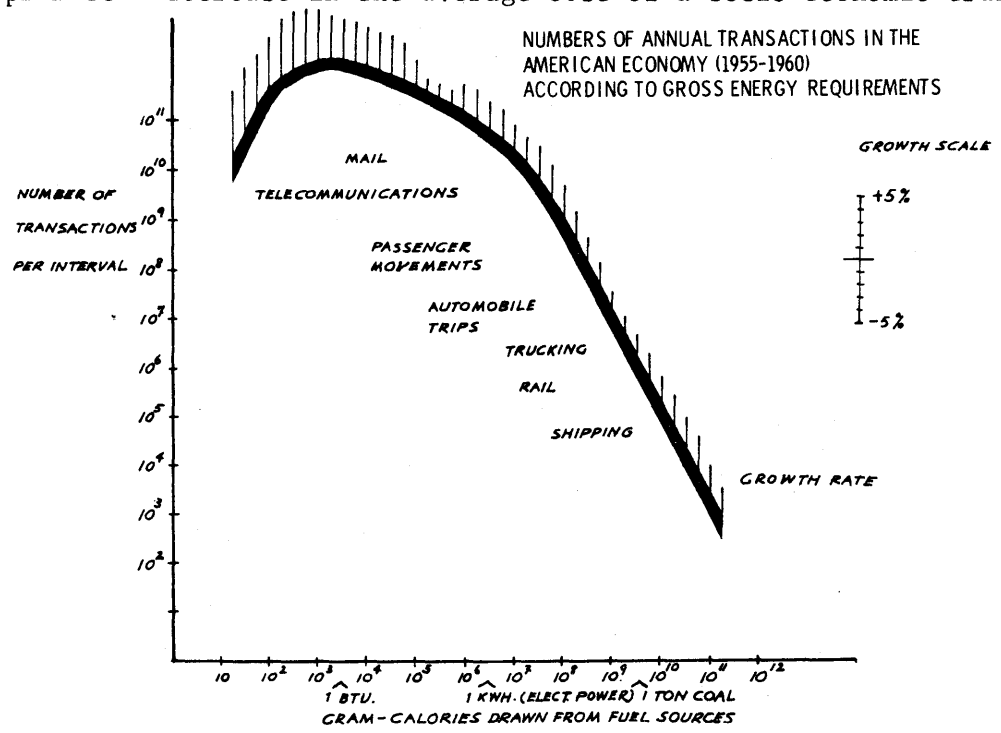


Figure 7

These data require a large number of rules and conventions regarding the proper allocation of energy costs before a distribution like the above can be plotted. However the substitution of telecommunications transactions for energy-intensive transactions at the growing margin does become apparent, and it is evident that henceforth social and economic activity should expand at a more rapid rate than energy consumption.

---

[13] Julius Rezler, Automation and Industrial Labor, Random House, New York, 1969, p. 35.

[14] Richard L. Meier, A Communications Theory of Urban Growth, M.I.T. Press, Cambridge, 1962, p. 61.

The Software

There can be little doubt that the hardware of the communications revolution, the "electronic highway", is fast developing. But as the amount of information available for decision making increases there develops a great incentive to use such information efficiently and effectively. An "information explosion" can be counterproductive unless accompanied by an increasing ability to structure and pattern information. Decision making can no longer be left to informal processes, to the "intutition" of governmental, corporate, and institutional managers. Rather it must be systematized and integrated with information flows. Software, a decision making technology in the form of programs, models, knowledge, and education becomes the central and organizing factor in an organization's and society's "information capital". In effect, the increasing flow of information restructures the decision making process. Decision making is transformed into a technology. It becomes a consciously designed human activity amplified by man-made instruments. Problem solving, once the domain of intuition and the skill that comes with years becomes part of a rationalized man-machine system. Problem solving and decision making skills are no longer allowed to evolve as the result of numerous and unintended trial and error experiments, but rather are consciously designed, machine amplified, and modified as a result of consciously articulated organizational needs.

A crude indication of a soceity's software is its educational level. Thus from 1950 to 1964 the proportion of college age youth in college rose from 25.5 to 43.3 percent. Within the same period the number of Ph.D. degrees conferred increased by more than 100%.[15] By 1970 the proportion of such youth in college far exceeded the corresponding proportion for the Western European countries.[16] Thus, there is clear if crude evidence that society's requisite software, the human component of man-machine decision systems, has developed apace of the communications hardware.

But our data suggest an even more tantalizing hypothesis. As the proportion of labor power engaged in information transmission and decision making increases, and as decision making itself becomes a technology, decision making replaces production as the "central" socio-economic activity. Decision making and information processing, rather than physical production, set the tone and define the modalities of the work experience. Decision making rather than physical production becomes the area for the application of human intelligence. Most important, decision making becomes the new arena of political conflict. The distribution of information rather than the distribution of commodities becomes the "leading" socio-economic conflict. Access to decision making prerogatives rather than access to physical capital become the leading and critical demand.

---

[15]Daniel Bell, "The Measurement of Knowledge and Technology", in E. B. Sheldon, W.E. Moore, Indicators of Social Change, Russell Sage Foundation, New York, 1968, pp. 204-206.

[16]Zbigniew Brzezinski, Between Two Ages: America's Role in the Technetronic Era, Viking Press, New York, 1970, p. 27.

In effect, the increases in communications software and hardware, in the information infrastructure and information capital, mean that society's channel capacity, its ability to structure and pattern information at a given rate increases. Like physical capital that amplifies labor power's physical capabilities, information capital amplifies labor power's information processing and decision making capabilities. Decision making is thus transformed into a technology, a consciously organized and machine amplified activity, and political conflict centers around the distribution of information and decision making prerogatives.

A Note for Guidance

Each chapter in this thesis explores the concept of information capital from a different perspective. We try to show how information capital emerges out of natural market processes, and in the process we build toward an analytic or framework for discussing the consequences of the "information revolution". Note that our interest here is not in expanding the literature of information systems *per se*, i.e., on the efficient design of problems solving, planning and information networks. Like the microeconomist who assumes that the firm applies its engineering knowledge efficiently, and then explores the more macro-consequences of this assumption, we assume "optimality" and then explore the more macro features of organizational and social structure.

The reader is advised to consult the appendix for the most minimal introduction to "Information Theory". Only the most minimal is required to read the mathematical formulations of Chapter Two and Three.

INFORMATION CAPITAL

from the Perspective of Market Equilibrium

Chapter One

THE MARKET AS AN INFORMATION SYSTEM

§1. It is by now a truism to note that the critical function of markets is to coordinate massive numbers of decentralized and independently arrived at decisions. Coordination is achieved through the combination of a price system and the profit motive that serves as a homeostatic device to insure over-all patterns of equilibrium and efficiency. But if we are to have a deeper understanding of the market as an information system, we have to ask: a) What kinds of information are required for market decisions? b) How does the cost of information affect patterns of coordination? c) What impact does better or "more perfect" information have on market coordination?

General equilibrium models probably constitute the most explicit statement of the price system as the market's "data base". Such models usually assume that each decision maker knows, and need only know the present and future price of every input and output. An auctioneer then takes all bids and offers, and through some recontracting device arrives at a set of present and future equilibrium prices. It is readily conceded that such models are massive abstractions but nevertheless demonstrate the "possibility" of equilibrium. But it is difficult to determine the relevance of such "equilibriums" if critical assumptions about information in the model cannot be justified.

Since all prices are known to all decision makers, it is implicitly assumed that information is costless. Price information is presumably printed in a kind of "stock market" page and distributed freely by the auctioneer. It is also assumed that future prices are in fact obtainable (an assumption made more tenuous if we let the commodity structure change over time), and that the firm is interested only in prices.

Information is of course not costless. Stigler[1] shows that where search is costly, there will be a distribution of prices for some given homogeneous commodity. In the interest of economizing on search costs producers and consumers can persist in inefficient patterns of production and sale, by relying on modes of goodwill and reputation. More importantly, the firm cannot afford to know "everything", every price. Comprehensive knowledge comes at a price, and as a consequence the decision unit must restrict its range of consideration.

Prices do not exhaust the category of information necessary for decision making. Inquiries into consumer tastes, technological and scientific development, political decisions, legal restrictions, and a whole host of data we associate with the "sensing" of opportunity are essentially non price data. Markets do not "provide" this data in any meaningful sense.[2]

---

[1]George Stigler, "The Economics of Information," J. Political Economy, June 1961, pp. 213-225.

[2]Urban agglomerations are usually credited with providing such data at low prices, particularly when the decision maker is not looking for a specific piece of data, but simply wants to stay "informed". See William Alonso, "Industrial Location and Economic Development",, working paper no. 13, Center for Planning and Development Research, University of California at Berkeley.

Finally, future prices are not known, particularly if patterns of consumption change over time. No "arbitrage" can exist between presently known products and future unknown products. Future markets will be functional only where the commodity is homogeneous and storage costs are low. (Money markets are the best example.) To be sure, capital markets represent generalized trades between the present and future, but no claim can be made that the structure of interest rates represents "real" trade-offs. An interest rate simply reflects both a lender's and borrower's guess about the future. Because such guesses are shot through with uncertainty, the "established"investor with a past record of success can purchase funds at a lower price.

General equilibrium model builders will readily admit to these objections, but will insist that their "equilibrium conditions" are of heuristic value. Perhaps. But the most curious consequence of these models is the modification if not the subversion of the original "invisible hand" concept. What puzzled Smith in 1776 and Hayek in 1937[3] was the fact that markets could coordinate independently arrived at decisions despite the fact that no one had the "full picture", the comprehensive view" of market patterns. Invisible hand meant just that, coordination took place "behind the backs" of the actors. Contrary to the general equilibrium assumptions no one knew all the prices. How then can fragmented views result in coherent and comprehensive patterns?

---

[3] Friedrich Hayek, "Economics and Knowledge," Economica, February 1937.

There is no mystery in the question or answer, but G.B. Richardson[4] provides a precise framework for considering the problem. Thus imagine that we listed the marginal revenue product and wage rate of factor J in all its N uses in "chain" fashion as below. (Profit maximization in each use insures $W_i = MRP_i$)

$$(W_1 = MRP_1) \rightarrow (W_2 = MRP_2) \rightarrow (W_3 = MRP_3) \rightarrow \cdots \rightarrow (W_n = MRP_n)$$

If each individual actor were "informed" about the wage rate in his use and his use only, then there is no presumption that these arrows will become equal signs. But as soon as there is a minimal "overlap" of information, so that W_i is known by two and only two actors, and each actor knows only two wage rates, (n - 1) actors will insure full and "optimal" equilibrium (i.e., the MRP of factor J is the same in all its uses). Thus with minmal overlap, the "division of knowledge" can lead to comprehensive and coherent patterns. Smith and company were right to marvel at the market as an instrument for coordination. Where information processing capabilities were limited, and information search was costly, fragmented information could "add up" to coherent over-all patterns.

This adding up process -- the "arrival" at equilibrium states is of course always tendential. Where the overlaps in knowledge are small, i.e., the information processing capabilities of the individual units are limited, and the transmission of the information down the chain is

[4]Gerald B. Richardson: "Imperfect Knowledge and Economic Efficiency," Oxford Economic Papers, June 1953, pp. 136-156; "Equilibrium, Expectations, and Information," Economic Journal, June 1959, pp. 223-237.

slow, long periods of disequilibrium can persist. "In the long run", initial conditions can themselves change, so that the "tendency" toward equilibrium remains just that, only a tendency.

There is, however, no one single equilibrium position. Equilibrium is always a "subjective" state. It exists when all plans are fullfilled and there is no incentive to change behavior. Thus imagine that a "subset" from the middle of the chain was broken off at both ends so that information flows only within the three emerging subsets but not between them. Then the system will be in equilibrium when each subset is in equilibrium, even if this implied different MRP's in different uses for the same factor. So long as there are no information flows to inform the actors of these differences, there will be no tendency toward a new more "efficient" position. This is simply the extreme case of "immobilities". In general, the greater the degree of segmentation, the greater the likelihood that the overall system will reach an equilibrium state.

We can show this more rigorously, and simultaneously deepen our understanding of the relationship between complexity, information processing capabilities, and equilibrium, by using a matrix rather than a chain to visualize all forms (price and non-price) of information.

Thus let us perform the following "thought" experiment. Let an (N $\times$ N) matrix list the names of all decision makers down the rows and across the columns. Give each decision maker "his" row vector and let him check a column, if and only if, during some specified period

of economic activity he is "informed" of the general state and behavior of the decision maker represented by the column. (Presumably, information about the state and behavior of another decision maker can include price information and/or considerations of oligopolistic interactions.) The resulting pattern of checks throughout the matrix represents the flow of information between the decision makers. No information flows whatsoever, would result in checks along the main diagnol only, while total information flow would be shown by a completely "filled" in matrix. More realistically, we would expect a pattern of "submatrices" to emerge, block diagnol in form, representing independent subsystems within the economy. This segmentation is then the analogue of our broken chain.

What can we say about the tendency toward equilibrium in such a system? To repeat, equilibrium is a subjective state. So long as plans are fulfilled and no new information intrudes, a system is in equilibrium regardless of "real" marginal equilivancies. More importantly, price need not be the only or even major form of "information" signal. Price is both a resource cost and a factor income. Because a price is someone's income there are always institutional pressures to stabilize prices and use other kinds of information signals as homeostatic devices. (To function as a homeostat, a variable must change.[5])

[5]Romney Robinson develops a model of "disequilibrium price", in which the price is set at the revenue maximizing point for the industry, but above individual firms' marginal cost. "Share of the Market" then functions as the industry's homeostatic device. Since industry price is greater than firm marginal cost, each firm has an incentive to expand production. Competitive equilibrium ensues when as a result of these efforts market shares stabilize. A falling market share is the "signal" for the firm to change its behavior. (Note that Robinson's model allows for meaningful competitive equilibrium whereas the general equilibrium models do not. In the latter the firm expands production to the point where price equals marginal cost. At that point it has no incentive to increase production by "stealing" its competitors' customers. The neo-classical equilibrium is strictly non-competitive.)

Now consider an independent subsystem of our matrix. Presumably, it will arrive at its own internal equilibrium by some process of "iteration" (analogous to a linear programming solution). We can think of two extreme forms of this "iterative" process, one resulting in the shortest time to reach equilibrium, and the other in the longest.[6] Let N be the number of decision making units, let T be the time unit under consideration (e.g. days), and let P be the probility that any unit will "hit" upon its correct equilibrium position. ("Correct" in that it will be consistent with over-all subsystem equilibrium. Assume that there is only one set of equilibrium values.) Then we postulate that equilibrium will be reached most quickly when each unit in succession arrives at its correct position. Since it will take each unit about T/P to arrive at equilibrium, total subsystem time will be TN/P. At the other extreme, we postulate that the system will reach equilibrium only when all the units simultaneously arrive at their correct position. (This is clearly the most extreme since there is no such thing as being "almost" there; (N- 1) units may be at the right point, but if the N'th unit is a laggard it's still "no go".) In this case it will take on the average T/P^N for the subsystem to arrive at equilibrium. Thus we can say that the actual equilibrium E^t will be given such that

$$TN/P \leq E^t \leq T/P^N$$

---

[6]The following discussion is an adaptation and extension from, W. Ross Ashby, Design for a Brain, Chapman and Hall, Ltd., London, 1960, p. 151.

Let us assume now that

$$E^t = \tfrac{1}{2}(TN/P + T/P^N)$$

Then if we have J independent subsystems the entire system will reach equilibrium in an amount of time given by

$$\max(E_i^t) \ , \qquad i = 1,\ldots,J$$

because by assumption the subsystems proceed to equilibrium independently of each other. Now it is clear that

$$dE_i^t/dN_i > 0$$

Thus for a fixed number of decision makers, as the number of independent subsystems fall $\max(E_i^t)$ will rise, or at least not fall. If we let S_J be the number of independent subsystems, we have in summary,

$$d \max(E_i^t)/dS_J \leq 0$$

Intuitively, as the degree of "complexity" in the entire system increases (i.e., as "interdependence" increases with the fall in the number of independent subsystems), the time required to reach equilibrium also increases. The system is characterized by longer and longer periods of "disequilibrium" as its degree of complexity increases. This assumes, however, that $dP/dS_J = 0$. If we let $dP/dS_J < 0$ then the sign of $d \max(E_i^t)/dS_J$ is ambiguous. What are the determinants of P?

P is the probability that the decision unit will be able to hit upon its correct equilibrium position. This will be a function of its information processing capabilities, i.e., its ability to observe

and process information concerning other decision making units, and its "modelling" capabilities, i.e., its ability to "structure" its information so as to anticipate possible equilibrium positions. Now a decrease in S_J implies an improvement in the former capability, since subsystems can be joined only when information flows across subsystem "borders" (i.e., subsystem A will join subsystem B, only when A can absorb information about the actions of units in B). But a decrease in S_J need not imply an increase in the latter capability. A unit can observe and process new information without being able to model it.[7]

Thus the system can handle its own complexity if the ability of the firm to structure its increased information flows improves along with the increases in system complexity. We say that this ability is a function of the amount of the firm's "information capital". Thus "information capital" emerges as the sine qua non for equilibrium patterns as system complexity increases. In other words, an increase in system complexity requires an increase in information capital if markets are to continue to perform their equilibrating and coordinating function.

---

[7]To give an analogy, if we let C be the former capability and M the latter, then the international arms race is high on C but low on M, i.e., the system tends toward increasing disequilibrium as the number of arms increases but the amount of defense falls. For a more detailed discussion see, Kenneth Boulding, Conflict and Defense, Harper and Bros., New York, 1962, p. 151. We can say more quixotically that improvements in C unaccompanied by improvements in M, mean that the increasing flows of information are simply disorienting. The decision unit can observe his inter dependence with other units but is unable to act upon this information, to adapt to the newly perceived interdependence, a process analogous to Toffler's description in Future Shock. See, Alvin Toffler, Future Shock, Random House, New York, 1970.

Within this framework the "entrepreneur" emerges as the "knowledge gap filler".[8] Equilibrium of the entire system is based on a kind of pact of "mutual ignorance". Subsystems "agree" to ignore each other, and in this way limit the variability of their environments (solving the problem of their limited information processing capabilities). But the entrepreneur upsets the applecart. He has the more "comprehensive vision" and takes advantage of "differences" he perceives over the entire system. Subsystems are thus joined, but in the process old patterns of information flow and decision making are destroyed. The destructive process is critical. Members of the old subsystem face new environments. Old patterns of decision making will no longer be appropriate, and unless the firm can adapt to its new "field" it could very well go under.[9] This destructive process is itself an incentive for the firm to "build up" its information capital stock. To adapt to larger subsystems and to insure its future adaptability, it will want to increase its information processing and modelling capabilities.

Note that as in Knight our entrepreneur is rewarded for strict uncertainty, since he responds to information differences outside the normal patterns of information flow.

To briefly summarize, information processing and modelling capabilities determine both the complexity and equilibrium tendencies of

---

[8]This is a concept developed by Leibenstein. See Harvey Leibenstein, "Entrepreneurship and Development," American Economic Review, Papers and Proceedings, May 1968, p. 75.

[9]This is analogous to Schumpeter's process of "creative destruction". See, Joseph A. Schumpeter, Capitalism, Socialism and Democracy, Harper and Row, New York, 1962, chapter 8.

market systems. We say that the general equilibrium model evades some of the critical problems associated with the nature of these capabilities by assuming information to be costless. All the relevant data was available to everyone. The Richardson framework solves these problems by showing how fragmented information -- the division of knowledge -- could result in over-all coherence. Nevertheless, such coherence is only a theoretical tendency, and equilibrium was more likely to result as a consequence of market segmentation. We then showed how increasing complexity, cetirus paribus, led to longer periods of increasing disequilibrium, instability, unless accompanied by concomitants increases in information capital. We noted that information capital was likely to emerge if firms sought to insure their future adaptation. Thus we can argue that information capital is both a result of and a necessary condition for market complexity.

§2. One of the more accepted dictums of economic theory is that perfect information leads to perfect markets. This dictum is no doubt responsible for the scant attention paid by economists to the recent and rapid changes in communications and information technologies. The growth in society's "data base" is seen simply as an extension of a process begun with the industrial revolution -- the growth and perfection of markets as social mechanisms for coordination, and arbiters of social change. But before this proposition can be accepted one has to rigorously define "perfect information".

There are in general three dimensions to the "information set" of a decision making unit. There is the range of variables, the precision with which they can be measured, and the perceived degree of correlation between them. An "improvement" along any one of these dimensions is a movement toward better or increasingly perfect information.

The economist's dictum is perhaps clearest if we consider only the dimension of precision. Thus imagine a stable equilibrium of independent subsystems described in the last section. As the homeostatic variables are known with greater precision, the iterative process improves, periods of disequilibrium are shortened, and subsystem markets become "more perfect". But as we saw improvements along the other two dimensions, were a consequence of developing system complexity and emerging information capital.

It is the central contention of this thesis that the emergence of information capital has social and economic consequences that far

transcend the increasing perfection of markets. In the coming chapters we will argue that information capital, as the locus of society's new decision making technology, destroys old decision making patterns, old organization structures, and unleashes new conflicts surrounding decision making prerogatives. The "distribution of information" becomes a dimension of social conflict. We will argue that ultimately the growth in information capital modifies the role of markets as mechanisms for social coordination, as consciously designed information systems increasingly guide social and economic behavior.

INFORMATION CAPITAL

from the Perspective of Organizational Growth

Chapter Two

CHANNEL CAPACITY, ORGANIZATIONAL GROWTH AND THE EMERGENCE OF INFORMATION CAPITAL

Introduction

Economic theorists have long wondered why there should be long run diminishing returns to firm output.[1] A rising average cost curve is explicable as a short run phenomena. Some factor, most likely the capital stock, will be "fixed". But fixed factors would seem a logical impossibility in the "long run". In the "long run" as output grows everywhere, there are no resource constraints on any given input, so that constant returns to scale should prevail at the point of minimum average cost. What then is the long run "scarce" factor? The answer is most often the entrepreneurial or managerical input itself. Coordination costs are subject to the law of diminishing returns. Managing too large and complex an enterprise imposes costs above and beyond the cost of production. The answer is plausible enough but is never rigorously justified. What are the determinants of managerial "coordination capacity"? If they are technological in origin can't their implicit costs be reduced through "technological progress"? What would be the substance of this "progress"?

The central theme of this chapter is that "coordination" capacity is a function of information processing capabilities, and that long

---

[1] For an extended discussion of this problem, see Kenneth Arrow, "The Firm in General Equilibrium", Technical Report #3, May 1969, Harvard University (unpublished). Also see, Chandler H. Stevens, Information and the Division of Labor, Doctoral Dissertation, M.I.T., Jan., 1967, chapter 4.

term improvement in such capabilities gives rise to what we have loosely called "information capital". We will show that information capital emerges out of the limits imposed by the two major forms of organizational growth, growth through scale and growth through the structural differentiation. However, as we shall touch on lightly here, and in more depth in Chapter Five, the emergence of information capital is a destructive process, that sets the stage for further conflicts over decision making power.

§1. The Organization as an Information System

Any organization can be viewed as a sum of its component parts, each part performing a distinct and unique function. The functional coherence of any given part makes it a relatively independent subsystem within the organization. As a consequence, communication within the parts, the subsystems, is greater than communication between them.

These relationships can be shown with the matrix used in our "thought experiment" of Chapter One.[2] Thus, if we let A and B be the units of subsystem I, and C and D be the units of subsystem II, and if we let the entries in the matrix be the indices of communication between units we would have

	A	B	C	D
A		60	2	2
B	60		2	2
C	2	2		60
D	2	2	60	

Figure 1

This matrix of communication is "nearly decomposable". If we let the communication indices between the subsystems go to zero, our matrix would become block diagnol in form composed of two completely independent subsystems.

---

[2]For a more complete exposition of this approach see, Herbert Simon, The Science of the Artificial, M.I.T. Press, Cambridge, p. 102.

However, the smaller the degree of communication between the subsystems, the more difficult will be the task of organizational coordination. The organization must in the end function as a coherent whole. If communication between the subsystems is minimized, the organization could degenerate into its component parts each pursuing its own goals. Subsystem actions would become inconsistent with each other and the organizational "rate of error" would increase. Thus, if a production unit did not communicate with the sales force too few or too many goods would be produced.

To counter a rising rate of error as subsystem communication falls, the subsystems must "buffer" themselves by various forms of resource "stockpiling". Thus, as in our previous example, the production unit will seek some optimum level of inventories to protect itself against variability in sales orders. In general, "buffering" will result in some form of "excess capacity". A production and design unit of an electronics firm may each have two draftsmen, though at any given moment none are used to "full capacity". Greater communication between the units could cut the number of draftsmen required in half, but because of low "information exchange" each unit retains two for unexpected periods of "peak activity".

Now buffer costs could be reduced if subsystem communication were increased, but communication also costs. Thus any organization confronts a "trade-off" between buffer and communication costs, the former increasing as subsystem communication falls, and the latter increasing as subsystem communication rises.

Emery summarizes this trade off in the following diagram.[3]

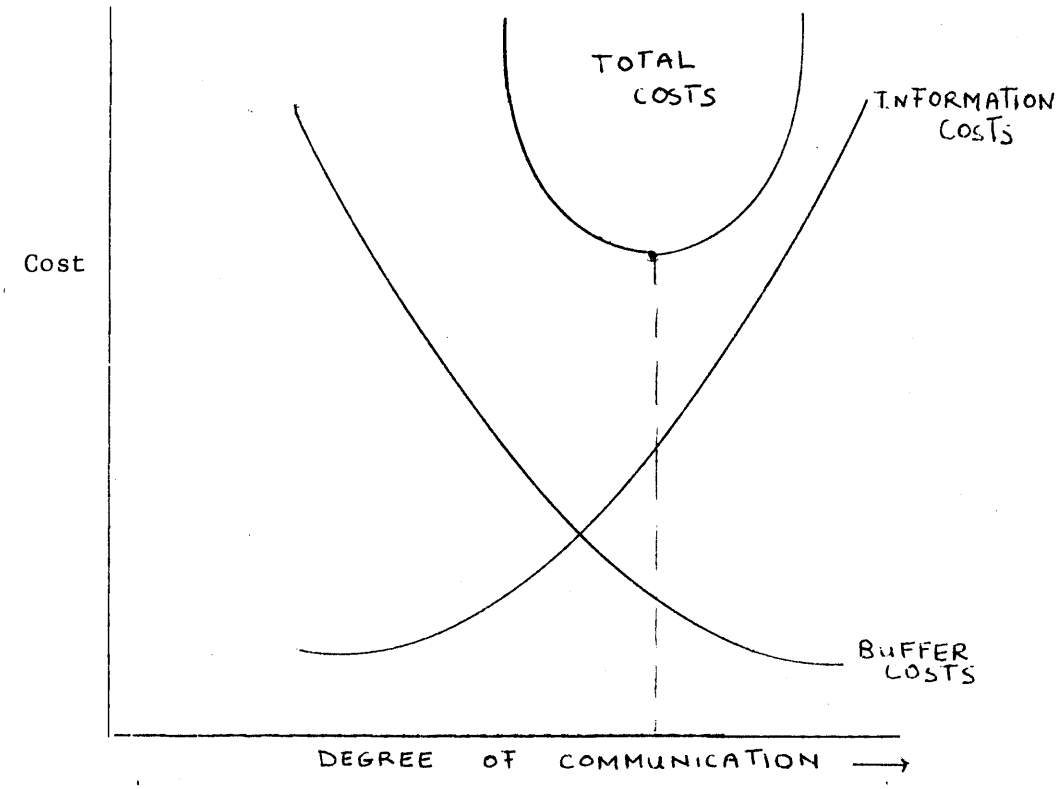


Figure 2

The cost curves in the above diagram will be determined by the information processing or coordination technology of the firm. Thus with the technology fixed, the firm will derive its optimum level of subsystem communication and thus subsystem "independence".

---

[3]James C. Emery, Organizational Planning and Control -- Theory and Technology, Doctoral Dissertation, M.I.T., Jan. 1967, p. 39.

The Emergence of Hierarchical Organization

Since communication between subsystems is costly, there is an incentive to transmit only the most "necessary" and the most "relevant" information between them. Thus the production unit of our previous examples, will not inform the sales unit of the daily maintenance decisions it must make, but will notify it in the event of a serious breakdown in machinery. Similarly the sales unit will notify the production unit of only major changes in sales patterns. Thus the "channel" will transmit selective information of a general nature. Hierarchic information flows are structured to do just that.

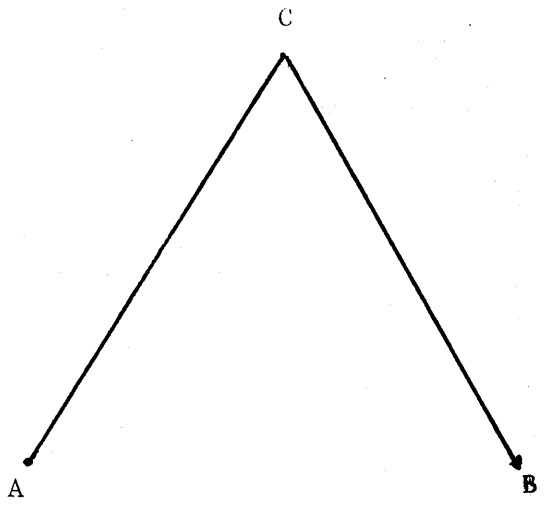


Figure 3

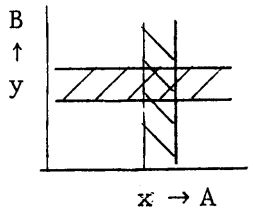
In Fig. 3, A-C-B, represents the "between system" information channel. C represents the point at which information from A to B is "matched" or "coordinated". Thus information that flows within A and B is more detailed but less general than information that flows to C.

Note that in the abstract, hierarchical information flows do not necessarily imply an inequitable distribution of "power". The information processors and decision makers at C can be the "elected representatives" of the members of A and B. But there is a clear tendency for hierarchic flows to generate or magnify such an inequitable distribution. This flows from the more general, more comprehensive nature of information available to decision makers at C.[4]

Now we define the "span" of an organizational unit as the number of subordinates with which it directly communicates. Thus, in Fig. 2, the span of unit C is two. If we assume that the span in unchanging throughout the organization, we can deduce some of the simple arithmetic properties of "hierarchic" or "tree structure" models of organization.[5]

---

[4]We can make this notion more explicit in the following way. Imagine that information from A and B is required for organizational decision. Information from A can be represented as some region along the X axis of a graph while information from B can be represented as a region along the Y axis. (We assume that the information comes in the form of a region rather than a point, since all information is subject to some margin of error.) Information to C then consists of the intersection of these two regions. This is shown in the following diagram.



Thus C has a more "exact" notion of the organization's "position" than does either A or B.

[5]The arithmetic results are drawn from Emery. See, James C. Emery, op. cit., pp. 52-53.

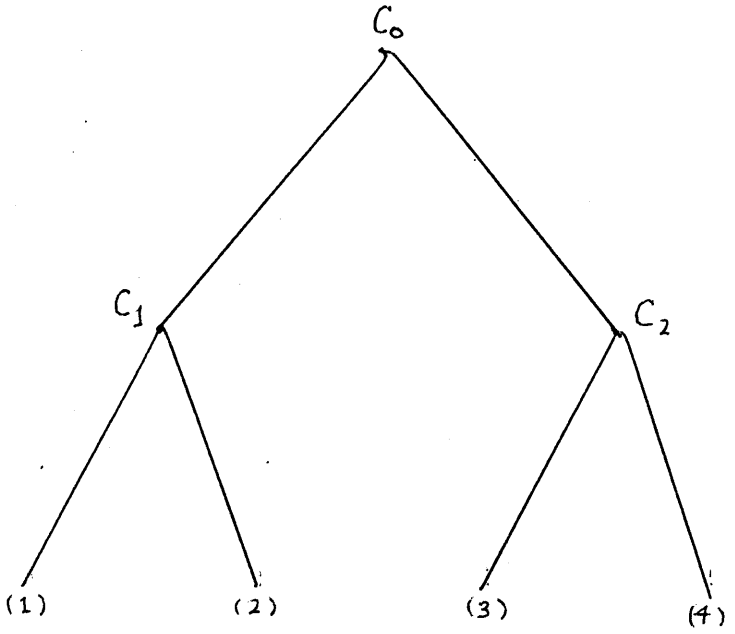


Figure 4

Figure 4 represents a sales organization, where the bottom four nodes represent the "primary" sales activities (e.g., sales in four different regions) and the remaining nodes represent coordinating and decision making units. The span of the organization is two.

If we let L be the number of levels in the organization, S its span, N the number of primary activities, and M the number of decision making nodes, then it can be shown that

1) $S^L = N$ or $L = \log_S N$

2) $M = N - 1/S - 1$

therefore, if N is assumed constant

1a) $dL/dS < 0$

2a) $dM/dS < 0$

Now we can heuristicall "prove" the following theorem

"For a fixed N as S falls, the 'amount' of information exchange between subsystems falls."

To prove this we need only recall that as information moves from one level to the next, it is coarsened. Detail is supressed as information moves up the tree. Since we know that $dL/dS < 0$, as span falls a piece of information from a given unit must move through a greater number of levels before it can be "matched" with a piece of information from any other unit. In turn we would expect, that as it moves through a greater number of levels it becomes progressively more coarsened. But presumably, more coarsened information is less information. Thus, in effect, less information will be exchanged between the two units at the "point of matching", i.e., at some node in the hierarchy.

We can show this more explicitly by using information functions to depict the amount of information flow over the tree.

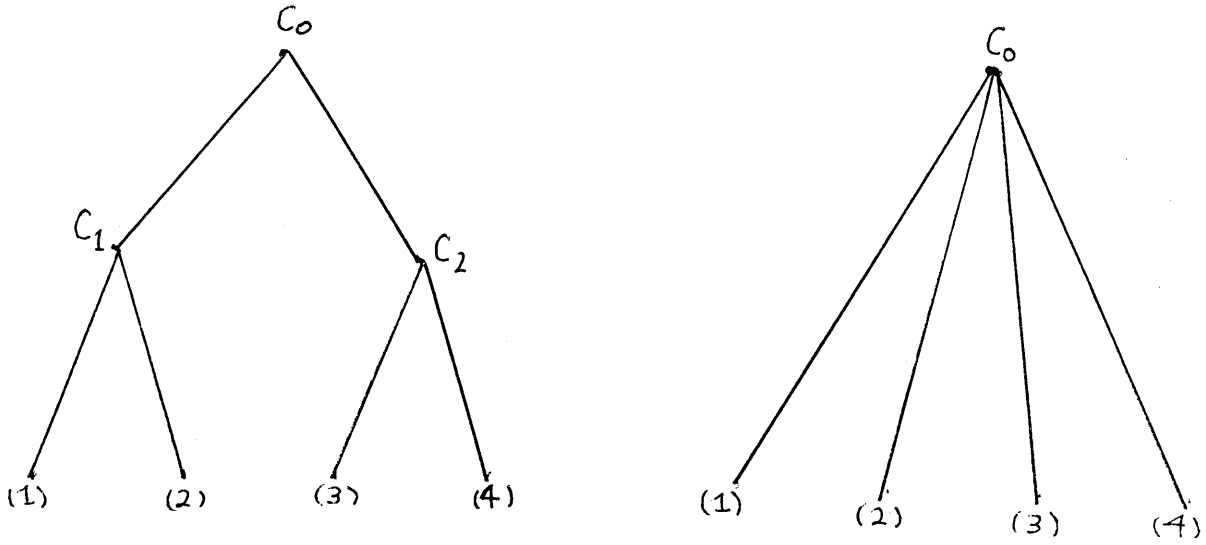


Figure 5

Consider organization A in Fig. 5. Imagine that the activity of each primary node can be described as being in either one of two states. (We will give a more specific example shortly.) Then the total set of possible activities will be of the size $2^4 = 16$. Now in every period let each primary node signal its activity state to its immediate "superior", who in turn transmits that information to the central coordinator. We assume that total "transmission" time up the hierarchy takes only one period and every node "fires" only once per period. We can see then that the central coordinator will be able to receive information from a set of size $2^2 = 4$ possible states, even though the primary activities are in a set of 16 such states. Thus the central coordinator receives in effect less information as a result of information coarsening.

Another way of viewing this is to describe the communication mechanism between primary nodes on and two (the left-hand nodes), and primary nodes three and four. Clearly they can communicate only through the central coordinator. Now nodes one and two can be in one of four states, but the central coordinator perceives them as being in only one of two states. Thus any information he gives to three and four about the states of one and two will be coarsened. As a result of the vertical movement of information up and down the hierarchy information is lost. Nodes three and four have only an incomplete picture of the activities of nodes one and two. However, the more centralized organiztion B of Fig. 5 has no intervening nodes between the primary activities and the central coordinator. The coordinator receives full information from a set of 16 possible activity states.

We can make this more rigorous by defining one of the possible activities of the primary nodes (that is, each node engages in many information transfers per period, one of which we will describe now).

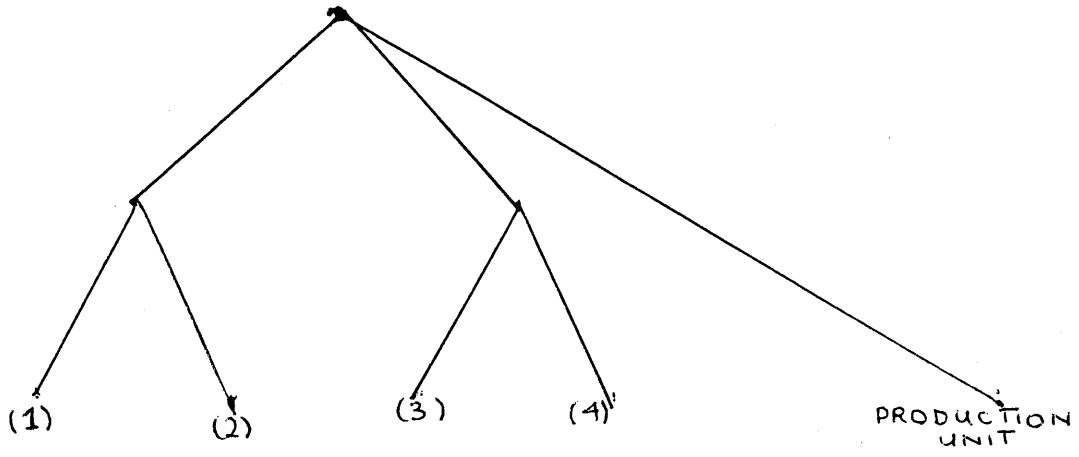


Figure 6

Let the organization in Fig. 6 represent an "expanded" organization where the left-hand side is again our sales organization, and the right the firm's production unit.

We imagine that in every period, the sales organization scans a market area divided into four region, to determine which regions within the area are about to order. The results of this scanning are reported to the central coordinator who in turn unstructs the production unit to produce a certain mix of goods for delivery. We assume that the products to be delivered to each region are tailored to the needs of each region so that for any order the production unit must order a specific mix of raw materials. In effect, each primary node specializes in scanning a given region of the market. Now assume that in any period any combination of regions can order,

but that there can be only one order per region. For simplicity, we let

$$p_i = .5, \qquad i = 1,\ldots,4$$

be the probability that any given region will in fact order in some specified period.

Under these assumptions there will be 16 equi-probable sets of orders. Thus we can say that in each period the organization must process $\log_2(16) = 4$ bits of information to make the correct production decisions. (Or alternatively, since $1/16 = .0625$, then $H(.0625_1,\ldots.0625_{16}) = 4$ where $H = -\sum p_i \log p_i$)

How will the information be coarsened as it moves up the tree? Intutively we can see that coordinators C_1 and C_2 must each transform a set of four possible states into a set of two possible states as he transmits information from the field to the central coordinator. We can represent one such possible transformation with the following matrix. Let I represent will order and O represent will not order

	I	O
IO	1	
OI	1	
OO		1
II	1	

Figure 7

The left-hand column represents the four possible combination of orders from the field transmitted to coordinator C_1 (or C_2). The

entries in the matrix represent the probability that a given set of orders from the field will be transformed into a given set of orders to the central coordinator. Thus order (I,O) from the field will be translated as (I) to the central coordinator with a probability of one (where I represents the statement "orders from both regions"). Note that in the translation process, coordinators C_1 and C_2 are not simply "parrots". They get information from the field which they must interpret to determine which set of orders actually "holds". In turn their "recommendations" along with the requisite information must be interpreted by the central coordinator.

Now in this transformation three out of four times C_1 and C_2 will report orders from both regions. Thus the amount of information transmitted from C_1 and C_2 will be $2(-.75 \log .75 - .25 \log .25) = 1.6$ bits. Thus four bits of information are transformed in 1.6 bits.[6]

But as we saw before the organization represented in Fig. 5B, the organization with the greater span will not coarsen its information flow. At the central coordinator node where information from the entire field meets, four bits or 100% of the information flow can be exchanged, while only 1.6 bits or 40% can be exchanged in organization 5A.[7]

Now as Fig. 5 suggests a greater span for a fixed number of primary activities implies a greater degree of centralization of organizational

---

[6]Note that the particular transformation matrix chosen in arbitrary. But any matrix chosen must coarsen the information flow since a set of two possibilities (I,O) cannot carry as much information as a set of four possibilities.

[7]For a similar approach to coarsening that does not use information functions see, Jacob Marschak, "Information Economics Reconsidered", working paper #149, Western Management Science Institute, U.C.L.A., June 1969, §8

activities. Thus we can say that the greater the communication between the subsystems, i.e., the greater the degree of information exchange between them, the greater the degree of centralization within the organization. Returning then to Fig. 2, depicted here again in Fig. 8, we can now write on our X axis, "the degree of centralization" rather than the "degree of communication".[8]

So now we see that with a fixed information processing technology, the trade off between buffer and communication costs will result in some degree of organizational centralization, with a consequent organizational "average span".

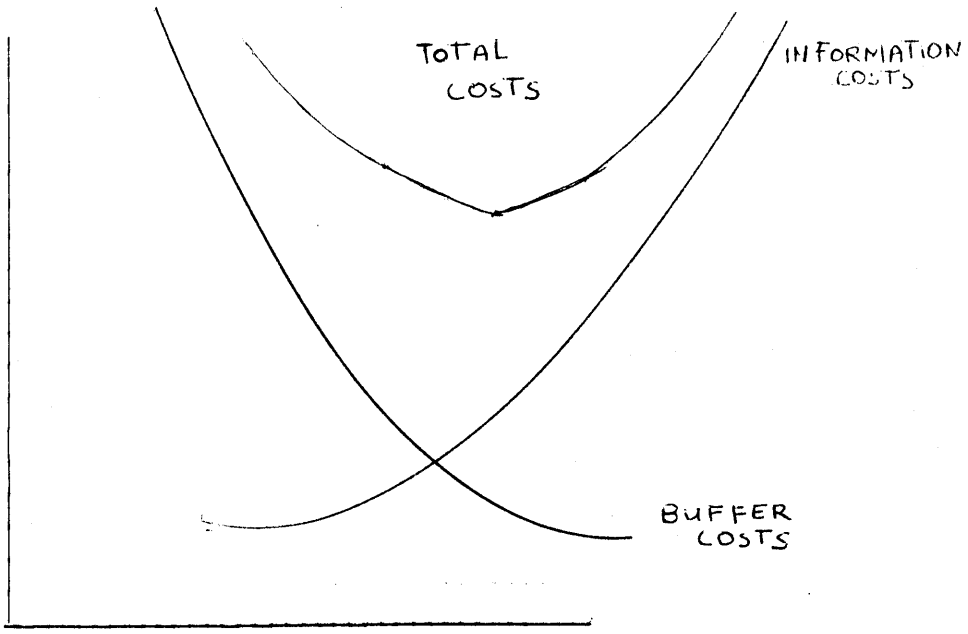


Figure 8

[8]This implicitly assumes that all communication is of the hierarchical form. For a more detailed discussion of the assumptions behind this graph see the appendix at the end of this chapter.

§2. Channel Capacity as Minimum Average Cost

We can now return to the question posed in the Introduction. In that way do information -- coordination -- costs, lead to a rising cost of production? To answer this question we must develop more rigorously, the relationship between the "rate of error", buffering, and information processing capabilities.

Intuitively, the concept of "channel capacity" suggests a limit to the amount of information that can be processed by a decision unit. If we think in information theoretic terms, this suggests that there is a limit to the "precision" of choice that any decision unit can attain. Thus we know that the process of "decoding" entails the selection of a single message from an ensemble of messages, where each message occurs with a given probability. The more nearly equal in probability are the messages the greater the information value of any selection or decision. Thus, interpreted broadly, the more difficult it is to detect differences between possible choices or decision, the more information there is contained in the choice, the more information is required to make the choice. But channel capacity is a "rate concept", it is the flow of information per unit of time. A difference may be difficult to detect, but if a great deal of time is spent in searching for the difference it will be found. Thus, the less distinct are choice possibilities, and/or the faster the rate at which such choices are made, the greater the rate of information flow, Channel capacity is thus a limit to the rate at which such choices can be made without error.

[9]Note our emphasis on the concept of choices. It is tempting to think of channel capacity as analogous to the maximal rate of water down a pipe if the pipe is not to burst. But this is more analogous to "effort capacity", i.e., how rapidly can a man perform a repetitive task that entails no choice but simply exertion.

Now we assume that a given organizational structure, the result of a balancing of information and buffer costs results in, or generates an organizational channel capacity with respect to the decision to produce a certain bundle of goods when orders are obtained from the "field". Thus let us look at Fig. 5A again. We can see that if orders came in at the rate of one every two periods, and if each could fire once per period, then there need not be any coarsening. Coordinators C_1 and C_2 could send the full four bits of information by firing twice for every given set of orders. Thus the channel capacity of the organization would be four bits every two periods, or two bits per period. At this rate there would be no error in the production decision. The firm would always produce the correct set of goods.[10]

More generally if we let A be the channel capacity of the organization and H be the bit value of the information field (in our example of four bits), and if we assume again that $P_i = .5$, $i = 1,\ldots,N$, where N is the number of regions (or primary activities) then,

$$R = A/H$$

---

[10]This does not imply that if orders came in at once every two periods, there would be no need for any kind of organizational buffering. Information transmitted for the production decision is only one of the many kinds of information that must be transmitted in the course of organizational operations, e.g., information concerning the internal allocation of the firm's resources, the purchase of new materials, etc. Rigorously speaking, even the production decision should be subject to the information buffer trade-off. That is, irrespective of the rate of output the firm could trade some information flow from t the sales to the production unit for an inventory stock. However, for the sake of our discussion we separate out the production decision. We imagine that the buffer-information tradeoff is determined for the sales organization of Fig. 4 which then determines the channel capacity for the production decision of the expanded organization in Fig. 5.

where R is the maximal rate of output, or delivery of orders sustainable without error.[11]

Now let us say that the firm wishes to sustain a level of output L greater than R. Theoretically it would have to process $LH > A$ bits of information. However, were it to try to do this we would expect its rate of error to rise, i.e., it could be increasingly less successful in predicting which combination of regions was about to order. If we let P(e) be the probability that any decision will be wrong, then intuitively we would expect

$$dP(e)/d(L - R) > 0$$

We can show this more rigorously in the following way. The organization must process LH bits of information, but in fact can only process A. Thus as L rises H must fall. The only way in which H can fall is if the organization implicitly "bunches" some of the sets of regions together. That is, it begins to ignore differences so that "processed" H is less than actual H. Bunching will reduce processed H since the organization implicitly reduces the number of sets or orders it must choose from (which in our original case was 16). By reducing the number of sets it constructs larger sets whose probabilities are larger than the original sets. Thus in our original case we had an information field of value $H(.0625_1,\ldots,0625_{16}) = 4$ bits. But after

---

[11] In units we have,

(Orders/Decision)(Bits/time)/(Bits/Decision) = Orders/Time

where decision means choosing one of the K possible combinations of orders. Note that since $p_i = .5$, (Orders/Decision) = 1, i.e., if there are four possible combinations, (IO, II, OI, OO), then there are four I's.

exceeding A the new or processed information field could look like H(.25, .25, .25, .25) = 2 bits, where each new set of probability .25 is composed of four old sets.

More generally processed H will be $H(\sum^{k} p_i, p_{k+1} \cdots p_n)$ while actual H will be $H(p_1 \cdots p_n)$. But we know from the definition of information functions that

$$H(p_1 \cdots p_n) = H(\sum^{k} p_i, p_{k+1} \cdots p_n) + \sum^{k} p_i \, H(p_i' \cdots p_k')$$

where

$$p_i' = p_i / \sum^{k} p_i \quad [12]$$

Thus processed H is less than actual H, or symbolically,

$$H(s) < H(a)$$

Now clearly $d(H(a) - H(S))/d(L - R) > 0$, and we can assume $dP(e)/d(H(a) - H(s)) > 0$ so that $dP(e)/d(L - R) > 0$.

More generally if we let Q be the level of output we can write

$$dP(e)/dQ = 0 \qquad Q \leq R$$
$$> 0 \qquad Q > R$$

However, as the probability of error rises the firm will be less and less successful in delivering its goods to the proper customer. (Remember we assumed that the goods were tailored to the specific

---

[12] For the proof of this theorem, and for a good introduction to information theory in general see, Elwyn Edwards, Information Transmission, Chapman and Hall, London, 1964. The theorem is found on page 41.

needs of each region.) Thus if it is to deliver its goods, it will have to maintain an inventory of finished goods so that when the unexpected customer drops in it will not let him leave empty-handed. In other words, levels of output greater than R require some stock of inventories. But how much? We can get a feel for the requisite size and cost of the inventory stock by constructing a simple inventory model.

Thus imagine that there are only two rather than four regions, but that again the probability of order from each region is .5. We can then construct a table that relates the true or actual set of orders in any period to the firm's guess about which set will be ordered.

	IO	OI	OO	II
IO	(1-P)	P′	P′	P′
OI	P′	(1-P)	P′	P′
OO	P′	P′	(1-P)	P′
II	P′	P′	P′	(1-P)

Figure 9

P is the probability of error and $P' = P/3$. Thus we have assumed for simplicity, that the error is distributed equally among the different possibilities. Thus, if the combination II "holds", i.e., both regions order, then the probability that no goods will be produced that period, i.e., the firm guesses, OO, will be P/3. We can

view this table as an "input-output" matrix where the order is the input and the guess about the order is the output.[13]

Now the firm's rate of information transmission will be given as H(input) + H(output) - H(input, output).[14] Now H(input) and H(output) are each two bits each so that transmission is zero when H(input, output) is four bits. Since there are 16 entries in H(input, output), it can be four only when all the entries are equiprobable, or $1 - P = P/3$ or $P = 3/4$. That is, when the probability of error is $3/4$, the probability of being correct is equal to the probability of a mistake so that information transmission is zero.

Now let us consider the following set of "inventory policies" for the orders of one region only. (Call it region A, and the goods produced for that region

If decisions are made at a rate of D per period, then we say that there are D decision intervals per time period. We define an inventory policy X as the average percent of intervals in which a unit of goods A will be stored as inventory. Or, X represents the probability that in any given interval a unit of goods A will be kept on the shelves. Clearly the "maximal" policy will be represented as $X = 1$, since in this case a unit of goods A will be stored every

---

[13]Note that this matrix is not to be confused with the matrix transformation we constructed in §1. There it exhibited the coarsening process as information went from one node to the next. It involved the reduction of a set of four orders to a set of two orders. This matrix presumably represents the final effect of all coarsening, judgments, guesses, etc. made along the whole hierarchy.

[14]For a more careful definition of transmission see Chapter Three. Transmission is also defined in Edwards, op. cit., p. 54.

interval. In this case the rate of error will have no effect on the ability of the organization to deliver orders for A, since any error will be compensated for by drawing down on the "unit stock" of inventories of A.

Let us assume that the organization determines some optimal rate of "order completion", i.e., the proportion of orders it wishes to fill. In addition, we assume that since no provision is made for storing the goods produced for the other region (call it B), units of B produced as a consequence of error are disposed of. We make a parallel assumption for goods A. In other words, the inventory stock is built up only as a result of planned decisions.

Now let D be the rate of decisions, X the inventory policy, and P the rate of error. Now as can be seen in Fig. 9, there will be eight times when an order is not anticipated correctly, four for A and four for B. Thus the expected number of missed orders per decision interal will be given as

$$8(P/3)(\tfrac{1}{4})\ ^{15}$$

Then if the total number of decisions made per time period is D, then

$$8(P/3)(\tfrac{1}{4})D$$

will be the number of orders missed, and since the total number of orders is equal to the number of decision intervals per time period (since the probability of order is .5) then, $8(P/3)(\frac{1}{4})$

---

[15] $\frac{1}{4}$ appears in the formula because each combination of orders occurs with a probability of $\frac{1}{4}$.

will be the proportion of orders missed. Then when $P = \frac{3}{4}$ and transmission is zero only half the orders will be filled.

If we let $4(\frac{1}{3})(\frac{1}{4})$ be written as F and if we remember that X is the inventory policy for goods A and we let Q be the level of orders, then FPQ/2 will be the number of orders of B missed while FP(1 - X)Q/2 will be the number of A missed. The proportion of orders completed will therefore be

$$1 + \frac{-FP - FP(1 - X)}{2} = 1 - \tfrac{1}{2}FP(2 - X)$$

Now we assumed that the firm fixes some rate of completion of orders, call it RC.

Then we know that

$$1 - FP \leq RC \leq 1 - \frac{FP}{2} \qquad \text{since} \quad 0 \leq X \leq 1$$

Thus we can write RC as

$$1 - YPF \qquad \text{where } 0 \leq Y \leq \tfrac{1}{2}$$

Then since RC is fixed, we have

$$dRC = \partial RC/\partial Y(dY) + \partial RC/\partial P(dP) = 0 \ , \quad \text{or}$$

$$-FPdY = FYdP$$

so that

$$dY/dP = -Y/P \qquad 0 \leq Y \leq \tfrac{1}{2}$$

However, we know that

$$1 - YFP = 1 - \tfrac{1}{2}FP(2 - X)$$

so that

$$2(1 - Y) = X$$

and since

$$dX/dP = (dX/dY)(dY/dP)$$

we have

$$dX/dP = \tfrac{1}{2}(2 - X/P) > 0$$

so that the rate of buffering varies directly with the rate of error.

Now what will be the average cost of production? Cost will be the sum of actual production, storage costs and inventory production costs.[16]

If we assume constant returns to scale in production, then average cost of production can be given as K. Total storage costs will be given as costs of storage per decision interval or S times the number of decision intervals in which a unit of goods A is stored, DX or SDX. But since the number of orders Q = D, this can be written as SQX. But how many goods are actually produced? Since we assumed that the rate of error is distributed equally among all the possible sets of orders, this means that the number of orders is equal to the number of goods produced. (Though of course only a certain fraction of the goods ordered will be successfully matched with the goods produced.) So the unit storage cost per goods produced can be written SX. Average inventory production cost will be given as the inventory policy X, times the number of decision intervals divided by the amount

---

[16]We ignore "penalty" costs, since we are interested in deriving the resource rather than the opportunity costs of production. That is, we wish to discover the cost limitations to organizational growth, not whether or not growth is optimal.

of production, times the average cost of production or $KXQ/Q = KX$.[17] Thus total average cost of production will be given by

$$K + X(S + K)$$

Now as we saw before if A is the channel capacity of the organization and H the bit value of the sales region (in our case one bit) then $R = A/H$ was the maximal rate of output sustainable without error. Higher rates of order resulted in "information overload" and an increase in "bunching". This means that the information system begins to ignore differences so that potential orders from different regions can no longer be distinguished. As we saw from Fig. 9, a higher P means a lower rate of information transmission. Thus at the rate of orders exceeds channel capacity, the amount of information transmitted along the organization's information system falls and the rate of error rises. But as we say for a fixed RC, $dP/dX > 0$, so that as Q exceeds A/H, X must rise. But since average cost of production is given as $K + X(S + K)$ then the average cost of production must also rise. (Note that the expression for average cost can be thought of as the sum of average production cost plus average information cost, where buffers represent the implicit cost of information.

[17] Note that in our model there is no replacement of inventory. The inventory policy X does not state that there should be a certain stock at any given moment, but rather in some proportion of randomly chosen intervals there will be a unit of A.

So exceeding channel capacity results in a constant average cost of production but a rising information cost.

Graphically this can be represented in the following way:

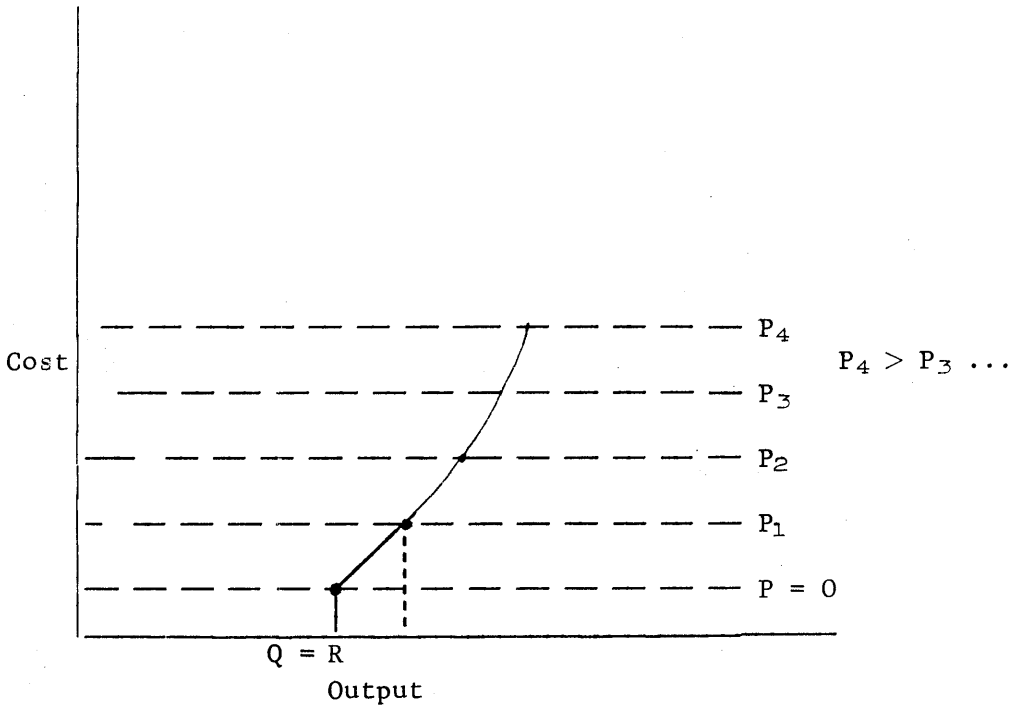


Figure 10

Each line represents average cost for a specified X and thus for an implied rate of error P. Then for a given RC as Q rises so does P. Thus average cost of production is a "crosscut" of these constant cost lines. The faster P rises with Q the steeper will be the average cost of production curve.

Thus we see that channel capacity is the "dual" of minimum average cost. In information theory, exceeding channel capacity will result in an information loss that varies directly with the

difference between the uncertainty of the input and the channel capacity. In our case, exceeding channel capacity results in a rate of error and average cost that varies directly with the point of production and the point of minimum average cost.[18]

Clearly our model is very simplified, but it does capture the way in which the buffering rate X rises with increases in the error rate, and under some simple assumptions we can see how this will generate a rising average cost of production. Richard Meier[19] shows that one of the major elements in "transactions overload", particularly for public service institutions, is the investment cost in "queuing" procedures as the rate of completion falls behind the order rate. While private institutions can simply sell to the highest bidder, too high a rate of output will strain internal channel capacities, so that queuing problems will develop within the organization leading to rising organizational costs.

To summarize, we have seen that the costs of information transfer and buffer stock maintenance will lead to an organizational structure that tends to minimize coordination costs. In turn such a

---

[18]Note that our model implicitly assumes that the cost of general information transfer and internal buffering is directly proportional to output, so that the average cost of such operations is constant. This will be strictly true if there are no fixed costs to such operations. Where there are such fixed costs, it is quite possible for the point of minimum average cost to be slightly to the right of the point of channel capacity. This point is taken up in more detail in the appendix to this chapter. For the purpose of our argument here, we assume that the average cost K includes both average production cost, and average information buffer cost.

[19]Richard Meier, A Communications Theory of Urban Growth, M.I.T. Press, Cambridge, Massachusetts 1962, p. 81.

structure will generate a channel capacity for the firm's production decision. Exceeding that channel capacity will result in a rising average cost of production, that in the end sets a limit to organizational size.

§3. Growth Through Structural Differentiation

We have seen that growth in the scale of production operations combined with a fixed channel capacity, i.e., a fixed information processing capability, will ultimately lead to a rising average cost of production. But information processing capabilities need not be limited. The number of decision makers can be increased. The lonely coordinators of Fig. 6 can find partners to assist them in the decision making processes. Thus an increase in the scale of both information and production operations should not *a priori* lead to a rising average cost of over-all operation. But it is clear that as the amount of information resources expands at each node in the hierarchy, or in each subsystem of the organization, the problem of intra-nodal coordination arises. How are the decisions within each node or subsystem to be coordinated, as the number of decisions, rate of decision making and information resources increase within each node? In other words, increases in scale generates a coordination problem, that if unresolved, will obviate any increase in channel capacity.

Perhaps the most likely solution will be the further differentiation of decision making labor, i.e., a division of decision making labor and a greater specialization of such labor. This proposition is of course simply a variant of the oldest theorem in economics, "the division of labor is limited by the extent of the market", or in our case growth in organizational output generates further structural differentiation or specialization within the organization.

The process of the division of decision making labor can be seen as part of the more general process of structural differentiation within the organization. Thus as sales, production, and design units develop semi-autonomous subsystems, specialized decision making processes will develop concomitantly to oversee and coordinate these units.

We can understand this process better by again employing the tree structure model of the sales organization depicted in Fig. 5.

In our sales organization model, greater specialization in the sales effort can be represented as a finer partitioning of the market area. That is each member of the set of N primary activities becomes a specialist for a smaller and smaller region within the area. This is shown in Fig. 11.

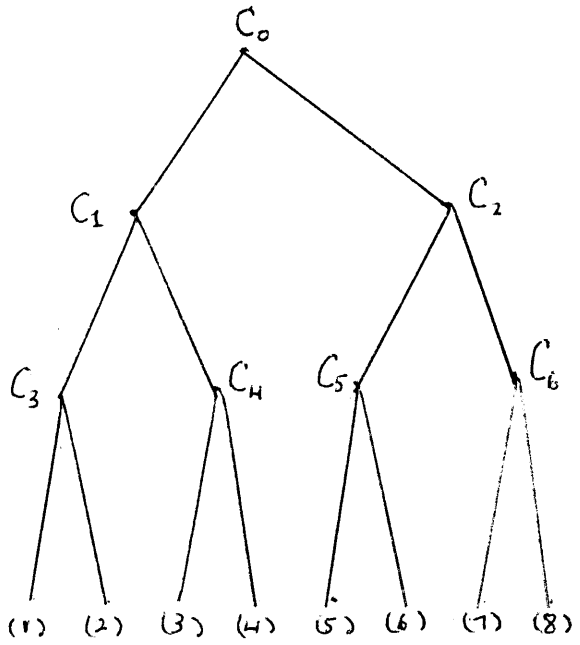


Figure 11

We continue to assume, however, that there are only four different products corresponding to the original four regions, and that in any decision interval there can be no more than one order per (original) region. (The probability of order from each region is again .5.) Under these assumptions, Coordinators C_3 through C_6 each scan the following information field.

Set	Probability
IO	.25
OI	.25
II	.50

Figure 12

Since we have assumed that the probability of order from each of the original four regions is still .5, the sum of the probabilities of the sets IO and OI (one order from each of the new "subregions") must sum to .5. Thus we assume that each of these subsets occurs with a probability of .25. The set II cannot occur since this would imply two orders from one of the original regions in a given decision interval -- contrary to our assumption of only one order per region in any given decision interval.

Each of the coordinators C_3 through C_6 will use the following transformation matrix in reporting to his superior.[20]

---

[20]Unlike the matrix in Fig. 7, this one is not "arbitrarily" chosen. IO and OI both mean a single order from one of the original four regions, while OO means no order from that region. Since from the perspective of production, we are interested only in orders from the original four regions, IO and OI naturally translate into 1 and OO to 0.

	I	O
IO	1	
OI	1	
OO		1

Figure 13

From the transformation in Fig. 13, it can be seen that both coordinators C_1 and C_2 scan a field of two bits each.[21] So after the first transformation the organization reduces its six bit field into four bits. But since there are only four different products corresponding to the original four regions, and since there can be no more than one order from each region per period, the organization will need only four bits for a correct production decision.

Then of what value were the extra two (6-4) bits of information that are picked up as a result of specialization? Presumably, specialization or the division of labor should increase the productivity of such labor. Or in our terms, the rate at which the individual nodes can process and transmit information should increase.

One way of seeing this is to view coordinator C_1 as being assisted by coordinators C_3 and C_4 in the scanning of the primary field. The task of scanning now has been divided up so that C_3 and C_4 do the initial scanning work, while at the same time the primary nodes need now survey on an even smaller region of the market area. Thus coordinator C_1 now receives more refined information which is presumably more easy to take up and transmit. Similarly the central coordinator will also receive more refined data so that the rate at which he can absorb information is also increased. In short, we would expect the increasing division of labor to increase nodal channel capacity. Thus let us imagine that the channel capacity of the organization

---

[21] That is, coordinator C_1 receives I and O each with a probability of .5, from both coordinators C_3 and C_4 for a total of (1+ 1) = 2 bits.

now rises from two to four bits per period, so that in effect the points of minimum average cost "shifts to the right". Under these assumptions there would seem to be no limit to organizational growth since continued structural differentiation should lead to greater and greater channel capacity and thus a constant average cost line.

What limits are there to specialization or structural differentiation?[22]

We can get some sense of the limits to specialization by recalling the complementarity between information and buffers. In Mathematics and Computers in Soviet Economic Planning,[23] Richard Judy notes that "as the economy grows more complex even the development of a feasible plan is beyond the power of the system".[24] He then goes on to quote the Soviet Economist, Aganbegian

> "...because of the growing interconnections of its (economy) sectors, the flow of information in the economy expands as the square (and sometimes the cube) of the expansion in the volume of production."[25]

If we interpret "complexity" and "interconnectedness" as the concomitants of structural differentiation this "squared" law of growth can be informally explained.

---

[22]J.H. Stevens deals explicitly with this question, but his results are not strictly applicable here. Stevens assumes increasing specialization but a fixed channel capacity, i.e., specialization without structural differentiation, so that the central coordination is not assisted in his scanning by intervening nodes. See, C.H. Stevens, op. cit., Chap. 3

[23]Richard W. Judy, "Information, Control, and Soviet Economic Management", in Mathematics and Computers in Soviet Economic Planning, ed., J. Hardt, Yale University Press, New Haven, 1967.

[24]Richard W. Judy, op. cit., p. 15.

[25]Richard W. Judy, op. cit., p. 15.

Intuitively we can see that part of this squared law derives from the mathematics of combinations. If we imagine a communications network composed of pairwise links then the number of links will be given by

$$\binom{N}{2} = (\tfrac{1}{2})(N^2 - N)$$

so that for large N the number of links will rise approximately with the square of the number of units. If we imagine that volume was proportional to the number of units then we can derive a rough variant of our squared law.

But this argument will not apply strictly for our organizational problem, since a central function of hierarchical information flows is to economize on the number of such pairwise links. More important, if information processing capabilities are fixed, no unit could process the amount of information required to maintain a growing number of such links.

Part of the answer lies in the complementarity between information and buffers. As Fig. 11 suggests structural differentiation with a constant span means a greater number of component parts in the organization. Thus each component or node communicates with a proportionately smaller part of the organization as the degree of specialization increases.[26]

---

[26]The assumption of a constant span is consistent with the notion that the information processing technology remains fixed. The channel capacity increases not because the information technology improves, but only because of specialization within the framework of a fixed technology.

One way of looking at this is to note that in the old organization of Fig. 5A coordinator C_1 scanned a two bit field in an organization that processed a total of four bits, or 50% of the information flow, while in the organization of Fig. 12, he scans two out of a total of six bits or 33% of the organization's information flow. In other words, as specialization increases the nodes or components in the organization process a smaller and smaller proportion of the organization' information flow. More generally, each node has a less comprehensive view of organizational processes. In other words, our coordinators find themselves in positions of interdependence with an increasing number of units but have no way of perceiving that interdependence. Information processing capabilities (expressed by the size of the average span, in this case two) do not keep pace with the growing interdependencies.

It is here that we would expect buffers to play a crucial role. The nodes or component parts of any organization must accommodate to their interdependencies, they cannot be caught short by unexpected strains and demands that arise from other parts of the organization. Thus we would expect that the amount of buffers maintained should rise with the increasing division of labor.[27] We can express this more rigorously in the following way.

Let G be the proportion of information flow processed by a decision making node. Let DS be the degree of specialization, i.e.,

---

[27]For a different approach to the problem of specialization and growth see, Harvey Leibenstein, Economic Theory and Organizational Analysis, Harper and Bros., New York, 1960, Chapter 8.

the "fineness" of the partition of the market area, N is again the number of primary nodes and M the number of coordinating nodes. I is the amount, in money terms, of buffers maintained by any decision making unit. Then we assume

$$I = f(G) \qquad f' < 0$$

but as we suggested, and as seems intuitively plausible,

$$G = f(DS) \qquad f' < 0$$

but by definition

$$DS = f(N) \qquad f' > 0$$

so that

$$G = f(N) \qquad f' < 0$$

but also

$$dM/dN > 0$$

so that

$$I = f(M) \qquad f' > 0$$

if we now write arbitrarily,

$$I = kM \qquad k > 0$$

and we assume k is the same for each coordinating unit we then have the total organizational cost

$$I = kM^2$$

Now how will such buffer costs affect the cost of production?

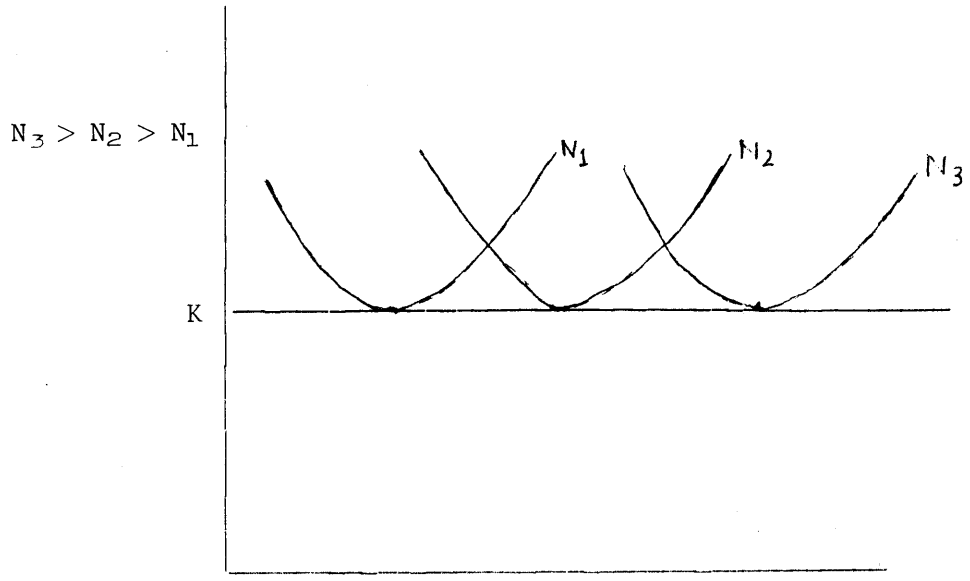


Figure 14

In the above figure we draw in different average cost of production curves corresponding to different degrees of specialization. K represents the sum of the average cost of production plus the average cost of information transfer also assumed constant. (See footnote 19, this chapter.) As we saw in section two of this chapter, a given organizational structure generates a channel capacity with respect to the production decision, and thus a rising average cost of production once channel capacity is exceeded. Thus each curve has a point of minimum average cost, corresponding to its channel capacity. As the figure suggests, organizations with a greater degree of specialization, will have a higher channel capacity, so that there points of minimum average cost will lie further to the right. (Curves with greater channel capacity will have higher fixed costs because to increase the number

of decision making nodes it must increase the amount of decision making labor. Thus at low levels of output they suffer from excess capacity and "overspecialization".)

If we let specialization be expressed as a continuous index, then it is clear that the long run cost curve for the firm will be the straight line envelope with height K. Each level of output on the long run curve represents a given value of N and thus some specified degree of specialization. If we let Q be output then we can write

$$Q = hN \qquad h > 0$$

but since

$$M = N - 1/S - 1$$

then

$$N = (S - 1)M + 1 \simeq (S - 1)M$$

so that

$$Q = h(S - 1)M \quad \text{or} \quad M = Q/(S - 1)h \ ^{28}$$

but since buffer cost I' is given by

$$I' = kM^2$$

we can write

$$I' = kQ^2/(S - 1)^2h^2 = VQ^2$$

where

$$V = k/(S - 1)^2h^2$$

---

[28] S is span which is assumed constant over the whole envelope. The formula relating M, N and S can be found on page

Thus long term average buffer cost will be given as

$$I'/Q = VQ$$

so that the long term <u>total</u> average cost of production will be

$$K + VQ$$

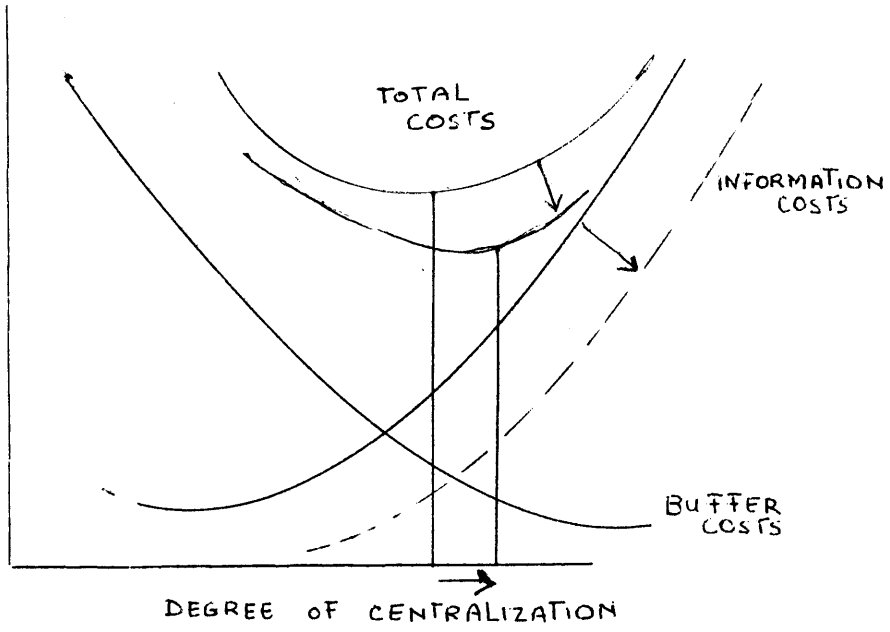
which clearly rises with output. Thus once again we can see that there will be coordination limits to organizational growth.

In summary we have seen that structural differentiation within the framework of constant information processing technology, expressed as some fixed average span, generates a higher channel capacity through the process of specialization. But the fixed information processing technology leads to coordination problems as the number of subsystems or nodes within the organization increases. These coordination problems result in rising buffer costs with a consequent rising average cost of production. This rising average cost will thus pose limits to organizational growth.

§4. Growth Through Restructuring and the Emergence of Information Capital

We have seen that a fixed information processing or coordination technology, expressed as an optimal degree of coordination, or optimal average span, imposes cost limitations on growth. A given organizational structure generates a fixed channel capacity with respect to the production decision, which when exceeded results in a rising average cost of production due to rising inventory costs. Similarly, structural differentiation with a fixed span permits a rising channel capacity, but at the cost of rising average buffer cost.

Thus we would expect that an improvement in the information processing technology should lead to increases in channel capacity. We can explore the dynamics of this process by returning to Fig. 2.



We represent a technological improvement in the information technology by letting the cost of information curve shift to the right. This results in a lower total coordination cost and a higher degree of centralization. This results in the elimination of some of the intervening nodes (between the field and the central coordinator -- the classic "middle level management") so that coordination is achieved with a smaller amount of information resources. Why should this improvement result in a higher channel capacity? The key lies of course in the increased centralization. There will be less coarsening as information moves up and down the hierarchy so that a greater proportion of the bit value from the field is relayed to the central coordinator. (This was explicitly demonstrated when we compared organizations 5A and 5B in Fig. 5.)

However, if we return to Fig. 5, we see that greater centralization means that the central coordinator must be able to process information at a faster rate, not only information with respect to the production decision but also information with respect to all other organizational operations. In our study of structural differentiation we found that the coordinator could process information at a faster rate because of greater specialization. The data was more refined and easier to "handle" as a result of the division of labor. But no such specialization occurs with centralization.

The key lies in our concept of Information Capital. An increase in the amount of information to the central coordinator is useless unless it is matched by an ability to structure or pattern this increased flow.

In Management Information Systems – New Directions, Z. Zannetos describes the case of the manager who when confronted with reams of computer output, but no guide to its perusal, simply chucks the whole affair in the wastepaper basket, and continues to make decisions in the old ad hoc fashion.[29] Thus if increases in the flow of information are to result in an increase in channel capacity, it must be matched by the development of a model, by the growth of information capital, that increases the productivity of "information labor". Thus centralization can lead to increases in channel capacity only if the technological improvements entail the development of information capital.

One way of seeing this is to note that organizational structure functions as an "implicit model". By generating certain kinds of information interactions and repressing others, it performs a selecting function, choosing to focus on certain relationships and ignoring others. As a model does, it structures information by choosing one set of "information matches" out of a large set of such matches. But as is suggested by Fig. 5, centralization reduces such organizational selectivity by reducing or simplifying organizational structure. Thus as the organization comes to structure information less and less an explicit model replaces organizational structure. We emphasize the word "explicit" since a consciously designed model reveals the full patterns of interaction to those members of the organization who have full access to the model. It is very much a "self image". An

---

[29] Zennon Zannetos, "Management Information Systems--New Directions", mimeo, Sloan School, M.I.T., March, 1967.

organization, on the other hand, is very much like the famed "invisible hand". No member of the organization has or need have a full view of organizational interactions. Like a market an organization can add up a series of "fragmented" views. Thus the emergence of information capital entails the substitution of a consciously designed model for organizational structure as the tool for patterning information.

Let us now return to Fig. 15 reproduced here as Fig. 16.

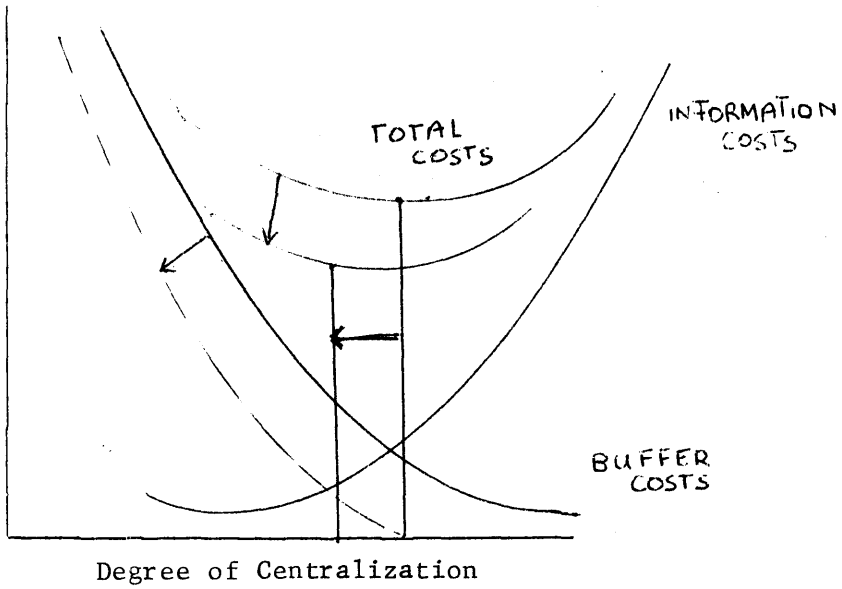


Figure 16

As we show in Fig. 16, organizational costs can be reduced by lowering buffer costs at every level of organizational centralization. Such a technological improvement will lead to a greater degree of organizational decentralization. How will this affect channel capacity, the rate at which the organization can process information without error?

We would expect that any technological improvement in information technology, such as represented by the downward shift in the buffer curve will lead to an increase in channel capacity. But how is this possible if as a consequence the degree of decentralization increases? Won't information coarsening increase, and as a consequence channel capacity fall?

Part of the key to this "paradox" lies of course in our original definitions. When we equated the degree of communication between subsystems with the degree of centralization (see §1, p.43), we implicitly assumed that all communication was of the hierarchic form only (i.e., information can flow only in a "vertical" direction). An increase in decentralization would be consistent with a rising channel capacity if, in fact, non-hierarchic forms of communication were permitted. This in fact is the most natural interpretation of the downward shift in the buffer cost curve.

The downward shift in this curve implies that at any level of centralization, the individual subsystems are better able to take into account the actions of other units on their own operations and thus need fewer buffers. In effect, the units are better able to coordinate themselves. The only way this is possible with an increase in the degree of decentralization, and thus a decrease in hierarchic information flows, is if non-hierarchic communication links are set up between the units, that substitute for and are cheaper than buffers. In other words, we interpret the buffer curve to include the cost of buffers and the cost of non-hierarchic (i.e., direct or "horizontal")

communications. A downward shift in this curve results in the progressive substitution of cheaper horizontal communication links for buffer. Thus improvements in information processing technologies are "double edged". They can be used to decentralize organizations by developing a technology of horizontal communication links between all the sybsystems or by improving the efficiency of hierarchical information flows.[30]

We can better understand the technology of horizontal communication links, if we assume that the subsystmes coordinate themselves through a "data base" that records the actions of all units and to which all units have access. The data base permits each unit to have a "total" view of organizational activities. Thus communicating through the data base permits direct information transfer without coarsening. It is for this reason that the channel capacity of the organization can increase. We can see that the data base is "integrative". It permits closer coordination of activities by reducing the need for buffers that separate the subsystems. Thus an increase in coordination is consistent with a decrease in centralization.

We can approach this process of decentralization in another way. If the subsystems are going to be able to process the increasing flow of information through the data base, they will require, as in the case of greater centralization, information capital, a model that enables them to structure this increased flow. Access to the "comprehensive"

---

[30]Admittedly, this ignores the costs of goal formulation. We are discussing the costs of coordination once goals are set.

or "total" will prove impossible without a model to pattern organizational information. Thus the process of decentralization that increases coordination results from a decentralization or democratic distribution of the organization's information capital.

There are thus two kinds of decentralization. An organization will be decentralized if given a fixed information processing technology, the organization is so structured so as to reduce the proportion of information flow to the central coordinator, without increasing the proportion of this flow to any other unit. Such a decentralization will result in a lower degree of coordination and thus a lower channel capacity. The other mode of decentralization consists in increasing the proportion of organizational information flow to each subsystem within the organization by democratically distributing the information capital.

However, whether information capital is used to centralize or decentralize organizational structure, its emergence is always a destructive process. It always destroys decision making prerogatives, whether these be the prerogatives of the "traditional elites" or the prerogatives of their subordinates. The emergence of information capital, thus entails the emergence of new levels of organizational conflict. (This is taken up in detail in Chapter 4.)

In summary, if an organization is to continue to grow, it must increase its channel capacity by developing information capital. But the development of such capital can be double edged. It can result in

greater centralization if the information capital is located increasingly in the central coordinating node, or it can lead to greater decentralization, to the progressive elimination of a "center" as the point where information from the entire organization is coordinated, if the information capital is democratically distributed.

This we shall argue sets the stage for political conflict around decision making prerogatives. The potential inherent in the developing information processing technologies points to a more democratic yet more coordinated decision making. Yet the prerogatives of those in power may lead to a suppression of this potential in the interests of less democratic and more centralized modes of decision making.

Summary

We can summarize our discussion by way of the implicit "taxonomy" of organizational growth developed in this chapter. The chart below lists both the forms of growth and the information limits to growth. We like to view the chart as outlining a sequential process.

	Form of Growth	Information Limit
1)	growth in scale, fixed channel capacity	rising rate of error leads to rising average inventory cost
2)†	growth in scale of information and material resources	problem of "intra-nodal" coordination
3)‡	growth through structural differentiation (implies an increase in material and information resources)	channel capacity increases but buffer cost rises more than in proportion to output
4)	growth through "restructuring", and reintegration through the development of information capital.	

†We think of this as a "transitional stage" as greater scale permits a greater division of labor.

‡Often it is at this point that the organization will "spin off" some of its subsystems. The costs and function of coordination are then taken over by the market.

The central analytic points that emerged from our discussion can be listed as follows:

1) Buffer costs are the dual of information costs. The balancing of the two leads to an optimum degree of centralization.

2) Channel capacity is the rate at which choices of a given degree of difficulty can be made. It is the dual of minimum average cost.

3) Continued organizational growth leads to the growth of information capital.

   a) Information capital generates a more comprehensive organizational "self image".

   b) Information capital permits an increase in organizational channel capacity.

   c) Information capital replaces organizational structure as the tool for patterning information.

   d) Information capital entails the restructuring of decision making prerogatives.

   e) Information capital can generate both a more decentralized yet more coordinated mode of decision making.

APPENDIX

This appendix is for the reader who would like to see all the implicit assumptions and details hidden in the footnotes of the text brought together in one place.

Our buffer-information curve diagram showed how an organization will arrive at its optimal degree of centralization. For the purposes of our argument we assumed that these curves include all tradeoffs except that between sales information and inventory stock. Thus the tradeoffs can include allocation of the firm's resources between the subsystems, new investments by the subsystems, forms of excess capacity, and operations with high fixed costs that are not shared effectively. Lack of communication between subsystems with respect to all these operations can lead to "suboptimization", duplication, and forms of excess capacity. All these results are encompassed under the general terms of "buffers".

We assumed that these tradeoffs determined an organizational degree of communication for the sales organization depicted in Fig. 4. We then argued that a given degree of communication implies a given degree of centralization, implicitly restricting communication to hierarchic information flows. A given degree of centralization then generates a channel capacity with respect to the production decision, i.e., for the expanded organization in Fig. 5.

Rigorously speaking, the information buffer curves of our basic diagram must be specified with respect to some level of output, since for some degree of centralization the higher the level of output, the

higher both the cost of information and buffers. However, we implicitly assumed that once the firm finds its optimal organizational structure, this structure will prove optimal, i.e., will result in the lower total coordination cost, over all levels of output. For this to be true, the shifts in the information and buffer curves must exactly offset one another. This is shown in Fig. 18, where as a result of these shifts the same structure, the same degree of centralization is preserved over all levels of output.

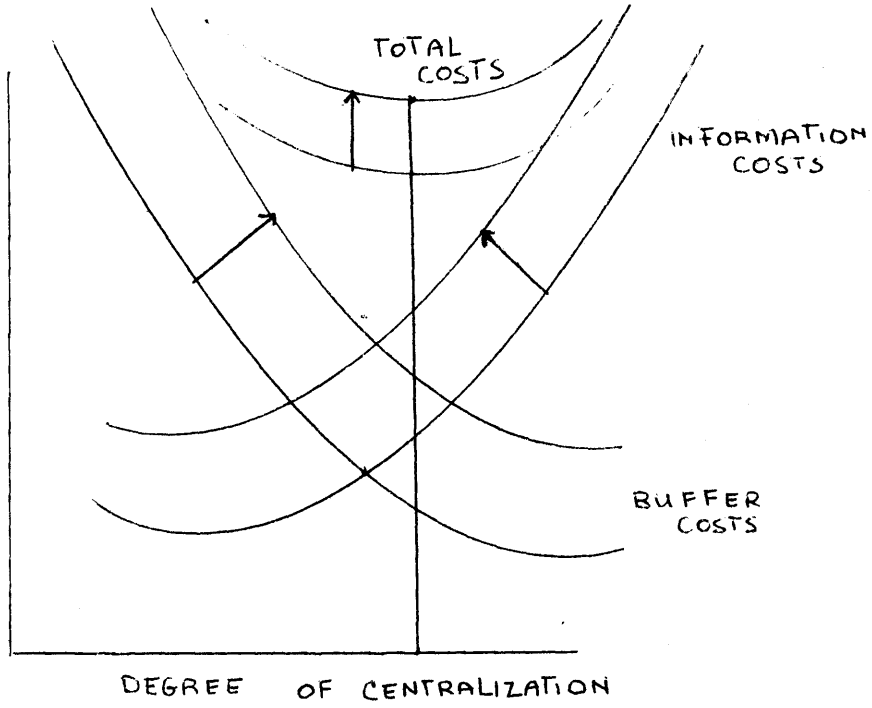


Figure 17

Is there any reason to expect such an offsetting shift in the curves?

We can understand the problem better by thinking in terms of cost curves. Thus if every organizational structure generates a

cost curve of coordination over different levels of output our assumption implies that the cost curve of the optimal structure lies everywhere below the cost curves of any other organizational structure. This is shown in Fig. 19.

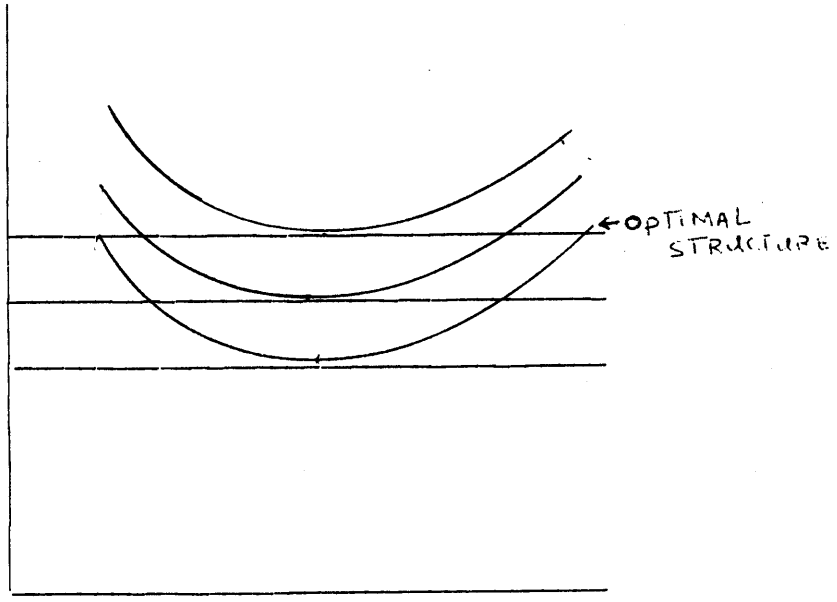


Figure 18

However, when we usually think of choosing the best cost curve from a set of curves we usually have the following "picture" in mind:

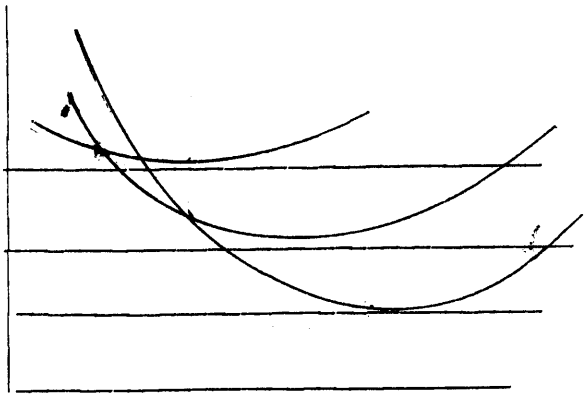


Figure 19

Thus we imagine that cost curves that are more efficient over higher levels of output are less efficient over lower levels of output. But this is because we associate this greater efficiency with higher fixed costs. We imagine that the greater efficiency is the result of more capital intensive modes of production that entail a higher level of maintenance expenditure.

In our case the parallel to different capital intensities that generate different modes of production are different organizational structures. But structure per se does not cost.[31] There are no fixed cost differentials to different structure. Different structures simply represent different ways of organizing a constant complement of human and material resources within the framework of a given coordination technology.[32] Thus we would expect that an organizational structure optimal at one level of output will be optimal at all levels of output.[33] It is for this reason that we would expect the shifts in the buffer and information curves to be mutually offsetting.

We cannot determine a priori how all buffer and information costs will vary with the level of output.

---

[31]This is a point that emerges very clearly from Leibenstein's studies of X-efficiency. See Harvey Leibenstein, "Allocative Efficiency vs. X-efficiency", American Economic Review, June 1966, 392-415.

[32]Note that this is not inconsistent with the assumption of higher fixed costs for structure of a greater degree of specialization. The specialization is achieved through the addition of human and material resources.

[33]To give an analogy, a smart fellow will solve problems better than a stupid one whether the problems are hard or easy.

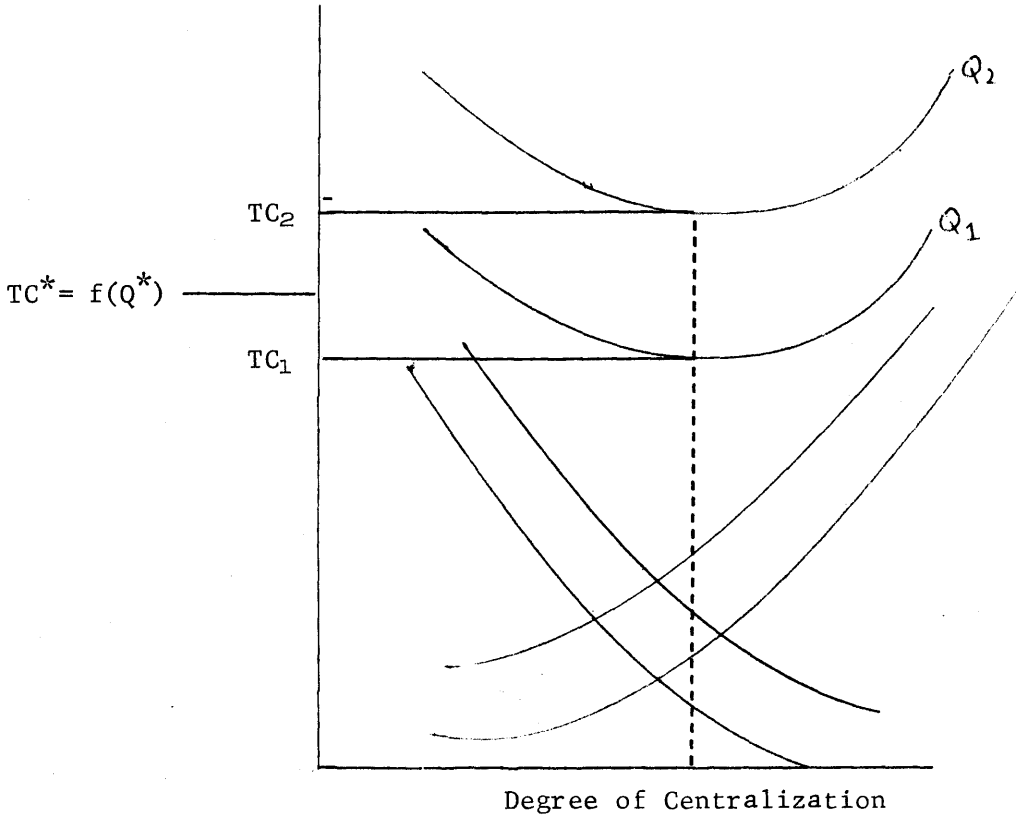


Figure 20

In Fig. 20, Q* represents the point of channel capacity with respect to the production decision. Now clearly, $(TC_2 - TC_1)/TC_1$ can be less than equal to, or greater than $(Q_2 - Q_1)/Q_1$, so that there can be increasing returns to such information-buffer costs particularly if there are fixed costs to such operations. However, just as the production decision has a channel capacity point, so we would expect other such decision or information transfers to have their channel capacity points. Beyond these points we would expect the information-buffer costs to rise more than in proportion to output, for reasons similar to those which we discussed with respect to the production

decision. However, these channel capacity points, call them $Q_1', \ldots Q_n'$, may not coincide with the channel capacity with respect to the production decision, Q*. Thus, if $Q_i' > Q^*$, then the point of minimum average cost would lie to the right of Q*, and minimum average cost would be a "weighted average" of the different channel capacity.

INFORMATION CAPITAL

from the Perspective of Software

Chapter Three

THE MODEL AS A CAPITAL GOOD

Introduction

We have emphasized throughout the previous chapters that an increased flow of information can be useless or even counterproductive, if not matched by an improved ability to structure or pattern such information. Hardware innovations require software development to harness the "information explosion". Software can be thought of as a set of models that pattern and channel information. Thus, we would expect that the "productivity" of information flows would depend on the quality of the model. Intuitively we would expect that models that reveal little of the inherent structure in the field or environment will be less productive, less useful than those that expose the interdependencies and interrelationships. We will say that the latter more "sophisticated" models contain more structure than the naive ones. We will show how models of greater structure possess a greater "channel capacity" and how as a consequence the average cost of sampling falls as model structure rises.

In §1 we present an intuitive view of these propositions anticipating in part the conclusions of Part Two. In §2 we establish a more rigorous exposition. The thrust of this chapter is to establish the relationship or "duality" between "better" models, the Shannon concept of channel capacity, and average cost. This student did not

possess the requisite skills to prove the "general case". Nevertheless, the results are consistent with our original intuition and are suggestive as to how proofs might be generalized.

§1. An Intuitive View of Structure and Transmission

We said that more sophisticated models possess a greater amount of "structure". This concept can be quantified. Thus imagine that a coordinator or decision maker must consider two variables in his or her decision making process. These variables might represent data on the internal processes of an organization, or data from the organization's external "field". The two variables are characterized by a joint probability distribution as shown in Fig. 1. Each variable can assume only one of two states

	X	x_1	x_2	
Y	y_1	.5	.25	(.75)
	y_2	.125	.125	(.25)
		(.625)	(.375)	1.0

Figure 1

The amount of "structure" in this joint probability distribution is given by the information analogue of the "degree of corellation". Thus, if we let Structure be equal to S, we have

$$S = H(x) + H(y) - H(x,y)$$

where H(x), H(y) are information value of the marginal distributions of x and y and H(x,y) is the information value of their joint distribution. Thus, we have in our example[1]

[1]The term structure is called the "transmission" between two variables by Quastler and McGill.

$$S = H(.75, .25) + H(.625, .375) - H(.5, .25, .125, .125) = .015 \text{ bits.}$$

Now we assume that the table given in Fig. 1 is in fact correct, that it represents the true probability distribution of the organization's relevant field. We define a model of complete ignorance as one whose structure is zero. Since in fact different configurations of probabilities will lead to a model of structure zero, we restrict our range of consideration to that class of models where the marginal distributions are unchanged. Thus, every model reflects a correct understanding of the marginal distributions of x and y but differing conceptions of their joint probability, of their dependence.

Under such conditions it can be shown that the model of zero structure is given, not unexpectedly, by the model where $P(x,y) = P(x)P(y)$. That is the probability P of the joints is given by the multiplication of the respective marginals, i.e., the variables are assumed to be orthogonal.

In Fig. 2 we present the model of zero structure for the field given in Fig. 1.

	X	x_1	x_2	
Y	y_1	.46	.29	(.75)
	y_2	.165	.085	(.25)
		(.625)	(.375)	1.0

Figure 2

so that Structure is given as

$$S = H(.75, .25) + H(.675, .375) - H(.46, .29, .165, .085) = 0$$ [2]

It can also be shown that structure will be given by

$$S = H(x) - H_x(y)$$

where $H_x(y)$ is the uncertainty of the conditional distribution of y given x. Thus as structure increases, the uncertainty of y given x falls since H(y) is assumed constant. Intuitively, once x is known the uncertainty associated with y falls as understood interdependence of the two variables approaches the true interdependence.

---

[2] To prove this for the general case note that,

$$S = \Sigma P(x) \log P(x) + \Sigma P(y) \log P(y) - \Sigma P(x,y) \log P(x,y)$$

but in the case of complete ignorance

$$\Sigma P(x,y) \log P(x,y) = \sum_x \sum_y P(x)P(y) \log P(x)P(y)$$

but this last right hand term can be written as

$$\sum_x \sum_y P(x)P(y) \log P(y) + \sum_x \sum_y P(y)P(x) \log P(x)$$

which in turn equals

$$\sum_x P(x) \sum_y P(y) \log P(y) + \sum_y P(y) \sum_x P(x) \log P(x)$$

which yields

$$\Sigma P(y) \log P(y) + \Sigma P(x) \log P(x)$$

so that structure is zero.

For all the basic definitions and properties of information functions see, Gardner, op. cit., Chapters One and Two.

The Concept of Transmission

Before we can rank the "productivity" of models of differing structure, we must first develop a measure of model performance.

Information theory provides a convenient measure in the concept of "transmission" over a channel.

Let us label the four states of Fig. 1 in the following way:

State	1	2	3	4
Probability	.5	.25	.125	.125

Figure 3

Now imagine that in any given decision making period, only one of these four joint probability states actually "holds". That is in any given period the variables x and y can be in only one of their two states. Assume in addition that the firm's decision making process requires that the firm determine which of the four states actually holds. (Or, we can put it in predictive terms and say that the firm must anticipate which of the four states will hold in the coming period.) To do this, the firm engages in some form of sampling (about which we will be more explicit in the coming section) and then makes an "educated" guess.

We can then record its guesses over some period of time, and then construct an "input-output" transmission table. The input represents the number of times each state actually held, and the output, the number of times each field was guessed to hold. The series of joint input and output states represents the distribution of the

input, or the actual occurrences, over the output, or the guesses. Thus the "source" is the field, the channel is the sampling process, and the output is the guess. Figure 4 presents such an input output table for the field represented in Figs. 1 and 3.

	Output	1	2	3	4
Input	1	4			
	2		1	1	
	3			1	
	4	1			

Figure 4

Reading the table we see that out of eight "occurrences", state 1 occurred with a probability of .5 and was correctly identified each time. State 2 occurred with a probability of .75 but was correctly identified only $\frac{1}{2}$ of the time.[3]

Now transmission is defined as the amount of "structure between the input and output. Thus, if we let T be transmission, I the input and O the output we have,

$$T = H(I) + H(O) - H(I,O)$$

In our case this yields

$$T = 1.75 + 1.5 - 2.0 = .25 \text{ bits.}$$

---

[3]Note we assumed that with a sufficient number of periods the states occur in proportion to their true probabilities. Thus state one occurred .5 of the time, state two .25 of the time, etc.

Note that when transmission is "perfect", i.e., each state is correctly anticipated, or identified, then we have

$$H(I) = H(O) = H(I,O)$$

so that

$$T = H(I) \quad [4]$$

Now we can imagine that if the firm spent more money in its sampling process, its guesses would be better and the amount of transmission "through" the transmission process would increase. Completely error free transmission would result in an output of 1.85 bits. Presumably this would cost an infinite amount per sample, thus perfect transmission could be approached only asymptotically.

This concept is presented by the following graph:

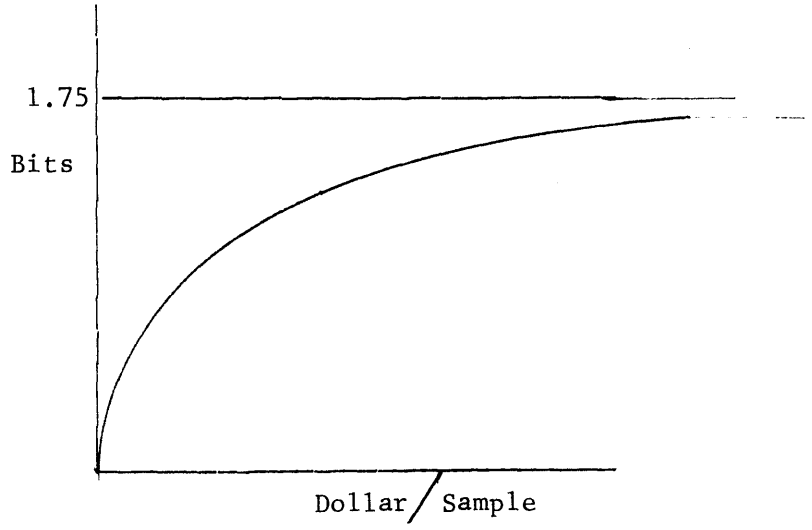


Figure 5

[4]Note that the definition of transmission does not distinguish between more severe and less severe mistakes. It is for this reason that it has been critical as a measure of performance. We implicitly assume throughout that all errors are equally serious.

Now we can express our intuitive concept of the relationship between structure and channel capacity in the following way. Given a fixed dollar per sample expenditure, a model of greater structure should result in a greater amount of transmission. That is, we feel that a greater understanding of the interrelationship between x and y should result in a more "efficient" sampling process that takes advantage of this increased understanding. Thus in Fig. 6 we have:

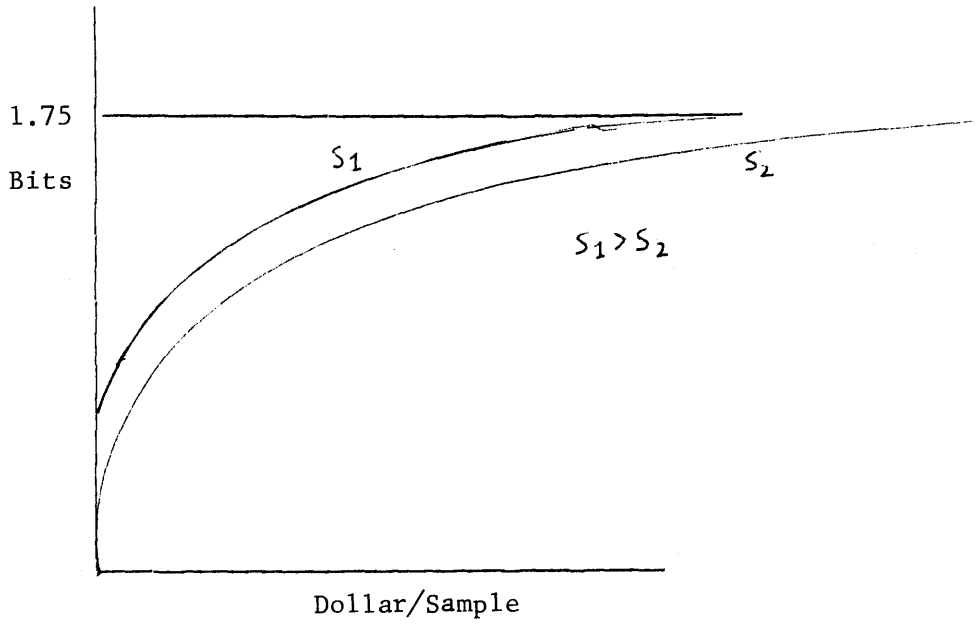


Figure 6

Model one possesses a structure S_1 greater than S_2.[5]

However, we later show that a better representation of this concept is developed in Fig. 7.

---

[5]Note that there can be such a thing as too much structure, i.e., attributing too much interdepednence to the variables then there actually is. Let us assume that we include only values of structure up to, and including, the true structure.

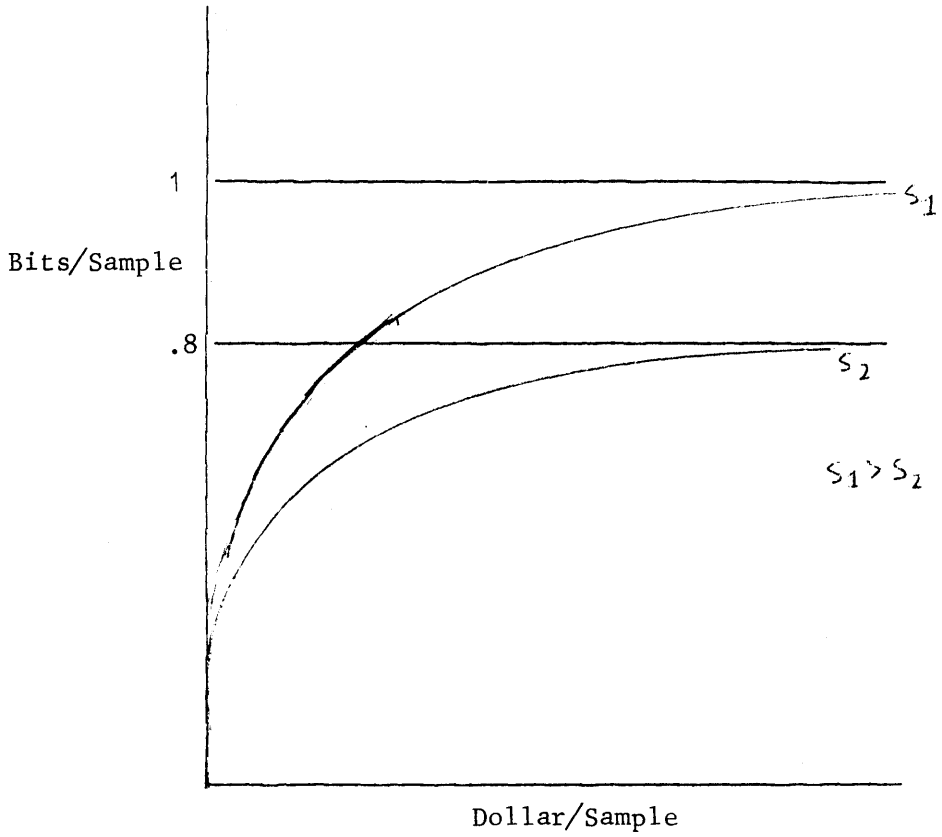


Figure 7

On the Y axis we plot the number of bits per sample transmitted, and on the X axis the cost per sample.[6] The model of complete structure approaches an asymptotic value of one bit per sample, while models of lower structure approach lower asymptotes. (One bit per sample is the maximum due to the "binary character of the sampling process.) If we think of bits/sample as the analogue of channel capacity, normally expressed as bits/time, our graph captures the notion that models of

[6]Intuitively we can see that bits/sample rather than bits is a better measure of model performance. If enough money is spent, any model will permit the transmission of the full 1.75 bits. As we shall see, bits/sample is a measure of the efficiency of different models, and each model has a "maximum efficiency". This maximum efficiency is the analogue of Shannon's channel capacity measure.

greater structure increase channel capacity. The transmission limit of any model, expressed as its asymptotic bit/sample value, is overcome by models of greater structure. It is this sense that greater structure "amplifies" the sampling process by "building" channel capacity. If we say that models of greater structure possess more information capital, we can see that information capital increases the productivity of the sampling process.

§2. The Sampling Process, Rank Value and the Cost of Perfect Information

We would like to develop in a more rigorous fashion the relationship between sampling cost, bit per sample and the amount of structure. In this section we present the sampling process as a "coding problem".

Imagine that a firm faces a problem analogous to that of Chapter Two. There are four regions to which it sells products, and in any given period one and only one region will order.[7] To produce the correct product, it must anticipate which of these regions is about to order. The probability that any given region will order is given by the distribution presented in §1, where each region corresponds to one of the states.

Region	A	B	C	C
Probability	.5	.25	.125	.125

Figure 8

We assume that the firm engages in sampling to make the best guess as to the sources of the coming order. Its sampling process will be of the Bayesian variety. Thus imagine that its sampling process consisted of one sample only (an assumption to be modified shortly). Each region has an *a priori* probability attached to it,

---

[7] We assume that one and only one region will order per period rather than some combination of regions. In this way we can stay within the simple confines of binary arithmetic. The problem is analogous to the toss of a die. Only one face can come up in any given throw. Note we implicitly rule out "cross-period" probabilities, that is, the probability of region A is considered to be .5 irrespective of events of previous periods.

and on the basis of the sampling result obtained the likelihood that this result could have been obtained is determined for each region. Combining these likelihoods with the a priori probabilities would yield the a posteriori probabilities. The region with the highest a posteriori probability would be chosen.

Let us now assume that the firm knows the correct distribution as shown in Fig. 1. On the basis of this distribution, what could constitute the firm's most efficient sampling process? Clearly, since A occurs half the time we would like to devise a sampling process that was relatively inexpensive when in fact A occurred. It would be foolish to have to spend the same amount to discover A as to discover D, since D will occur with only one fourth the frequency of A. Phrased in this way, the problem of the most efficient sampling process, the process that minimizes sampling cost is the "dual" of the coding problem over a binary channel. (A binary channel is one that is either off or on, analogous to our assumption that each region will or will not order each period.)

The Coding Problem in Brief

Imagine that we wish to send a series of four messages over a binary channel where the messages occur with the following probabilities:

M_1	M_2	M_3	M_4
.5	.25	.125	.125

Figure 9

Since the channel will be either off or on, a code will consist of an assignment of zeroes and ones (off and on) for each message.

To find the most efficient code partition the set of messages into two sets, each with the same cumulative probability. Thus the first partition will be:

Message	M_1	M_2 M_3 M_4
Probability	.5	.5

Figure 10

Then assign to one set the value zero and to the other the value one. Continue by partitioning the next subset in a similar fashion again assigning zeroes and ones. When the partitioning is complete each message will be associated with a set of zeroes and ones.[8]

This coding process is best represented by a tree structure. (See Fig. 11)

This code is most efficient in that the expected number of zeroes and ones that will be sent is at a minimum. The expected number can be derived by multiplying the code length of each message by its probability. When the code is the most efficient possible,

---

[8]This is exactly analogous to the "find the name in the telephone book" parlor game. The most efficient way to proceed is to divide the book in half, ask which half it is in, divide that section in half, ask again, etc. In this way you minimize the number of guesses required to locate the name.

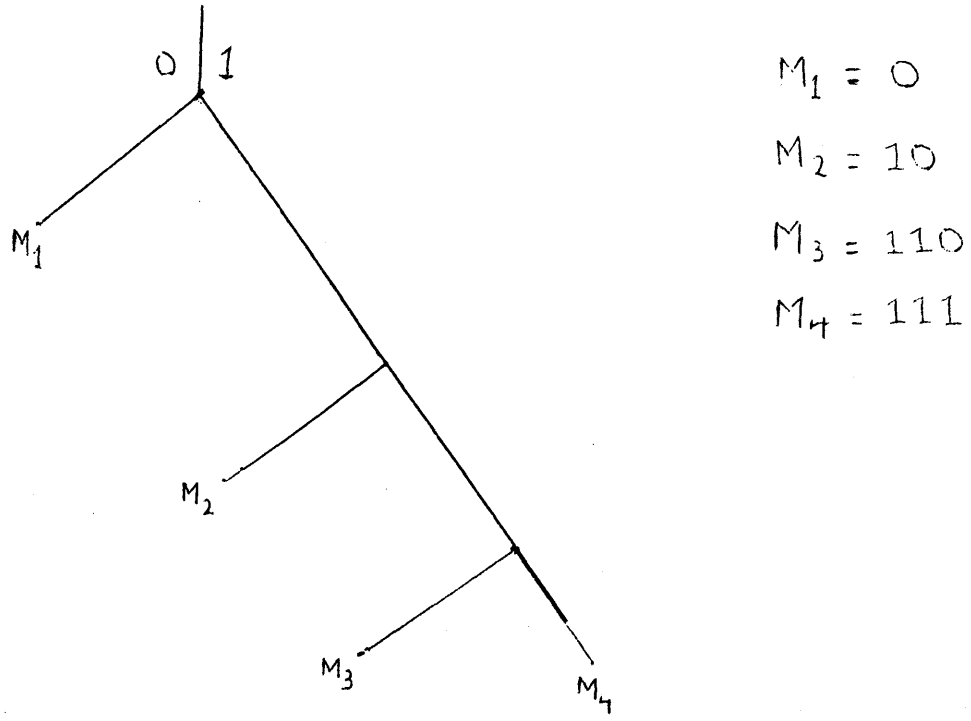


Figure 11

this expected number will be equal to the uncertainty value of the distribution, in this case 1.75.[9]

How can this tree structure be interpreted as a sampling process? Let each node on the tree represent a single sample. Let the symbol v mean "and/or". Then the first node represents a sample that enables the firm to distinguish between region A and region BvCvD. The second node will be a sample that distinguishes between B and CvD, the third between C and D.

How in fact will the firm distinguish or choose a particular branch at any given node?

[9]Not all probability distributions can be so divided, in fact only where the probabilities are negative powers of two can the field be so exactly divided. In cases where this is not possible, we divided our message set into subsets so as to come as close as possible to equiprobable divisions. Coding theory shows that this may in fact not be the most optimum mode of division. Specifically, we can code for longer sequences of messages (i.e., rather than just code for M_1 and M_2, code for M_1M_2, etc.). But this has no natural interpretation in our problem. It is equivalent to having the firm predict orders for more than one period. We will simply ignore this problem, using only exactly divisible probability sets in our example.

Let us consider the first node. If the firm understands the underlying market forces that shape the probability distribution, it will be able to select an "indicator" that will distinguish efficiently between the possibility of A and the possibility of BvCvD. If we let the value of the indicator be I, then the sampling process will result in the following Bayesian calculations,

$$P(A/I) = P(A)P(I/A)/P(I)$$

and

$$P(BvCvD/I) = P(BvCvD)P(I/BvCvD)/P(I)$$

The likelihoods, P(I/A), P(I/BvCvD), reflect the firm's understanding of the relationship between the indicator and the two possible events. An efficient indicator would result in divergent likelihoods for the same event. The firms will choose that region whose a posteriori probability is greatest.

Thus each node can be regarded as a sampling point, where a selected indicator is sampled and used to distinguish between the two branches on the node. Different tree structures will result in a different set of indicators, since the possible regions to be compared at each node will be different. What will be the most efficient tree? In our coding theorem the most efficient tree was one that minimized the expected or average code length of the messages. But as is clear, code length is equivalent to the number of samples required to determine if a given region will be ordering next period. Thus the efficient

tree structure would be one that minimized the expected number of samples and will be given by the binary division process described above.

Assuming that the firm uses the most "efficient" tree structure what will be the cost of "perfect" information? Let us assume that the indicators at the different nodes are equally "reliable", e.g., they have the same variance. Then let S be the amount in dollars per sample required for "perfect" (or as perfect as possible) information. Then the expected cost of perfect information will be given by S $\times$ H where H is the uncertainty of the distribution itself.

We can get a clearer idea of the meaning of an efficient structure by asking, what would the firm's tree structure look like if it had an incorrect conception of the relevant distribution?

Thus imagine that it posits the following model:

Region	A	B	C	D
Probability	.125	.125	.25	.5

Figure 12

while the true model is still

Region	A	B	C	D
Probability	.5	.25	.125	.5

Figure 13

If the firm follows the binary porcess described above it will construct the following tree:

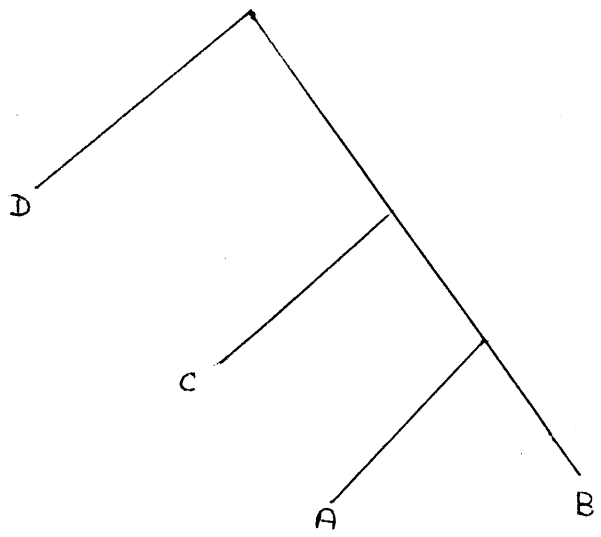


Figure 14

What will the cost of perfect information be to the firm? We simply multiply the number of samples required to reach each region by its probability of occurrence, times the cost per sample of error free information, S. So we have,

$$1 \times (\tfrac{1}{8}) + 2 \times (\tfrac{1}{8}) + 3 \times (\tfrac{1}{4}) + 3 \times (\tfrac{1}{2}) = 2.62 \times S$$

while for the "correct" tree structure we would have

$$1 \times (\tfrac{1}{2}) + 2 \times (\tfrac{1}{4}) + 3 \times (\tfrac{1}{8}) + 3 \times (\tfrac{1}{8}) = 1.75 \times S$$

This can be generalized in the following way. The cost of perfect information for model j, call it PI_j, will be given as

$$PI_j = -S \sum p_i \log_2 p_j$$

where p_i is the <u>actual</u> or <u>true</u> occurrence of the region and p_j is its <u>supposed</u> occurrence in the model j. (A little counting will show that $-\log_2 p_j$ is the number of nodes required to reach a region on the tree structure generated by model j.) In the case of the correct model this collapses to

$$-S \sum p_i \log p_i = H(p_1 \cdots p_n)$$ [10]

This calculation is presented for three different models.

Region	Model 1	Model 2	Model 3	True Model
A	.25	.25	.125	.5
B	.5	.125	.125	.25
C	.125	.5	.25	.125
D	.125	.125	.5	.125
Expected Cost	S(2.00)	S(2.37)	S(2.62)	S(1.25)

Figure 15

We can now define an "information distance" between any model and the "true" representation as,

$$-\sum p_i \log p_j - (-\sum p_i \log p_i) = \sum p_i(\log p_i - \log p_j)$$ [11]

Note that since each model implies a given tree structure which in turn generates a set of indicators that compare events "pairwise",

---

[10] Note that we have implicitly assumed that the cost per sample required for error free information per sample is the same for different models. We might argue that "more incorrect" models should also generate less efficient or less reliable indicators. This is possible. To the extent that this is true any cost differences we generate with respect to perfect information for different models must be regarded as minimum differences. We will continue to assume that S is the same for all models. This point will be discussed when we discuss transmission with error.

[11] Note that this can in turn be written as $\Sigma p_i \log (p_i/p_j)$ which is strictly analogous to the measure of the additional information obtained between the a priori and a posteriori probabilities, i.e., the information gained in any sampling process in general. Our measure can thus be regarded as a measure of the additional information gained by having a correct model. Strictly speaking this information distance measure will hold only for the cases of exact binary division. In other cases $\log_2 p_j$ will not correspond precisely to the number of nodes required to reach the region. But to the extent that $\log_2 p_j$ can be used as a measure of the amount of information in an event with probability p_i (and this is not strictly kosher since only sets have information values, not single events), our measure is an extention of this use. It is a suggestive way of measuring how "wrong" a model is, though only in the case of exact binary division will it correspond to cost differences.

we can think of more efficient or better modelling as one that selects the better subset of the large set of pairwise relationships that exist between relationships. (Thus the tree structure generated by models one, two, and three on the previous page, is depicted in Fig. 14. It permits twelve possible pairwise comparisons between sets of regions. Each model selects only a subset of four.)

In other words, we can use many different concepts (indicators) to partition a conceptual field, there are many ways of looking at a problem, but there will be one subset of concepts, associated with the most efficient sampling process, which like Occam's razor permits of the most economical understanding.[12]

Let us now call the expected number of samples for any model its "rank value". Thus the rank value of model one in Fig. 15 is 2.00.

We can now interpret more fully the sources of "efficiency" of the models with lower rank values. Thus let us consider the tree structure of model three in Fig. 15.

The probabilities written along side the "branches" of the tree are the <u>true</u> relative frequencies of the subsections being compared at a given sampling node. The probabilities written at the end points are the <u>true</u> probabilities of the regions at these points.

---

[12]The concept of an efficient tree structure for a model is exactly analogous to the concept of an efficient tree structure for an organization. In both cases the efficiency is a function of the way in which information is organized. In the model certain concepts -- indicators -- are developed, while others are suppressed. In the organization certain interactions are recognized while others are ignored.

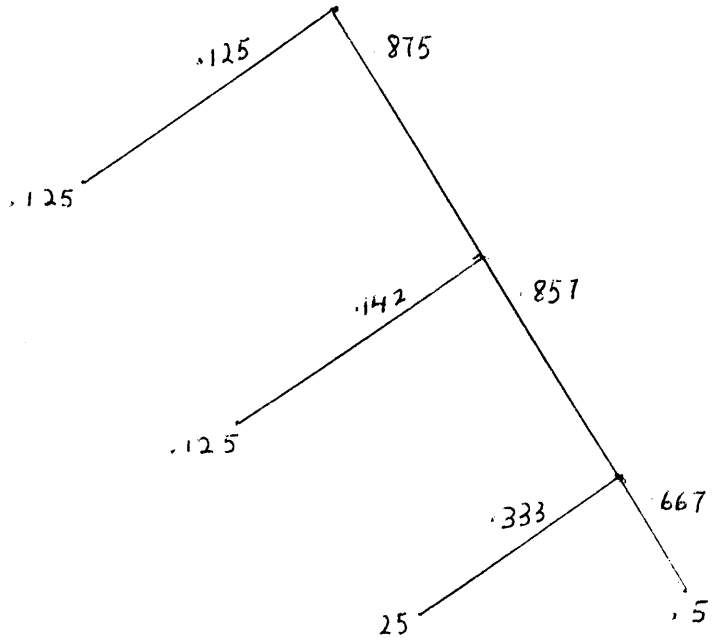


Figure 16

Now we know that the maximum bit value of any binary choice is one, and is given where the two choices are equiprobable, i.e., H(.5, .5) = 1. Thus the bit value of each sampling choice of the tree of model three will be less than one. In other words, an incorrect model does not make the "most" out of a binary choice. It loses bits per sample at each node.

Can we calculate the average or expected bit loss? First take the bit value at each node and subtract it from to get the bit loss at each node.

Sampling Node	(.125, .875)	(.142, .857)	(.333, .667)
Bits Loss	.4566	.4081	.0822

Figure 17

The bit loss in the determination of any region will be the sum of the bit losses over the nodes that must be "passed" to reach that region. The expected bit loss will then be the weighted average over the regions. Thus,

Expected Bit Loss

$$(.4566)(.125) + (.4566 + .4081)(.125) + (.4566 + .4081 + .0822)(.75)$$

which equals .87.

But note that, not surprisingly,

$$2.62 - 1.75 = .87$$

where 2.62 is the rank value of model three and 1.75 the rank value of the true model.

Thus the information distance between a model and its true representation is equal to the expected bit loss of the model. Because the incorrect model does not maximally exploit the binary choice, it must spend that much more to get the correct information.

There is another fruitful way with which to interpret this bit loss concept. Consider any pairwise comparison of a posteriori probabilities at one of the sampling nodes, and let the log of the ratio of the two probabilities be the measure of their divergence.

Thus let R and R' be the two sets of regions under consideration at a given sampling node, and let I be the value of the indicator. Then the divergence of the two a posteriori probabilities will be given as

$$\log (P(R/I)/P(R'/I))$$

which can be written as

$$\log (P(R/I)) - \log (P(R'/I))$$

but

$$\log P(R/I) = \log P(R) + \log P(I/R) - \log P(I)$$

and the same will be true for $\log P(R'/I)$.

Now when the correct or true model is used, and the tree is most efficient, the probabilities of the two branches at the given nodes are equal. (Each is equal to .5.) Thus $P(R) = P(R')$ and the measure of the divergence of the a posteriori probabilities becomes,

$$\log P(I/R) - \log P(I/R')$$

or simply the differences in the likelihood, the prior probabilities are eliminated. In effect, with correct binary division, we eliminate our a priori biases in the determination of the correct region. We rely solely on the likelihoods derived from the sampling information itself. We let our sampling information "do all the work", thus getting the maximum out of the sampling information.

§3. Transmission with Error

Thus far we have considered the performance of different models with respect to the cost of perfect information only. How will the different models perform when we assume some rate of error in the sampling process? Let us once again write down the different models and their rank values.

Region		A	B	C	D	Rank Value
Model	1	.25	.5	.125	.125	2.00
	2	.25	.125	.5	.125	2.37
	3	.125	.25	.125	.5	2.50
	4	.125	.125	.25	.5	2.62
true		.5	.25	.125	.25	1.75

Figure 17

Now for every model the following tree structure is used in the sampling process.

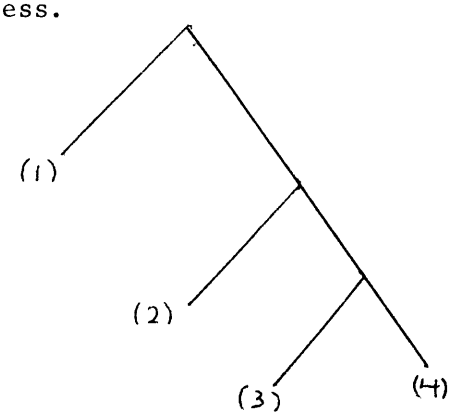


Figure 18

The numbers identify the different "branches" of the tree.

When the model is specified the regions are distributed along the tree by the rule of binary division.

Now turn to the tree structure of the true model and let the probability of making an error be P. Thus for example, if the region A is in fact about to order, then at the first node the probability that the firm will not choose the left branch of the first node, region A but rather the right branch will be P. However, once it has made the mistake and goes on to lower nodes, we assume that there is a .5 chance that it will choose either branch of these lower nodes (that is, once it has already made an error, the probability that it will choose one false alternative over another is .5). Note that the lower P, the higher will be the dollars per sample spent, where each node represents one sample. Error or P will vary between 0 or perfect information, and .5 or completely random information.

Now we can construct an "input-output" table that gives the probability that a specified input, or region, will yield a specified output or guess. This is shown in Fig. 19.[13]

Branch	1	2	3	4
1	1-p	.5p	.25p	.25p
2	p	$(1-p)^2$	.5(1-p)p	.5(1-p)p
3	p	p(1-p)	$(1-p)^3$	$p(1-p)^2$
4	p	p(1-p)	$p(1-p)^2$	$(1-p)^3$

Figure 19

[13] Note the distinction between this input output table and the one developed in Chapter Two, §2. There, because we assumed nothing about the firm's model of the field we could let error be distributed evenly over all guesses. Here because we assume a model, the error is distributed over the guesses in a structured way.

Thus, for example, the probability that the region on branch two will be mistaken for the region on branch one is p. Similarly, the probability that the region on branch two will be correctly identified will be $(1-p)^2$.

Thus to specify the input-output table for model one in Fig. 17, we multiply each row by the <u>true</u> probability of the region placed at the respective branch (as determined by model one). For example, for model one region B will be placed on the left-hand branch of the first node. Region B occurs with a true probability of .25. So the "weighted" probability of correctly identifying B when it in fact occurs will be .25(1-p). If p = .20, then .25(1-p) = .20.

We present the input-output table for model one p = .2.

	1	2	3	4
1	.2	.025	.0125	.0125
2	.1	.32	.04	.04
3	.025	.02	.016	.064
4	.025	.02	.064	.016
Sum	.350	.385	.1325	.1325

Figure 20

The sums over the columns represent the distribution over the output or guesses. Thus, the probability that in any sampling process the guess will be the region on branch one, in this case region B will be .35; or .35 of the time the firm will guess region B.

Now we can calculate the amount of transmission for model two with p = .2.

We have the input uncertainty, 1.75 bits, we can calculate the output uncertainty over the column sums, and we can calculate the joint uncertainty from the entries in the input-output matrix itself. Thus for this case we have

$$T = H(I) + H(O) - H(I,O)$$

$$.409 = 1.75 + 1.833 - 3.174$$

In other words the "expected" amount of transmission for the model will be .409 bits.

However, our analysis is not complete. The transmission of the expected number of bits requires a certain number of samples, and since each sample costs a certain amount, the more samples required to achieve a given level of transmission, the less efficient the model. So we like to know not the "expected transmission" but rather the "expected transmission per sample".

To calculate the expected number of samples, simply multiply the frequency of the output states, or guesses, as given in the column sums, by the number of samples required to reach that state or region. Thus it takes only one sample to reach region B on the tree of model one.

In our case the expected number of samples comes to 1.92. Thus we can say that model two sill transmit .409 bits using on the average

1.92 samples for an efficiency of .409/1.92 or .212 bits/sample. This last number is thus a measure of the efficiency of model one with p = .2.

The next table presents similar calculations for the four models and "true" representation -- here called model zero, for error, p = .2, .25, .33.

As can be seen from the last column, model efficiency falls as we go from models of low rank value to models of high rank value, and as we go to higher levels of error. In other words, the greater the information distance between a model and its "true representation" the less efficient it is at some specified level of error.

We graph these results on the next page. For a given p we arbitrarily choose some level of dollars per sample, assuming that "perfect" information is approached only asymptotically.[14]

[14]Note that this continues the assumption expressed in footnote 10 of this chapter. In other words the same dollar per sample generates the same error level for different models. To the extent that this is not true the above table will represent minimum differences in efficiency. This suggests an interesting way of distinguishing between modes of improving knowledge. Thus let us assume the higher the rank value of a model, the less divergent the a posteriori probabilities (i.e., its indicators are not as efficient) and thus the greater the chance that sampling variance will result in wrong choices. Thus the difference between the efficiency of the model and the true representation will be greater than that given in Fig. 20. Thus the firm can improve model efficiency by improving the indicators so that the difference approaches its minimum value. Further improvements require a change in the model itself. Now improving the indicators, so as to get more divergent a posteriori probabilities means getting a better understanding of the "field". However, as its understanding of the field increases sufficiently it should begin to develop a new and better model, i.e., one with lower rank value. Thus improvements in indicators, within the framework of a given model, is analogous to increases in quantitative knowledge, while changes in the model itself, in the structure of the model itself, is analogous to qualitative knowledge. Sufficient increases in quantitative knowledge, i.e., improvements in the indicators, should lead to increases in qualitative knowledge, i.e., an improvement in the structure of the knowledge. Increases in quantitative knowledge only can increase model efficiency to the minimum difference between itself and the true model.

p = .2

Model	RV	Transmission	Sample	Transmission/ Sample
0	1.75	.435	1.75	.248
1	2.00	.409	1.92	.212
2	2.37	.392	2.09	.187
3	2.50	.370	2.18	.170
4	2.62	.353	2.24	.157

p = .25

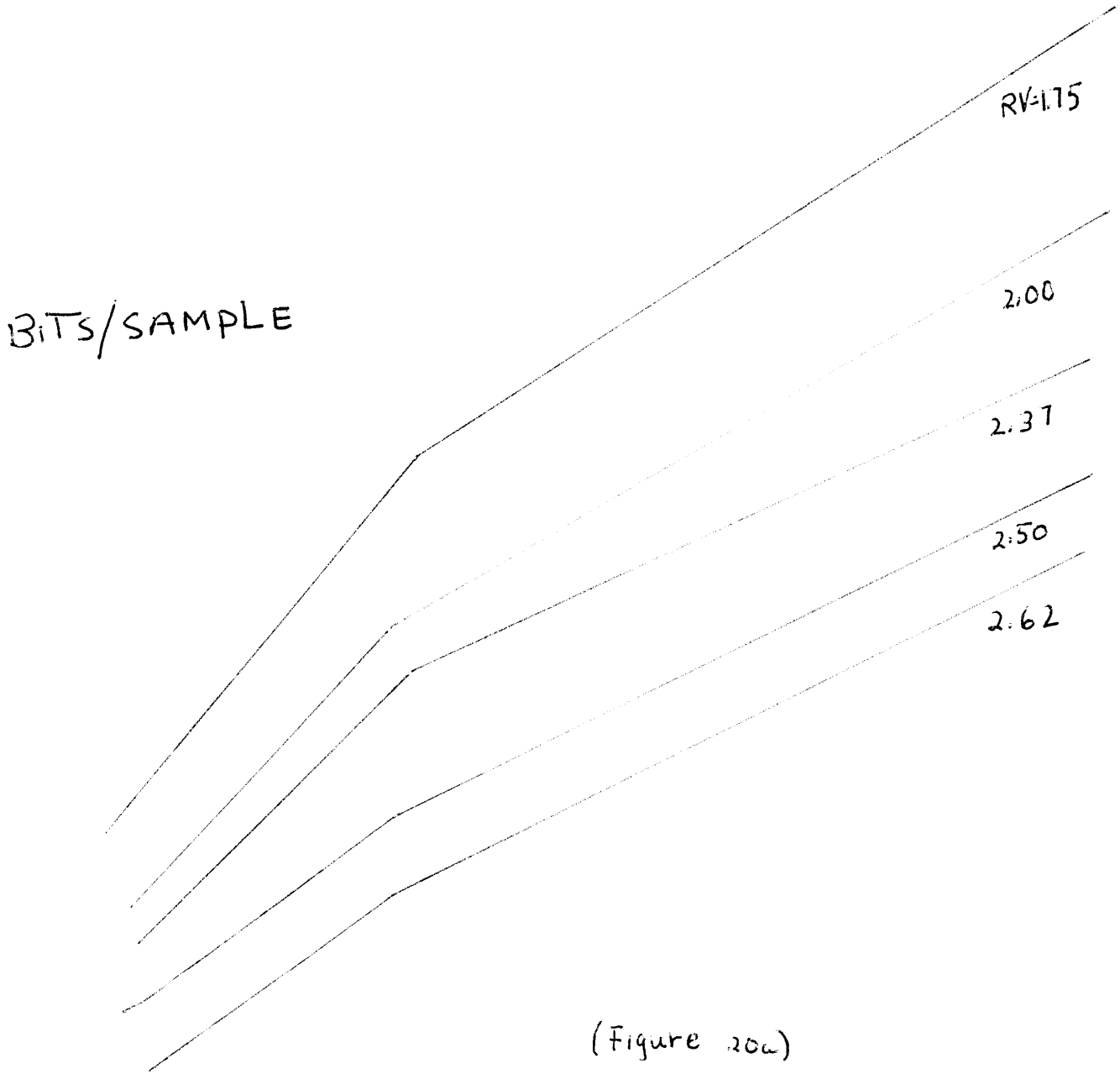
0		.294	1.75	.168
1		.268	.189	.141
2		.252	2.04	.124
3		.235	2.10	.112
4		.220	2.15	.102

p = .33

0		.120	1.75	.068
1		.109	1.85	.059
2		.104	1.93	.054
3		.009	1.98	.045
4		.008	2.00	.040

Figure 20

BITS/SAMPLE = f(MODEL; DOLLARS/SAMPLE)



(Figure 20a)

Now what will be the asymptotic value by each curve? Clearly, it must be the maximum number of transmitted bits divided by the number of samples required to reach that number, or 1.75/Rank Value. Ore more generally, for a field given of information value H, the maximum productivity of this model will be given by H/RV.

This is clearly the analogue of Shannon's channel capacity, which is an expression of the maximum transmission rate of a channel. The analogue of "rate" in our example is "sample" (which, as we shall see, translates into cost).

We list the channel capacities of the different models in Fig. 21.

Model	True	1	2	3	4
Chan. Cap.	1	.87	.74	.70	.66

Figure 21

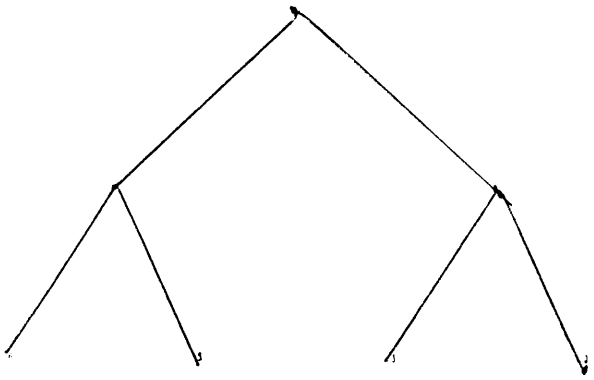
In other words, the lower the rank value, the closer the model is to the "truth" the greater the channel capacity. Getting "better" models is equivalent to building "channel capacity".

Now our graph suggests, but does not prove, that models with lower rank values will be everywhere more efficient than models with higher rank values, i.e., at no point do these curves cross. We can partially confirm this last proposition by calculating the transmission for the following model:

Region	A	B	C	D
Probability	.25	.25	.25	.25

Figure 22

The following tree structure is used to generate the intput-output matrix.



Proceeding as before, we find the transmission rates for the different errors.

Error	.20	.25	.33
Rate	.208	.133	.057
Rank Value	2.00		

Figure 24

Note that the rank value of this model is the same as that of model one. Thus this model and model one approach the same asymptotic value of transmission per sample. However, error has a greater effect

on this model, thus it lies everywhere below model one and every where above model two. In other words, the ranking of models by transmission rates at one level of error is unchanged when we move to a different level of error. This is consistent with our intuition that at no points should the curves cross.

§4. Structure and Rank Value

We are now in a position to begin to relate our concepts of transmission and structure to the material developed in the last section.

When structure is zero, then $H(x) + H(y) - H(x,y) = 0$, or $H(x) + H(y) = H(x,y)$. Since structure cannot be less than zero (simply a special case of the fact that transmission is non-negative), the model of zero structure is the model where the joint uncertainty is at a maximum.

Now we can prove the following theorem:

> "The rank value of the model of zero structure is equal to its joints uncertainty."

If we imagine that the true model is given by a matrix of dimension (I,J) where y takes on J states, for a given x, and x takes on I states for a given y, and if we let a_{hk} be the probability of the (h,k) state, then the probability of the (h,k) states of the model of zero structure will be given by:

$$a'_{hk} = \sum_{h=1}^{I} a_{hk} \sum_{k=1}^{J} a_{hk}$$

The rank value of this model will be given as

$$\sum_{h=1}^{I} \sum_{k=1}^{J} a_{hk} \log a'_{hk} \quad \text{or} \quad \sum^{I} \sum^{J} a_{hk} \log \sum^{I} a_{ik} \sum^{J} a_{hj}$$

or

$$\sum^{I} \sum^{J} a_{hk} \log \sum^{I} a_{ik} + \sum^{I} \sum^{J} a_{hk} \sum^{J} \log a_{hj}$$

But since the order of summation is irrelevant

$$\sum_{h}^{J}\sum_{k}^{I} a_{hk} \log \sum_{i}^{I} a_{ik} + \sum_{k}^{I}\sum_{h}^{J} a_{hk} \log \sum_{j}^{J} a_{hj}$$

Now the first term in this expression is the uncertainty of the marginal distribution over the columns or H(y), and the second is the marginal distribution over the rows or H(x). But we know that for the model of zero structure H(x) + H(y) = H(x,y), and since Rank Value = H(x) + H(y), the rank value of the model of zero structure is equal to its joint uncertainty.

Now for the model of complete structure rank value will also be equal to the joint uncertainty of the true model itself. But since we are confining our attention to the class of models where the marginal probabilities are constant, this means that the rank value of the model of complete structure is less than the rank value of the model of zero structure. In other words, if we let RV′ be the rank value of the model of zero structure, and RV the rank value of the model of complete structure we have

$$RV' = H(x) + H(y)$$

but

$$H(x) + H(y) > H(x,y)$$

but

$$H(x,y) = RV$$

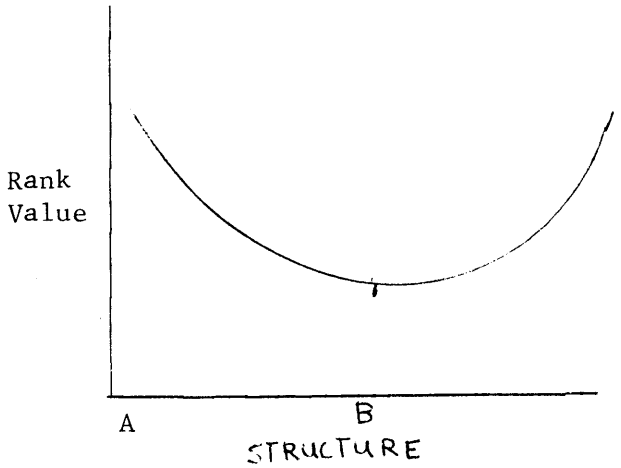
so

$$RV' > RV$$

This suggests that as we go from a model of zero structure to a model of complete structure, rank value valls, and as we saw in the last section, models of lower rank value have higher channel capacities. This is partial justification for our original intuition that increases in structure lead to more efficient models.

Nevertheless this is incomplete, since we would have to prove that increases in structure are monotonically associated with decreases in rank value. We know that over a certain range of models this will not be true. As we move through models possessing more than the true amount of structure, rank value will increase, i.e., attributing more interdependence to the variables than there actually is, is just as bad as not attributing enough. (It is in this sense that rank value is a much better measure than structure as a general index of "truth".

Our intuitive "picture" of the relationship between structure and rank value thus looks like Fig. 25.



We know that at point B, the point of complete or actual structure, rank value is at a minimum. (This follows from the basic coding theorem, i.e., correct binary division results in the more efficient code). Also, we have been able to derive the rank value of point A, the point of zero structure. Clearly our discussion in Part One referred only to models between points A and B. Thus we must prove monotonicity between these two points. Unfortunately, we have not been able to complete this proof, and present only a partial development here.[14]

Since we are considering the class of models where the marginals are preserved, any change in structure is a result of a change in joint uncertainty. Now since structure varies inversely with joint uncertainty, we want to prove that rank value varies directly with joint uncertainty reaching a maximum at the point of zero structure or maximum joint uncertainty. Now the joint uncertainty of any model between A and B can be written as

$$- \sum (P_i + e_i) \log (P_i + e_i)$$

where the P's represent the true probabilities and the e's the additions or subtractions to these probability values, to arrive at the specified model. Now this expression can in turn be written as

$$- \sum P \log (P + e) - \sum e \log (P + e)$$

---

[14] Rigorously speaking, structure is unidimensional in the two variable case only. That is only in this case will there be just two models that correspond to a given rank value, one that overstates and the other that understates the amount of structure in the model.

But the left-hand part of this expression is the rank value. So if we let $H^m(x,y)$ be the joint uncertainty of model m, we have

$$RV^m = H^m(x,y) + \sum e \log (P+e)$$

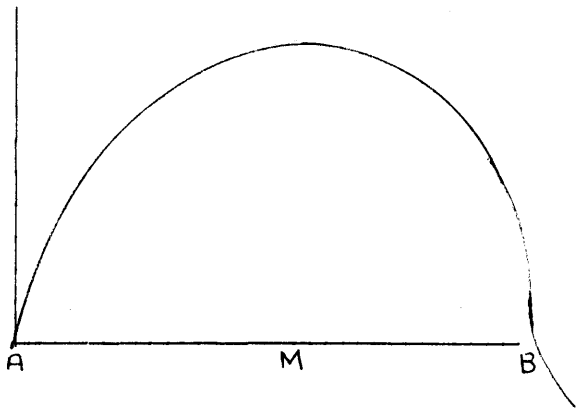
so that

$$dRV/dH(x,y) = 1 - d(-\sum e \log (P+e))/dH(x,y)$$

so that the latter term must be less than one between the points A and B if the graph of Fig. 25 is to be monotonic. I have not been able to prove this.[15] Note that this expression cannot be greater than one everywhere, since that would imply $dRV/dH(x,y) < 0$ every-

---

[15]Note that we know that at both points A and B $-\Sigma e \log (P+e)$ is zero; at point A because the additions and subtractions cancel each other out, and at point B because all the e's are zero. Thus, by Rolles theorem this expression reaches a maximum or minimum between points A and B. Call that point M. Then we would expect this expression to rise between points A and M, thus lowering RV, to fall between M and B but not as fast as the fall in H(x,y), thus slowing down the decline in RV. Finally, the decline in this expression brings RV to its minimum. Beyond B this expression falls more quickly than H(x,y), and very likely turns negative at point B. This intution is expressed in the following graph.



where, which is impossible since we have already located the two points A and B. So if there were a region where this expression were greater than one the derivative of the curve of Fig. 25 would have to change sign at least once, which intuitively seems implausible.

Let us summarize. We have seen that models can be ranked in information distance from the true model, and we saw that this ranking generates an efficiency of transmission measure such that models closer to the truth are more efficient. In addition, we tried to argue that the concept of structure can be related to our general measure of information distance such that models of greater structure are more efficient. However, as is clear, our model is not complete, specifically:

a) We did not analytically derive the relationship between transmission and rank value. Our graph was suggestive but not a precise proof.

b) We have not succeeded in determining the relationship between structure and rank value. We determined the relationship for two extreme values only.

c) We have employed discontinuous distribution only where exact binary division was possible.

In other words the "structure" of our model can be thought of as developing the following three links.

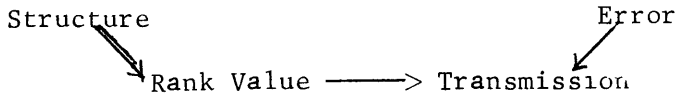


Figure 26

However the links are only partial. I suspect that the use of continuous distributions would complete these links and permit analytic proofs of our suggested propositions.

We can now easily establish the relationship between transmission rates, channel capacity and average cost. Thus we return to Fig. 21 which is reproduced in its essentials here.

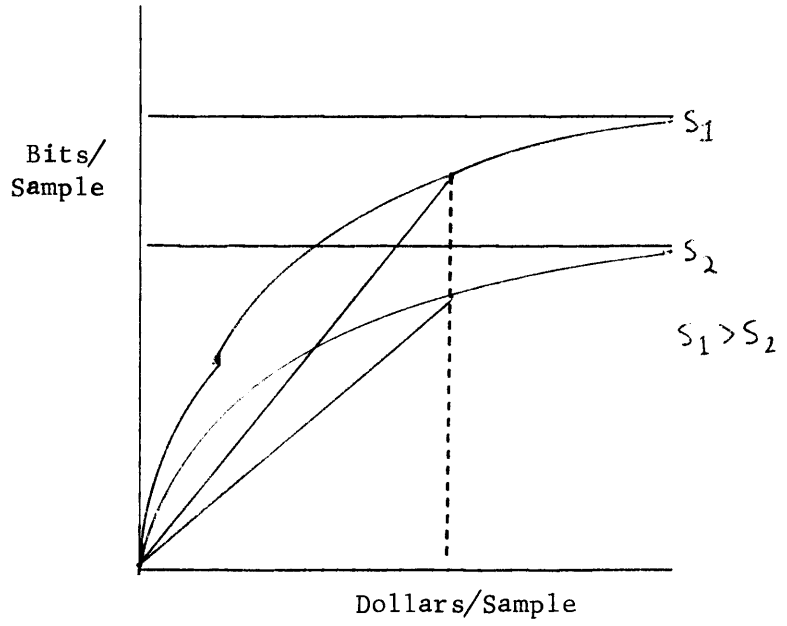


Figure 27

Take the ray from the origin to the graph. Its slope will be given as,

(Bits/Sample)(Sample/Dollar) = Bits/Dollar.

Since the slope of the ray falls as dollars/sample increases that means that bits/dollar falls or dollars/bit, the average cost of a bit rises. Thus, to approach the channel capacity of a model requires an increasing expenditure per bit of information. (This flows

from the assumption that perfect information can be approached only asymptotically.)

However, as is clear as the amount of information capital increases, i.e., as the amount of structure rises, the average cost of information falls for any specified level of dollar expenditure per sample. In other words, Fig. 27 translates into

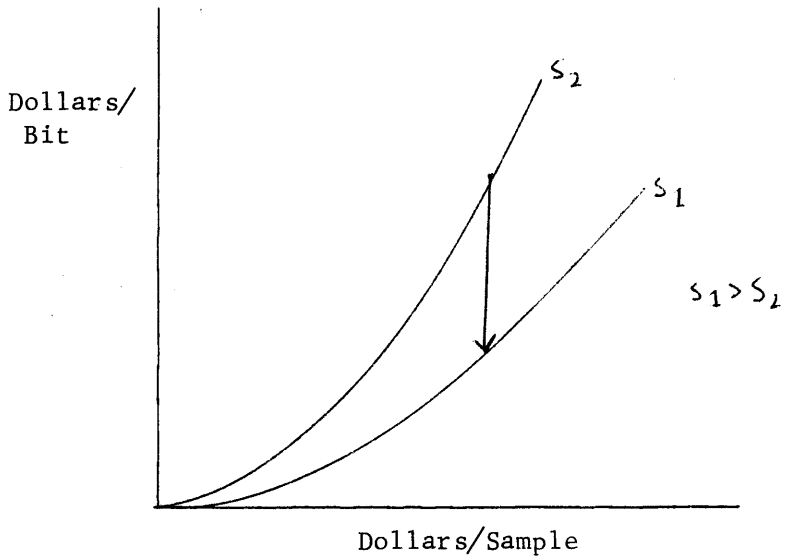


Figure 28

where the lower cost curves are associated with more information capital. In other words, increasing information capital raises channel capacity and lowers the average cost of information.

Summary

We have sought to give greater rigor to our original intuition that models that reveal more of the implicit patterns in the data, that expose their inherent complexity are more productive than "naive" models that do not capture this complexity. We then measured complexity with the information measure of the amount of structure between two variables. In turn, the amount of structure can be thought of as a measure of the amount of information capital in the model. The term "capital" is appropriate here since structure represents "congealed" information that determines the cost of processing information "flows". We saw that mathematically the measure of structure is a measure of the transmission between two variables. Thus, information capital represents congealed transmission that reduces the cost of flow transmission. Transmission within the model substitutes for transmission through the model. We found that we could measure the maximum productivity of models with a term analogous to Shannon's channel capacity. A model's channel capacity was given by the maximum bits per sampling dollar it could produce. This maximum was equal to H/RV, where H was the uncertainty of the field and RV was the rank value of the model. We saw that a given transmission rate curve was in turn associated with an average cost curve of information, and that to reach the point of channel capacity of a given model required an increasing expenditure per bit of information. However, we saw that increasing the amount of information capital lowered the average cost of information for any specified level of error. Thus building channel capacity was equivalent to lowering the average cost of information, a result consistent with our thoughts in Chapter Two.

We can assume that to build models of greater structure requires an investment of money for research and that the greater the amount of structure the greater the investment costs. If we term sampling costs as the cost of "information labor", then our results are succinctly stated as:

> "An increase in the ratio of information capital to information labor, increases the productivity of information labor, and lowers the average cost of information."

§5. Some Special Properties of Information Capital

Information capital does not share all the properties of physical capital. Its rate of depreciation or obsolescence, is not a function of its utilization but is rather a function of the rate of change in the field itself. Thus, if the model of the field is correct, its channel capacity will be one bit per sample. But as the field changes its rank value will rise and its channel capacity will fall.[15]

We can in fact argue that information capital is replenished rather than depreciated by use. Knowledge not used is soon forgotten. This is simply a special case for the general proposition that structure--organizational or informational--is replenished by use. Role differentiation is legitimized in the daily acts of role fulfillment. When roles are no longer followed, the role structure breaks down.

Unlike physical capital there are no fixed cost differentials associated with different amounts of information capital. We often associate a high fixed cost with a highly capital intensive production

---

[15]We can hypothesize that the rate of depreciation of information capital is a function of the stock of information capital in society. This follows from the notion that as the amount of information capital increases, the response time to new innovations and opportunities decreases, the horizon of the organizational unit widens, and as a consequence the environment becomes both more complex, and less stable. Thus, as the amount of information capital grows, the rate of change of the "field" and its complexity increases. Thus any given model has a shorter useful life as its channel capacity falls more quickly. Thus firms cannot for long rely on a given model but must anticipate new models, new configurations. Paradoxically, while information capital can reduce the level of risk for a given organizational unit with respect to a given field, the "composition" effect results in an over-all increase in uncertainty. In other words, it is a "fallacy of composition" to believe that over-all risk is reduced with the growth in society's stock of information capital.

process, attributing the high overhead to maintenance and the like. (The term fixed cost should not obscure the fact that overhead costs are flow costs, as opposed to investments which are one shot affairs.) Maintenance cost, however, are simply a variant of depreciation costs, and as we have already argued depreciation of information capital is independent of its use.[16]

The difference between the initial investment cost and reproduction cost is greater for information capital than the corresponding difference for physical capital.

There are always cost differences between the first addition of a new form of physical capital to the capital stock, and subsequent additions, due to presence of bugs, and gaps of knowledge. As these are overcome, reproduction cost, the cost of subsequent addition falls toward the resource costs of the new capital. Thus as the proportion of information investment costs to total investment costs rises, the greater will be the difference between the initial investment costs and the reproduction cost. Since information capital is "pure" information, this difference will be greatest with respect to information capital.

Finally, as many have pointed out, the consumption of information is not a zero sum process. A "bit" of information can be consumed many times without "disappearing". However, the private utility of information

---

[16]Not coincidentally, this is the analogue of the "theorem" stated in the appendix to Chapter Two, where we said that there are no fixed cost differentials to different organizational structures, so that the cost curve of one organizational structure lies everywhere below the cost curve of another.

depends on its distribution. The ability to profit from a piece of information is a function of who else does not have access to it.

The consequence of these particular properties of information, and information capital, is that information is very diffifult to "privatize". There is an inherent "publicness" to information. Depreciation through use, high fixed costs, high reproduction costs, and zero sum consumption, are all scarcity limits to the full and equitable distribution of physical capital. All economic systems must devise a mode of distribution for physical capital, but these natural scarcity characteristics of physical capital tend to support any particular system of unequitable distribution. The distribution of information capital, however, is not constrained by these limits.

However, we know that the private utility of information is a direct function of its maldistribution. Similarly, as we saw in chapter two, full access to the firm's information capital would imply more democratic decision making. Thus, at the same time that the private appropriation of information capital becomes necessary for private gain and power, so does the natural characteristics of information push in the direction of a full and democratic access to information capital. This is perhaps in its "purest" form an example of Marx s contradiction between the social nature of production and its private appropriation.

Equally important, the maldistribution of information capital limits its productivity. Information disseminated widely can lead to

---

[16]Not coincidentally, this is the analogue of the "theorem" stated in the appendix to Chapter Two, where we said that there are no fixed cost differentials to different organizational structures, so that the cost curve of one organizational structure lies everywhere below the cost curve of another.

to more information, more insights, a greater number of"new combinations" than if its dissemination were restricted. Thus as in Marx, private forms of appropriate are fetters on the development and growing productivity of information capital. We can thus expect the conflicts around the appropriation of information capital to be particularly severe.

INFORMATION CAPITAL

from the Perspective of Conflict

Chapter Four

ON GROWTH AND CONFLICT

§1. It is perhaps one of the unwritten assumptions of economic theory, that growth can ameliorate economic conflict. The distribution of income is a critical parameter only when resources are fixed. Conflict will then be strictly "zero sum" and for every loser there will be a gainer. Growth, however, so the argument goes, is a "positive sum game" where, if the institutional rules are reasonably constructed, everybody is the winner. This proposition is reasonable, however, only if one focuses on the growing income stream, to the exclusion of the resultant distribution of capital. After all, capital broadly defined, consists of an enforceable claim on this income stream, whether it be in the form of the ownership of physical capital, "tenure" arrangements, or access to certain decision making processes. It is the economic correlate of "power" and "status" the ability to insure a continuous stream of benefits through the exercise of certain economically or socially sanctioned prerogatives. Growth that entailed the redistribution of such prerogatives, the destruction of old and the creation of new claims, would contain the seeds of conflict. Economic theory, however, sheds little light on these "capital effects" of economic growth.

Two of our more subtle economic historians, J.A. Schumpeter and Karl Polanyi, saw "capital destruction" as an intimate part of the

process of economic growth. To Schumpeter, growth was a process of "creative destruction".[1] The creation of new products, new processes, and new forms of economic organization, meant the destruction of old "capital values" of old structures and organizations. Growth meant the restructuring of old modes of production and consumption, and the demise of those organizations that could not withstand the "gale of perrenial destruction". To Schumpeter this process was the only meaningful form of economic "competition". The continuous fluctuation of prices about their long term equilibrium values were of only marginal significance to the competitive process. What really counted were the technological changes, the product innovations, the introduction of new organizing and coordinating principles, the traditionally "exogeneous factors" of neoclassical thinking.

Polanyi extended this concept of "destructive" growth.[2] In his analysis it lies at the heart of the strictly utopian idea of the perfect and unregulated market. Unfettered mobility in the market for "labor" (the fictitious name for man) and perfect markets for "land" (the fictitious name for nature) could mean only the continuous upheaval if not destruction of all forms of social community and social status. The idea of the perfect market could mean only the dissolution of society. It was the confrontation with this idea, and its partial realization in the initial stages of industrialization that in Polanyi's view, gave rise throughout Europe to the spontaneous demands for protection from the unregulated market.

---

[1] Joseph A. Schumpeter, Capitalism Socialism and Democracy, Harper and Row, New York, 1962.

[2] Karl Polanyi, The Great Transformation, Beacon Press, 1957.

Curiously enough, Schumpeter also discovered an institutional response to the "excessively" destructive winds of economic growth. But unlike Polanyi he located it not in the political sphere, but in the very "monopolistic" practices regarded as so inimical to competition by the classicisits and neo-classicists. The various "restraints" on trade, the inflexibility of prices were ways of insuring a degree of stability in the vortex of market fluctuations. The demise of a firm meant the destruction of society's "organizational capital". The productivity of the new had to be insured before the old could be allowed to disappear. Restraints on trade were thus a way of insuring both "orderly advances and retreats".[3] In Schumpeter's view they served to cushion the process of transition, and minimize the costs of dislocation.

Thus, in the schemes of both these theorists, growth entailed the destruction of old physical, organizational, and social capital. The resistance to this destruction, to change and the rate of change, could crystallize in political conflict. However, some analytical and empirical reflection will show that not all growth can be classified as "capital destroying". In fact, we can fruitfully distinguish between three modes of growth, capital preserving, capital augmenting, and capital destroying growth. In turn, this taxonomy will help us analyze the impact of the emergence of Information Capital.

Capital preserving growth can be associated with a growth in the scale of economic operations, with the composition of output and the

---

[3]Joseph A. Schumpeter, op. cit., p. 90.

structure of capital (both physical and organizational) constant. It can be empirically associated with the initial stages of growth of an economy or urban region, and resembles most closely the metaphor of constant returns to scale that lies at the heart of such neo-classical theory.

Capital augmenting growth is associated with the process of structural differentiation, or the specialization of economic activity. We call it capital augmenting because it creates new centers of economic activity that are complementary rather than competitive with the old. In the parlance, this is associated with "external economies", as the addition of new activity lowers the cost curves of the old. Urban economists have found this process of structural differentiation to be central to the process of urban development.[4] It is capital augmenting because it creates new physical and organizational capital while simultaneously raising the present value of the old.

Finally, capital destroying growth is associated with Schumpeterian competition, in which new capital is competitive with the old. In the process of raising over-all output and productivity, such competition results in a redistribution of capital claims. In addition, it can be fruitfully associated with the process of the integration of economic activity, as backward and forward linkages (i.e., suppliers and buyers) once coordinated through the market are coordinated through the firm.

---

[4]Edgar Hoover and Raymond Vernon, The Anatomy of a Metropolis, Doubleday and Co., New York, 1962.

It should be clear now that the taxonomy developed here parallels the taxonomy developed in our schema of organizational growth. The process of structural differentiation corresponds to the process of greater specialization of organizational activity. We saw that in the "tree model" or organizational structure, an increase in the degree of specialization with span fixed resulted in an increase in the number of decision amking and processing nodes (i.e., $dM/dN > 0$). As in structural differentiation within the economy, new loci of decision making power, of organizational prerogatives, were created.

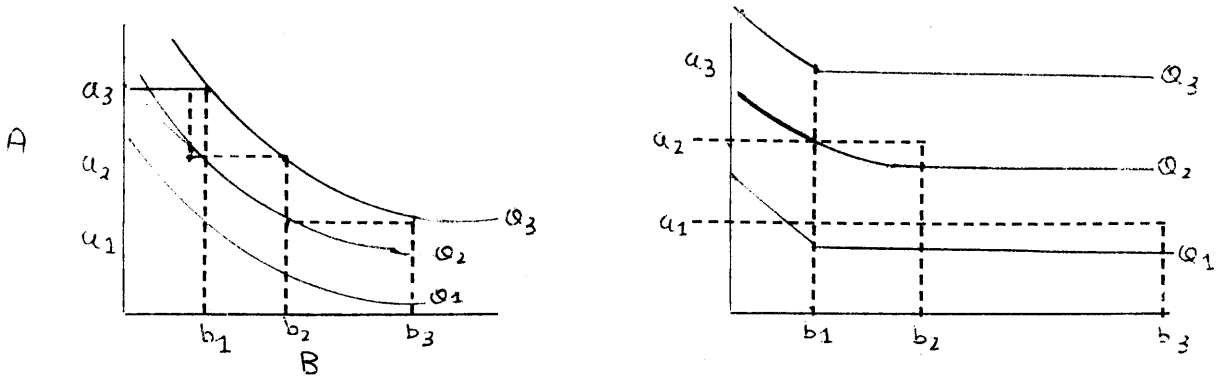
Similarly, the emergence of information capital as our "final stage" of organizational growth corresponds with the process of Schumpeterian competition. Both involve the destruction of capital. As we saw earlier, the emergence of information capital could result in either a centralization or decentralization of decision making power. But in either case, it involved the restructuring of organizational activity, and a redistribution of the "organizational capital".

This parallelism between our two taxonomies lends to a more basic insight, for there is a more fundamental symmetry between Schumpeterian competition and information capital. Schumpeterian innovation entails the discovery and utilization of new information. Whether it be a new product, or process, the entrepreneur must integrate hitherto compartmentalized markets or materials. He must discover patterns that transcend the given structure of production and consumption.

In effect, the entrepreneur embodies information capital. But of course in the Schumpeterian scheme the "number" of entrepreneurs is exogenously given. Schumpeter did not suggest that the ability to imagine new economic patterns and organizational structures could be technologically determined. But as we have already argued the ability to comprehend new patterns and structure information on an ever widening scale is in fact a function of the information processing and modelling technologies. This suggests that as these technologies improve the societal capacity for "entrepreneurial activity" increases. But we suggests here that as these technologies improve, the entrepreneurial capability, the ability to imagine new economic and social structures (in a non-utopian and empirical fashion) is no longer located strictly within the business - market sector. Rather, it is found increasingly in the education, government, and other sectors where the skill for appropriating information improves along with the development of society's information technologies. Thus an increasing proportion of the population lives within the nexus of "future images". This accounts in part for the psychological sensation of an increasing "rate of societal change", and the demands for societal "restructuring". More importantly, where the "entrepreneurial" capability does not coincide with financial or political power, we can expect intensifying political conflicts over the rate and direction of change. But we get ahead of our story. We want to turn now to a more detailed analysis of the dynamics of conflict surrounding the emergence of information capital. Once again we utilize an organizational perspective to inform our analysis.

§2. We have suggested that the defined set of prerogatives and tasks of a unit within an organization constitutes its "package" of organizational capital. The performance of such tasks and the exercise of such prerogatives insured a continuous stream of benefits. The unit's position within the organization defined the quality of its organizational capital. What however "protects" the unit? What insures that its claim to organizational output will be preserved? Its claim is in part insured to the extent that it can exercise its "power" within the organization. In turn such power rests on the fact that it monopolizes certain organizational resources, because the organization is in fact based on a series of task specializations. Structural differentiation means that tasks are performed by specialized units. Thus, for example, each unit will process some subset of the organization's information flow that is inaccessible to the other units. As a consequence, the tasks of one unit cannot substitute for the tasks of another in the production of organizational output. Specialization implies a range of "fixed coefficients" in the organizational "production function". Power based in the structure of organization derives in the end from the nature and degree of task specializations.

We can show this relationship between power and "fixed coefficients" through the use of the traditional isoquant diagram. In Fig. 1 we let each axis represent the "amount" of input from each



1a 1b

Figure 1

unit. The isoquants represent the amount of organizational output. Figure 1a is the case of complete substitutability, while Fig. 1b is the case of a high degree of specialization in A. Assume now that the "goal" of unit B is to keep output at Q_3. Let the organization begin at output A_3B_1, and now let A draw down its input to A_2, and then to A_1. In Fig. 1a B can always respond to a reduction in A's input so as to keep output at Q_3 (i.e., by moving first to B_2 and then to B_3). Thus for each move of A, B has a response that keeps output at Q_3. But in Fig. 1b, A clearly controls the outcome. At A_2 output must fall to Q_2 while at A_1 output must fall to Q_1.[5] B has no responses to A's moves and thus A controls the outcome. (Note that we have not been considering the implicit "costs" of response. We assume that the goal is the maintenance of a level of output irrespective of the cost of that output.)

---

[5] We are assuming no "intervening" outputs between Q_3 and Q_2.

We can summarize this "game" in the following tabular form.

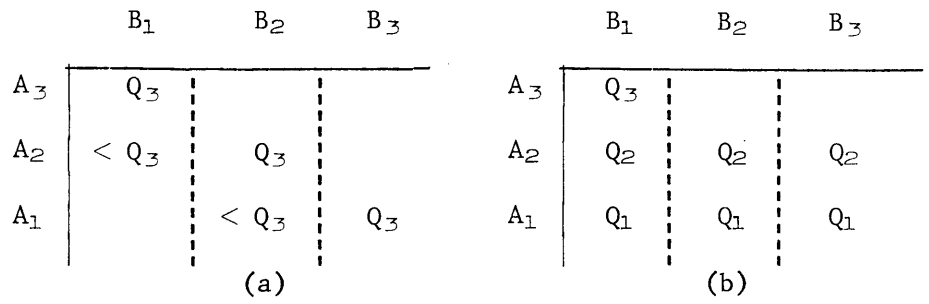


Figure 2

As is clear from the tabular form, B controls the outcome in the first case, while A controls it in the second. Note that in

there is no distinction between B_2 and B_3; they are equivalent to the same response. In other words, as the degree of substitutability falls, the "variety" of B's distinct responses falls. This corresponds with the cybernetic notion of "control" or "regulation". Ashby shows, in his Introduction to Cybernetics[6], that the number of possible outcomes in a "game" of move and counter move will be greater than or equal to (Variety of A/Variety of B) where the number of a unit's possible distinct moves is its variety. If we measure this in log to the base two, we then get that the lower limit of the information or uncertainty value of outcomes will be given by $\log A' - \log B'$, where A' and B' are the variety of A and B, respectively (i.e., the number of distinct responses). Thus, if A' is fixed, the lower limit to the uncertainty of the outcome will be

[6]W. Ross Ashby, An Introduction to Cybernetics, John Wiley & Sons, New York, 1963, Chapter 11.

determined by B. Thus we can say (though a more rigorous model would have to be constructed to prove this statement) that the less the degree of substitutability the greater the uncertainty of the outcome, the less power does B have over A. This is the sense in which specialization, the monopolization of organizational resources, confers power on the organizational units.

This concept of a range of non-substitutability suggests that there will be "threshold" effects in conflicts over the distribution of income, i.e., there will be limits to A's and B's share of organizational income. We can demonstrate this concept by modifying Herbert Simon's model of the "employment relation".[7]

Let B be the employer, A the employee, and let the X axis represent A's "effort input" per unit of time. Assume the stock of physical capital is constant. As effort input per unit of time (E/T) increases, the organizational output rises. On the Y axis we place the wage rate

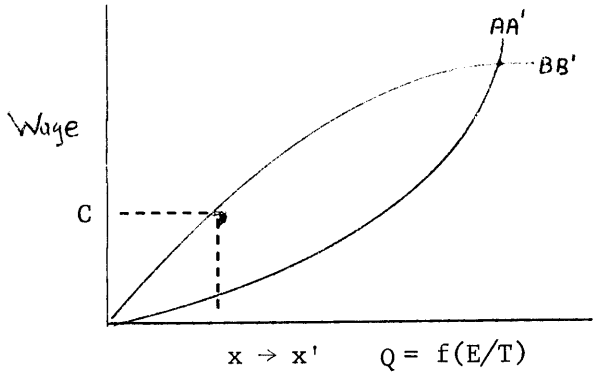


Figure 3

of A. Then we say that A has a minimal acceptance curve AA' and B

[7]Herbert Simon, Models of Man, John Wiley and Sons, New York, 1957. See "A Formal Theory of the Employment Relation".

has a maximal offer curve BB'. The position of these curves will be determined by some of the traditional market factors (i.e., market scarcities) and by the relative power of A and B within the organization. Thus, for example, at any given level of output A will determine what damage (as in a strike) he can inflict on B without generating too great a cost to himself. Similarly B will calculate how expensive the cost of buffering against actions by A will be. Such calculations will generate the offer and acceptance curves. Initially these curves will represent subjective evaluations of power, but over time and with experience such curves will come to represent the objective correlates of relative power.[8] As can be seen, conflict over the distribution of income will take place within the area bounded by the two curves. This zone of conflict is the analogue of the uncertainty measure $\log A' - \log B'$. If B cannot control the outcome then he cannot "dictate" the wage offer to A. Variety in the set of outcomes means a zone of conflict over the distribution of income.

Now let the organization begin at output X and at wage rate C, and let output rise to X'. Since the condition of employment has changed (i.e., E/T has risen), there will presumably be a new wage level. But so long as the structure of the organization remains unchanged, neither the acceptance nor offer curve will shift positions. Thus even though the wage rate changes, the zone of conflict is stable.

---

[8]Note that in contrast to our previous model of power, we assume that the goal of the two units also includes the cost of sustaining a certain outcome.

In short, conflicts over the distribution of income are stable long as there is no threat to the structure of the organization, to the distribution of organizational capital.

It should be clear now that the emergence of information capital introduces a new level of conflict. It results not in a struggle over output, but in a struggle over the claims to output, over the zone of conflict itself. Struggle over the claims to output introduces a new level of uncertainty to organizational conflict. As we have just seen, a given conflict zone corresponds to the uncertainty value with respect to the possible outcomes. But as we saw the emergence of information capital entails a change in the structure of organization and a redistribution of power. Thus in the course of transition from the old to the new structure there are no bounds to the conflict, there is no uncertainty value attached to the possible set of outcomes. The organization finds itself operating in a region of "meta-uncertainty". (This distinction precisely parallels Knight's distinction between uncertainty and risk.) It is this new level of uncertainty, the conflict over the zone of conflict, that distinguishes conflict surrounding the emergence of information capital, from conflict over the distribution of income.

As a consequence, income per se is no longer a reliable guide as to "enemies and allies". The emergence of information capital can "twist" what might appear to be the "natural" set of alliances. Thus when an information system restructures a medical service in a

hospital, administrators, lower level clerks, and senior doctors, all stand to lose, while "para-professionals", young doctors, and urban blacks (who will gain from better service) all stand to win. Thus the redistribution of organizational capital, as opposed to organizational income, generates new alliances. The emergence of information capital means a "new politics".

The conflict surrounding the emergence of information capital can be particularly severe because the generalized "uncertainty", the absence of bounds to the conflict, renders strict material and utilitarian considerations inoperative. Michael Rose, in his Computers, Managers and Society, finds that fears of redundancy, uncertainty with respect to job ladders and structures, and the threatened loss in status brings a seemingly "irrational" dimension to the problem of introducing computers into offices.[9] The problem becomes a political rather than a technical one. Rose notes:

> It is worth noting that this fear of redundancy is not simply that of the loss of a convenient well paid or previously promising appointment, nor the worry of finding another to replace it, with the accompanying possibility of a period of unemployment, the need to move house perhaps, and other dislocations. The indignity of the severance itself, its assault upon individual feelings of self esteem, even sometimes of indispensability, is no less repellant than these mundane practilities.

In a different context, Polanyi deals with this same "economistic prejudice". He attacks economic historians for focusing simply on income data in their estimation of the impact of the industrial

---

[9]Michael Rose, Computers, Managers and Society, Penguin Books, Baltimore, 1969, p. 157.

revolution on the English peasant.

> Actually, of course, a social calamity is primarily cultural and not an economic phenomena that can be measured by income figures, or economic statistics. Not economic exploitation, as often assumed, but the disintegration of the cultural environment of the victim (English peasant) is then the cause of degradation.[10]

In short, information capital means the dissolution of old structures. It is the consequent, threatened loss of power, fears of redundancy, and generalized uncertainty that affects all who stand to lose with "restructuring" that brings a new qualitative dimension to organizational and, by inference, more general social conflict. The emergence of information capital means that conflict no longer centers around the distribution of income, but rather around the distribution of organizational capital within a context of growth and change.

---

[10]Karl Polanyi, op. cit., p. 157.

ADDENDUM

We have suggested that the conflict surrounding the transition period can be particularly "severe". We can show this somewhat more rigorously through the use of some simple game theoretic considerations.

Thus let A be the employee and B the employer. Assume that B can decide to restructure or not restructure the organization. A is aware of this possibility and can either toughen its stance, i.e., bring its power to more fully bear on organizational operations, or behave as in the past. Thus we say that A can engage in either cooperative or non-cooperative behavior. Let us list in the squares present value of the proportion of organizational output going to A in the event of different combinations of behavior.

		B	
		Restructure (R)	No Restructure (NR)
A	Cooperative (C)	.2	.3
	Non-Cooperative (NC)	.4	.5

Square (C,NR) represent the present value of A's organizational capital before any "innovation". Square (NC,NR) represent a "tougher" stance on the part of B with the same organizational capital (i.e., without any improvement). A can achieve the highest share in the case of (NC,NR), since he exercises his power more fully within the old structure; but he achieves the lower share at (C,NR) where he cooperates in the loss of his organizational capital. In other words, he can raise his share from .2 to .3 by not cooperating

with the restructuring process. In effect, he resists the innovation, postpones the diminuition of his own power, and thus raise the present value of his organizational capital.

Now, if A follows a maximin strategy, i.e., he looks for the best of the worst, he will chooise point NC. If B pursues a minimax strategy, i.e., he minimizes his loss under the worst conditions, he will choose point R. The game will then settle down at (NC,R) or restructuring with non-cooperation, as opposed to (C,R) which would have implied a less severe transition.

The numbers chosen here are, of course, somewhat arbitrary, but the example does bring out the way in which under the impact of information capital, there will be a tendency on the part of potential lowers to "milk" the present organizational capital dry". (Our use of the employer-employee analogy should not obscure the possibility that the potential losers can be those who are presently in power.)

INFORMATION CAPITAL

from a Radical Perspective

Chapter Five

§1. The Demise of the Market

"The demise of the market" figures centrally in J.K. Galbraith's latest book, The New Industrial State.[1] Drawing on Berle and Means' classic, The Modern Corporation and Private Property[2], and somewhat unconsciously, on Schumpeter and Polanyi, Galbraith finds that price fluctuations no longer determine the allocation of resources. Market forces take a back seat to corporate planning as price elasticity drops precipitously to zero.

The economics profession was singularly unimpressed. Everywhere you look there are markets. After all commodity brokers match supply and demand daily, list price is always shaded a bit by the dealer when autors are moving slowly out of the show room, and "in the long run" market price is a good measure of resource cost. In the end "the demise of the market" seemed like too weighty a proposition to depend on the stability or instability of a few key industrial prices. But the concept persists with an underground life of its own. Some of the more stubborn social critics cannot help but imagine that since the Russian Revolution, The New Deal, the Keynesian victory, the Marshall Plan, the rise of the multi-national corporation, the growth in the public sector, etc., modern capitalism must differ in some way from its nineteenth century counterpart. But how?

It is our hypothesis here that the fall of markets is best understood when we view them as social information systems. Their "demise"

---

[1]John K. Galbraith, The New Industrial State, Houghton Mifflin, Boston, 1967.

[2]Adolph A. Berle, Gardiner C. Means, The Modern Corporation and Private Property, Harcourt Brace & World, New York, 1968.

is contingent on the rise of a consciously designed "information infrastructure" composed of societal information systems that suggest and permit new and anti-market modes of societal coordination and resource allocation.

Markets as information systems are of course the time honored explanation for their persistence in industrial society. The number of decisions required for resource allocation were deemed too numerous for society's limited information processing capabilities.[3] Allocation was achieved through the profit incentive as resources moved to their point of highest return. Exchange relationships could substitute for the transmission of information, since each actor needed only a limited amount of information for his own maximizing activities. Through the profit incentive, the invisible hand insured that these pieces of fragmented information would be integrated into a rational whole. No comprehensive knowledge was required for comprehensive coordination.

The theorists went further. Not only did they insist on the coordinating capability of markets, but they sought to prove that such coordination was in fact "optimal". Elaborating on an essentially moral problem in the Wealth of Nations, these theorists sought to prove the consistency of private selfishness with social selflessness. The profit motive meant not only complete coordination, but the "best" allocation of human and material resources.

---

[3]See Chapter One, "The Market as an Information System".

Now there can be no denying that markets can, in fact, coordinate economic activities. Two centuries of industrial growth testify to their power and resilience. But no theoretical or empirical evidence has proven that such coordination is "optimal". Whatever the generality we wish to accord to "static" models of optimal allocation (and they are by the admission of their authors sorely deficient), they clearly "miss the point". Industrial society is a society of change, of the elaboration of science, technology and social form, of changing "initial conditions". There can be no proof that market allocation over time is optimal since information about the future is always uncertain. Whatever the merits of decentralized decision making in a static context, decentralized investment decisions possess hidden and nasty complementarities. The invisible hand is blind to "externalities" between the past and the future.

However, if we take a different more historical perspective, the intution of the theorists seems essentially correct. In an historical perspective, markets, irrespective of their optimality properties, proved to be indispensible social instruments for allocation over time for the coordination of social change. The mobility of resources, an essential precondition for all industrial growth, means in the end the mobility of men, and the dissolution of social structure. By installing the "pecuniary" motive, markets weakened non-economic and semi-feudal formations that depended on a static configuration of technology and resources. They insured the minimal adaptability required for the successful diffusion of innovations.

However, the pecuniary motive could not suffice alone. Contrary to the simplistic litanies of traditional economic liberalism, exchange relationships are not always beneficial to both parties. When placed in a context of economic change, they involve the destruction of old organizational and social capital.[4] Adaptability meant subjecting oneself to the continuous upheaval and "revolutionary" impact of the capitalist dynamic, with its consequent destructive effects on the old social and economic structures. It is here that fragmented information, the consequence of limited information processing capabilities, played a crucial role. To the extent that markets were grounded in limited awareness, their mandates had the quality of "historical force". The invisible hand stood above human action and political will. To resist market forces was to resist the progressive force of history itself. Thus fragmented information provided an "ideology" of markets that insured its continued viability despite its destructive impact. More important, where a comprehensive awareness was absent, where the allocation of resources appeared as part of a grander more metaphysical scheme inaccessible in its details to human comprehension, resistance would prove all that more difficult. Nor, from a historical perspective, would resistance have proven beneficial. The industrial revolution in all its destructiveness was a necessary condition for the development of human freedom. Markets based on a fragmented awareness could carry through with this destructive process.

---

[4]See Chapter Four, Information Capital from the Perspective of Conflict.

Thus, irrespective of proofs of "optimality", markets were essential instruments for allocation "over time". But it is essential to see that the very justification for markets, limited information processing capabilities, was at the same time the necessary condition for their continued viability as instruments for social and economic change.

To be sure, no society has in fact been characterized by such a high degree of "information fragmentation" so as to be dominated in total by market "forces". The strictly "political" sphere was autonomous enough to "regulate" the market, and much of the political conflict of the nineteenth century was an exercise in limiting the more destructive aspects of market allocation.[5] It was in fact the long-term program of Marx's Critique of Political Economy to expose the social relationships behind the seemingly autonomous exchange patterns, so that market forces would be not only resisted, but ultimately transcended, and the means of production taken up and guided consciously. But we suggest here that the growth in society's information infrastructure, in the form of electronic media, government and corporate information systems, and the entire university complex, constitutes a "quantum jump" if not the culmination of the attack and destruction of "market forces". The growth in the information infrastructure means that comprehensive information and awareness become technologically feasible. As a consequence, both the ideology of markets and the grounds for their continued viability are shattered.

---

[5]For a full elaboration of this theme see, Karl Polanyi, The Great Transformation, Beacon Press, Boston, 1957. The writer's debt to Polanyi should be evident throughout this essay.

We can understand the impact of the electronic media and the information technologiesm by analyzing three inter-related developments. The growth in the information infrastructure "socializes" the means of production, encapsulates the future into the present, and leads to demands for democratic access to decision-making processes.

Marx pointed out long ago that the means of production in industrial society become increasingly social, that the distinction between the private and social becomes increasingly blurred. Economists would have to agree, since in anticipating the integration of markets, they have expressed the same idea. As mobility increases, as factor and product substitutability increases, the ramifications of private decisions spread further and faster throughout the economy. The private decision has increasingly social consequences. Information systems complete this process by providing comprehensive information that exposes these interdependencies. Different transportation technologies can be evaluated for their impact on racial residential patterns, while the reading capabilities of lower class children can be studied as a function of the impact of food processing on nutrition. Where the means of production were in fact social, they are now <u>experienced</u> as social. More important, the information technologies will reveal not only given interdependencies but create new ones. "The medium is the message", or in the economist's parlance, information technologies create "consumption externalities". Geographically diverse

groups, once compartmentalized, now interact "through" the societal information systems. The television in the Indian village means that western consumption patterns are visible to the peasant, and it thus transforms the aspiration level and mobility patterns of the urbanizing Asian masses. Thus by revealing given interdependencies and creating new ones the distinction between public and private is eliminated.

To be sure the concept of "inter-relatedness" is not a new one. John Donne knew that "no man is an island". What is new today is that this idea is not arrived at through the exercise of an apocalyptic or poetic imagination, but rather grows out of daily experience. Whether it be through the media, education, or government information systems, interdependencies are revealed in their full empirical complexity, to an increasingly broad sector of the population. Interdependence becomes a working proposition based on the "facts and figures".

It is this full empirical awareness that makes the future appear increasingly malleable. Just as the private becomes the public, so is the future collapsed into the present. Where interdependencies are understood the future consequences of present decisions can be grasped. The ecological consequences of forms of energy utilization can be studied and anticipated. Comprehensive information can expose the complementarity of investments decision that were hidden by market forms of coordination. In effect, "optimal" allocation between present and future seems increasingly possible. The growth in the "futuristic"

sensibility develops concomitantly with the technological capability for planning the future. The rational allocation of investments is no longer a "Utopian" prospect. The information technologies promise to provide truly optimal coordination precisely in the area where markets can only fail.

In effect, society in its present and future forms is experienced socially, and private decision-making prerogatives based on private property, or governmental secrecy appear increasingly unviable and illegitimate. The investment decision of a single entrepreneur is no longer his prerogative since it has perceivable social consequences. As a consequence, there develops in society a demand for democratic decision making based on full access to comprehensive information. As society is experienced socially, there develops a demand to understand it in its full social dimension. A comprehensive understanding, an ability to structure and pattern the increasing flow of information, becomes a necessary prerequisite for democratic decision making. Here then is the key to the new dimensions of social conflict in the age of information and communication. The industrial class antagonisms were based on the distribution of income within the interplay of market forces, but "post-industrial" conflict is based on the distribution of information, on access to comprehensive understanding, within a framework of societal planning,of a consciously elaborated human future.

However, it is important to see that the growth in this planning framework itself is not and has not been a result of consciously-willed political decisions. Rather, it resembles more an "unconscious" social revolution whose tempo has increased since World War Two. Throughout the "capitalist" world, governmental decision-making has increasingly guided economic growth and change.[6] Even in America, the last bastion of "free enterprise", the growth of the public sector, the rise of Keynesian policy, and governmental funding of much research and expenditure have increasingly politicized economic change, and delimited the private market. The growth in the new information and communication technologies gives this process an even greater momentum. As the technological capabilities for societal planning and coordination increase, capitalism must willy nilly unravel itself. The agents of this "revolution" need not be and most likely will not be conscious political agents but rather the force of the new technologies itself. Within the tumul of undirected and even less understood political conflict, these technological capabilities will be taken up and new modes of coordination will develop.

But such a process need not appear in a "liberating" context. It need not advance man's freedom. As the Russian experience all to well demonstrates, planned allocation of material and human resources can mean in the end the totalitarian command of such resources.

It is here that the conflict is joined. The conflicts over the new technologies of communication, learning and information become

---

[6]For a complete discussion of this development, see Andrew Shonfield, Modern Capitalism, Oxford University Press, London, 1965.

conflicts over their mode of development within the overriding framework of a consciously designed human future. The demand for democratic decision-making and the utilization of the new technologies for such purposes stand against the powerful totalitarian thrust implicit in conscious control and design.

The demise of the market means not the disappearance of all markets but rather the end of the market as the arbiter of social change. Markets no longer are the key to man's "social reproduction" and the industrial class antagonisms lose their force and power. Instead the new, vibrant and resonant conflicts develop within the technologies of information communication and learning; the technologies that form and structure the societal planning processes and that provide the information models and human labor required for conscious decision making. It is the relationship to these new "means of production and reproduction" that will determine the levels and quality of social conflict in the "post-industrial" age. Any program for radical research and action thus must transcend the old "industrial-class" framework of thought. Instead it must focus on the new technologies, and try to understand their transformative impact on communication, decision making, and learning. Only in this way can they be appropriated for human freedom.

§2. The End of Economics

Economics as a social science developed as a response to the new complexities of the industrial revolution. By analyzing market patterns and exchange relationships it hoped to provide insights into the elaborating forms of industrial technology and life. The demise of the market as the instrument for social change raises serious questions as to the relevance and centrality of economics as a science of society.

We can understand this problem better by distinguishing between three kinds of economic study and analyze, in turn, the impact of the "communications revolution" on each one. Economics can be divided into a social science of market forces, a policy science to guide governmental action (i.e., Keynesian economics) and an engineering science for the technocratic allocation of resources.

It is as a social science of market forces that economics is most seriously compromised. Its historic centrality as a social science rested on the centrality of markets as the tool for industrial and social development. When all social development and social conflict arose from the interplay of market forces, economics was the key for both contemplating and changing history. When market forces are increasingly delimited, and social change politicized within the new framework of communication information and learing, the study of economics as a key to history loses its centrality.

All this might appear particularly unjust if the analysis of "exchange relationships" were still a rich and fertile field of study. But it is probably safe to say that in the last half century no new and essential insights have been developed in this area. Despite the hope and program of the mathematical and neo-classical school, our concept of the invisible hand has not significantly changed. Instead general equilibrium theory has served to remind us of the remarkable set of mathematical restrictions required if market allocation is to prove "optimal". The general equilibrium theorists hoped that by developing and analyzing these restrictions, they could understand the real world workings of decentralized allocation but they have obtained few such insights. In fact, general equilibrium theory appears to be at a theoretical dead end, and one of its major theorists has wondered if the entire program was not misconceived.[7] One can surmise that if it had not been for the interest in the mathematics of maximization under constraints, the vacuum in neo-classical thought would have long been apparent.[8]

In retrospect, neo-classical economics will probably be judged as one of the more curious intellectual abberations of the twentieth century. Based on a nineteenth century mind set of utility maximization

---

[7]Frank Hahn, "Some Adjustment Problems", Econometrica, Jan., 1970, 1-17.

[8]The talmudic branch of economic theory known as the pure theory of international trade is the best example of the momentum provided by mathematics, independent of any contact with market realities. Based on a theorem enunciated by Ricardo almost two centuries ago, it rests on the myriad of permutations of the two-factor, two-commodity, two-country model. When confronted with the unfortunate "Leontief Paradox" international trade theorists were forced to redefine the concept of factor so as to make the entire theory tautological.

animal spirited entrepreneurs, and pure competition -- and set in a framework of mild mathematical interest -- it obscured the essential exhaustion of the analysis of exchange relationships.

Economics as a policy science, or Keynesian economics, has always rested on a strange amalgam of market and non-market orientations. As a science of governmental policy, it presupposed the ascendancy of a political governmental structure that stood above market forces. However, the specific application of monetary and fiscal policy was thought to take place <u>through</u> the market and thus stood outside politics. The economist was the "invisible dial turner" who affected only the level but not the structure of economic activity. Monetary and fiscal policy was thus politically neutral, and the economist was the invisible assistant of the invisible hand. In this way economists could come to grips with the rising non-market forces while at the same time insisting on the significance of their own skills as theoreticians of markets. Nor was this seemingly paradoxical resolution without foundation, since in fact neither market or non-market forces completely controlled the processes of resource allocation (though the vision of complete neutrality was an exercise in bad faith, since government spending can never be neutral, nor does monetary policy affect all sectors of investment spending evenly).[9]

However, the technological developments described in the last section -- the socialization of the means of production through

---

[9]One suspects, for instance, that the economist's opposition to wage and price controls flows not so much from his sense of their infeasibility but rather because they violate the sense of his own skills. Wage and price controls would entail the full acceptance of non-market control of economic development.

information, the developing future orientation -- destroy this uneasy resolution. Increasingly, the economic question becomes one of the structure and not of the level of spending, of the allocation of investments, of the pattern of social reproduction. Market forces appear insufficient and there develops a demand to take "technology in hand" and consciously guide the elaboration of the future.

Thus the economist is thrown willy nilly into a species of "social engineering", into a study of alternative investment patterns and their impact on future social structure, and it is here that his weakness is exposed. Not only is a broad interdisciplinary framework required if future social structures are to be anticipated -- a framework that has thus far eluded not only the economist, but all social scientists -- but the economist must confront the hopeless abstractions of his own neo-classical heritage (even in its "engineering" component of shadow prices, linear programming, etc.)[10] Moreover, the economist must confront the fact that no matter how many cost-benefit resource allocation studies he does, and whatever their merit, there exists as yet no force to take up and give political muscle and flesh-and-blood reality to his own recommendations and speculations. Just as a sensibility

---

[10]To quote the mathematical economist N. Georgescu-Roegen, "But no other profession has committed the sin of theorizing in a vacuum as often and with as complete an ingenuousness as the economist of the Classical and especially, the Neo-classical tradition". See Nicolas Georgescu-Roegen, "The Institutional Aspects of Peasant Communities: An Analytical View" in Subsistence Agriculture and Economic Development, ed., Clifton R. Wharton Jr., Aldine, Chicago, 1969. Leontief has hoped that this abstractionist tendency could be overcome with massive infusions of real live data. See, Wassily Leontief, "Theoretical Assumptions and Nonobserved Facts", American Economic Review, March 1971, 1-7.

always precedes the actual political development, so when the economist responds to the new sensibility -- to the future orientation -- he must confront the actual political vacuum, the absence of any force that can in fact carry out the process of rational resource allocation. Instead he must confront a government that sinks its investments into the military expenditure and expresses its power with air wars against semi-developed countries. All cost-benefit studies, all programs of efficient allocation can only appear weak and undirected when compared against a backgrop of eighty billion dollars of waste.

In effect, the vacuum that the economist faces in his own work, the exhaustion of neo-classical economics and the Keynesian pose, but the absence of any methodology and social structure to give a new discipline of social engineering real meaning, corresponds precisely to the present historical vacuum; the destruction of the ideology and efficacy of markets by the new information technologies, but the absence of any real political force to take up the new technologies in the interests of conscious and rational resource allocation. It is of course this gap that is responsible for the seemingly sudden dissolution of social structures and the appearance of all manners of "crises". A society as complex and uncoordinated as America cannot impose on itself new modes of coordination without deep political convulsions. The economist is, of course, part of this convulsion. The historical transition into "post-industrialism" destroys his own intellectual perspectives. He must confront fully

the end of economics if he is to avoid a creeping cynicism and skepticism as to the value of his own intellect. Only in this way can he develop new and future oriented skills that demonstrate full cognizance of his own changing social position.

Is there a Radical Economics?

One of the responses to the vacuum in economic thought has been the attempted development of a "radical economics", conceived of as both a political and methodological critique of conventional theory. Its political roots are abundantly clear, but its methodological substance is not. The radical economist shares in the general revulsion against the destructive use of American power and stands for its more humane and rational application. But he has not defined his methodological background. Many have gone back to the old Marxist texts in an attempt to find a structure that would satisfy their political sensibilities; but they face a closed chapter in intellectual history. Both the detractors and supporters of the labor theory of value have long ago filled out all areas of debate, and both long ago forgot that Marx was not an economist, and was not interested in contributing one more refined theory of price fluctuation. He sought instead a critique of political economy that would expose the social relationships that lay behind exchange relationships. In this way he hoped to bring a vibrant and developing industrial working class to a new consciousness, so that the means of production could be taken up and brought to their full potential.

In the absence of such a working class, the new radical economists are thrown into a kind of "sentimental" Marxism. They become obsessed with the "distribution of income", which while of social significance, can hardly provide the basis for a new analytic, a framework that will produce new insights and discoveries. They condemn themselves to rediscovering the same fact over and over, that in fact the rich are richer than the poor, and that power is distributed unequally.

Thus the conventional economists have probably been right. Radical economics has contributed more of a sensibility rather than a structure, though the conventional types have been all to unwilling to respond to that sensibility, to face the vacuum in their own work.[11]

In the end, the radical economist must face a central contradiction that lies at the root of his effort. He must realize that both political economy and the critique of political economy stand and fall together. In the nineteenth century, when industrialization became the central project of the West and markets the instruments for change, both political economy and its critique represented legitimate perspectives on historical change. They were the poles or antimonies through which the process could be contemplated and transformed. The demise of the market and the gradual disappearance of industrial class conflict means both the end of economics and its antimony.

---

[11] In their more cynical moments, the conventional will reply to a critique by a radical with "Das kennt jeder Esel", or "Every ass knows that", attributed apparently to Brahms. In other words, "we knew all along how silly we were being, so who says you're so smart anyway", which hardly answers the criticism.

The radical economist thus needs a new source of inspiration, and curiously enough, we can find it in Marx, but we must go beyond and deeper than the labor theory of value and the distribution of income. To Marx the central and motive contradition of capitalism lay not in wage labor, surplus value, or the distribution of income but rather in the tension between the productive and technological potential, and capitalism's inability to appropriate or realize this potential. The distribution of income, wage exploitation, and recurrent economic crises were significant only to the extent that they were indicators and signs of this fundamental tension. They were such indicators precisely because the problem of social reproduction, of the elaboration of technology was played out in the market. The demise of the market means that these indicators lose their social significance as signs of "revolutionary" tensions.

As we have already indicated the problem of social reproduction, of social change, is increasingly played out within the context of societal planning and within the framework of the new communication and information technologies. In fact, as the means of production are socialized through information, so does the technological potential become increasingly apparent to those who have access to the means of information and communication. Society learns of its technological potential in its full empirical complexity, it comprehends the gap between what is and what might be, as its own information technologies proliferate.

Thus, if radical social science is to explore the new dynamics of this old and fundamental tension, it must study the new technologies of communication information and learning. It must go beyond the simplistic "new working class" theories and explore the way in which these technologies are transforming the work, the skill structures and the sensibilities of those who have access to the new means of communication and information. It must analyze the dynamics of planned social change within the framework of these means of production and reproduction. It must understand the convulsive political process whereby the planned allocation of resource finally dominates the processes of economic and social change. It must confront both the totalitarian and anti-totalitarian potential implicit in this process.

The conventional economist is right when he points out that there is no "other" economics. But unfortunately, he does not go far enough; there is in fact no economics. There is a whole new area of study opened up by the new technologies that lies outside the traditional academic categories. If the radical economist commits his intellectual energies to this new field of study, he will help fulfill Marx's central program, appropriating the means of production for human freedom.

APPENDIX

THE BRIEFEST INTRODUCTION TO INFORMATION THEORY

The information measure is essentially a measure of dispersion. Intuitively, we can see that the greater the variance of the probability distribution of some variable, the more "uncertain" we are about it, and thus the greater the amount of information contained in any message that tells us of its exact location. Thus, where the variable is distributed with variance zero, there is no information in the message specifying its location. If we consider a six sided die, each face marked with a "snake-eyes", and if we let the variable be "the side that comes up with the toss of the die", then there will be no information in the message that specifies which face came up, since we know that snake eyes must appear.

Now consider a normal die with six different faces. How much information is contained in the message that specified which face came up? We might say that the message contains "six units" since it selects one out of six equi-probable possibilities. (Note that where the faces are not equally likely to come up, we could not in good conscience answer "six units" since the variance of the variable "which die will come up" will be less than when they were all equi-probable). Continuing in this way, we could say that the toss of two dice contained 6×6 or 36 units of information.

Now in the interests of "additivity" it will be more convenient to measure our information units in logs. Thus, the information

contained in the toss of a single die will be $\log_2(6)$, and in the toss of two dice $\log_2(36)$. Note then, that the information contained in the toss of two dice is simply twice the information contained in the toss of one die, since

$$\log_2(36) = \log_2(6 \times 6) = \log_2(6) + \log_2(6)$$

This is the sense in which information functions are additive.

Now it is often the convention to measure these logs to the base two. In this case the basic unit of measurement becomes the message that specifies the location of a variable in a set of two equi-probable states, since,

$$\log_2(2) = 1$$

This unit of measure is called a "bit" of information. The number of bits contained in the two of two normal dice will thus be $\log_2(36)$.

Now, how can we measure information when the "events" are not equi-probable. Imagine the following situation. Let our die have four faces marked with a snake eyes and two faces marked with a "six". There are thus two possibilities with probabilities $\frac{2}{3}$ and $\frac{1}{3}$ respectively. Now the amount of information contained in the die will clearly be less than $\log_2(6)$, but how much less?

One way of looking at this is to see that the toss of the die adds to the a priori information we have about the die. How much a priori information do we have? Well, intuitively we could say that since snake eyes occurs with a probability of $\frac{2}{3}$ and six with a probability of $\frac{1}{3}$, we have the following amount of a priori information,

$$\tfrac{2}{3}\log 4 + \tfrac{1}{3}\log 2$$

In other words, we have a set of four faces each occurring with equal probability and a set of two faces each occurring with equal probability. Taking a weighted average gives a measure of our a priori information. Thus the information gained in the message telling us which face came up is

$$\log 6 - \tfrac{2}{3}\log 4 - \tfrac{1}{3}\log 3$$

or more generally

$$\log N - (N_1/N)\log N_1 - (N_2/N)\log N_2$$

$$N_1 + N_2 = N$$

which can be written as

$$\frac{(N_1 + N_2)\log N - N_1 \log N_1 - N_2 \log N_2}{N}$$

which reduces to

$$-(N_1/N)\log N_1/N - (N_2/N)\log N_2/N$$

or alternatively

$$-P_i \log P_i - P_j \log P_j$$

where

$$P_i = N_1/N \qquad \text{and} \qquad P_j = N_2/N$$

So if we consider a probability distribution with states, i.j,... and probabilities P_i, P_j,... then the information value of this distribution, or alternatively its "uncertainty" value will be given as

$$H = -\sum P_i \log P_i$$

Note that H must always be positive since $\log P_i$ is of necessity negative.

The formula for the uncertainty of a distribution is suggestive of the following interpretation. The information value of a given state is proportional to the negative of the log value of its probability. Thus, the higher its probability, the lower its information value. (In other words, the more likely it is to occur, the less information we have when we are told that in fact it has occurred). The total information value of the entire distribution is then given by a weighted average of the states in the distribution.

Note that in the case where all the states are equi-probable, we would have

$$-\sum P_i \log P_i = -N(1/N) \log (1/N) = -\log(1/N) = \log N$$

which is consistent with our original definition. In other words, a distribution of N states will be at a "maximum" when all the states are equi-probable. Note that when we write the information value of a function, it is usually abbreviated as

$H(p_1,\ldots,p_n)$ where H is its information or uncertainty value.

The additivity of information functions means that we can "break down" the probability of its distribution into its constitutent parts. Thus, if we consider the following probability distribution of four states,

A	B	C	D
.25	.25	.25	.25

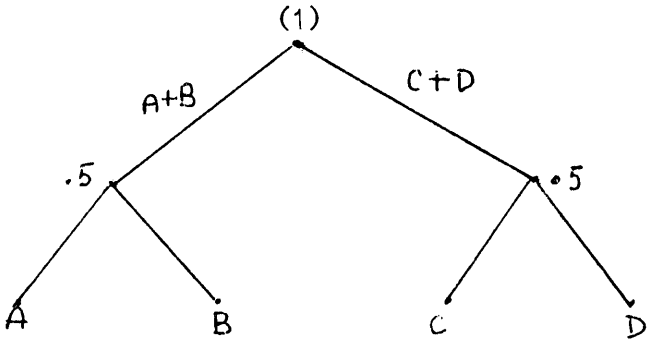
then

$$H(.25, .25, .25, .25) = H(.5, .5) + .5H(.5, .5) + .5H(.5, .5)$$

or

$$H(A,B,C,D) = H(A+B, C+D) + .5H(A,B) + .5(C,D)$$

This decomposition can be viewed as a tree structure



so that the total information value of the function is the weighted average of the information values at each node.

The formula for this specific case is easily confirmed

$$H(.25, .25, .25, .25) = -4(.25) \log .25 = -(-2) = 2 \text{ bits}$$

and

$$H(.5, .5) = -2(.5) \log .5 = -(-1) = 1$$

so that

so that

$$H(.5, .5) + .5H(.5, .5) + .5H(.5, .5) = 2 \text{ bits}$$

All the other formulations required for reading the text are explained in the body of the text itself.

BIBLIOGRAPHY

ALONSO, William: "Industrial Location and Economic Development", Working Paper No. 13, Center for Planning and Development Research, University of California at Berkeley.

ARROW, Kenneth: "The Firm in General Equilibrium", Technical Report #3, May 1969, Harvard University.

ASHBY, Ross W.: An Introduction to Cybernetics, John Wiley & Sons, New York, 1963.

-----: Design for a Brain, Chapman and Hall, Ltd., London, 1960.

BELL, Daniel: The End of Ideology, The Free Press, New York, 1960.

-----: "The Measurement of Knowledge and Technology" in E.B. Sheldon and W.E. Moore (eds.), Indicators of Social Change, Russell Sage Foundation, New York, 1968.

BERLE, Adolph A. and Gardiner C. MEANS: The Modern Corporation and Private Property, Harcourt Brace and World, New York, 1968.

BOULDING, Kenneth E.: Beyond Economics, University of Michigan Press, Ann Arbor, 1968.

-----: Conflict and Defense, Harper and Brothers, New York, 1962.

BRZEZINSKI, Zbigniew: Between Two Ages: America's Role in the Technetronic Era, Viking Press, New York, 1970.

BUSINESS WEEK: "A New Industry's Wild Ride", May 24, 1969.

-----: "Growing Wildly Without Blueprint", Dec. 6, 1969.

-----: "Who Has Ridden the Growth Curve", Jan. 4, 1969.

-----: "Two Gee Whiz Giants Go at Each Other", June 13, 1970.

THE DIEBOLD GROUP: Automation, Impact and Implications, Communications Workers of America, AFL-CIO, April 1965.

EDWARDS, Elwyn: Information Transmission, Chapman and Hall, London, 1964.

EMERY, James C: Organizational Planning and Control: Theory and Technology, Doctoral Dissertation, M.I.T., Jan. 1967.

GALBRAITH, John K.: The New Industrial State, Houghton and Mifflin, Boston, 1967.

GARDNER, Wendell R.: Uncertainty and Structure as Psychological Concepts, John Wiley & Sons, New York, 1962.

GREENBERGER, Martin: "Computer in the Organization", mimeo, Sloan School, M.I.T., 1966.

HAHN, Frank: "Some Adjustment Problems", Econometrica, Jan. 1970.

HAYEK, Friedrich A.: "The Nature and History of the Problem", in Collectivist Economic Planning, ed. by F.A. Hayek, Augustus M. Kelley, New York

-----: "Economics and Knowledge", Economica, Feb. 1937.

HURWICZ, Leonid: "Centralization and Decentralization in Economic Processes", mimeo, Harvard University, August 6, 1970.

JUDY, Richard W.: "Information, Control and Soviet Economic Management" in Mathematics and Computers in Soviet Economic Planning, ed. by J. Hardt, Yale University Press, New Haven, 1967.

LEIBENSTEIN, Harvey: "Entrepreneurship and Development", American Economic Review, Papers and Proceedings, May 1968.

-----: Economic Theory and Organizational Analysis, Harper and Brothers, New York, 1960.

-----: "Allocative Efficiency vs. X-efficienty", American Economic Review, June, 1966.

LEONTIEF, Wassily: "Theoretical Assumptions and Nonobserved Facts", American Economic Review, March 1971.

THE MANPOWER REPORT OF THE PRESIDENT, 1970.

MARSCHAK, Jacob: "Information Economics Reconsidered", Working Paper #149, Western Management Science Institute, U.C.L.A., June 1969.

-----: "The Theory of an Efficient Several Person Firm", Discussion Paper No. 82, Cowles Foundation for Research in Economics, Yale University, 1959.

MEIER, Richard L.: A Communications Theory of Urban Growth, M.I.T. Press, Cambridge, 1962.

POLANYI, Karl: The Great Transformation, Beacon Press, Boston, 1957.

QUASTLER, Henry and Edward McGILL: "Standard Nomenclature, an Attempt" in Information Theory and Psychology, ed. by Henry Quastler, The Free Press, Glencoe, 1955.

RADICAL SOFTWARE, No. 2, New York, 1971.

REZLER, Jues: Automation and Industrial Labor, Random House, New York, 1969.

RICHARDSON, Gerald B.: "Imperfect Knowledge and Economic Efficiency", Oxford Economic Papers, June 1953.

-----: "Equilibrium, Expectations, and Information", Economic Journal June 1959.

ROBINSON, Romney: "The Economics of Disequilibrium Price", Quarterly Journal of Economics, May 1961.

ROEGEN, Nicolas G.: "The Institutional Aspects of Peasant Communities: An Analytical View" in Subsistence Agriculture and Economic Development, ed. by Clifton R. Wharton Jr., Aldine, Chicago, 1969.

ROSE, Michael: Computers, Managers and Society, Penguin Books, Baltimore, 1969.

SCHUMPETER, Joseph A.: Capitalism, Socialism and Democracy, Harper and Row, New York, 1962.

SHANNON, Claude E. and Warren WEAVER: The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1949.

SHONFIELD, Andrew: Modern Capitalism, Oxford University Press, London, 1965.

SIMON, Herbert A.: The Sciences of the Artificial, M.I.T. Press, Cambridge, 1969.

-----: Models of Man, John Wiley & Sons, New York, 1957. See "A Formal Theory of the Employment Relation".

-----: The New Science of Management Decision, Harper and Row, New York, 1960.

SINGH, Jagjit: Great Ideas in Information Theory, Language and Cybernetics, Dover Publications, New York, 1966.

SMITH, Ralph L.: "The Wired Nation", The Nation, May 18, 1970.

STEVENS, Chandler H.: Information and the Division of Labor, Doctoral Dissertation, M.I.T., Jan. 1967.

-----: "Citizen Feedback Systems", paper presented at the Eastern Regional Conference on Science and Technology for Public Purposes, April, 1970.

STIGLER, George, "The Economics of Information", Journal of Political Economy, June, 1961.

TOFFLER, Alvin: Future Shock, Random House, New York, 1970.

VERNON, Raymond and Edgar HOOVER: The Anatomy of a Metropolis, Doubleday and Co., New York, 1962.

ZANNETOS, Zennon: "Management Information Systems: New Directions", mimeo, Sloan School, M.I.T., March 1967.

BIOGRAPHICAL NOTE

NAME: Lawrence A. Hirschhorn

BORN: July 2, 1945

ATTENDED: Brandeis University, 1963-1967; B.A. in Economics, 1967. Massachusetts Institute of Technology, Graduate Program in Economics 1967-1971, Ph.D. in Economics, 1971.

TAUGHT: Brandeis University, Instructors of Economics, 1968-1970.